Genomic Applications in Pathology

George Jabboure Netto
Karen L. Kaul
Editors

Second Edition

Springer

Genomic Applications in Pathology

George Jabboure Netto • Karen L. Kaul
Editors

Genomic Applications in Pathology

Second Edition 2019

Springer

Editors
George Jabboure Netto
Professor and Chair of Pathology
University of Alabama at Birmingham
Birmingham, AL
USA

Karen L. Kaul
Department of Pathology and Laboratory
Medicine, NorthShore University HealthSystem
Clinical Professor of Pathology
University of Chicago Pritzker
School of Medicine
Evanston, IL
USA

ISBN 978-3-319-96829-2 ISBN 978-3-319-96830-8 (eBook)
https://doi.org/10.1007/978-3-319-96830-8

Library of Congress Control Number: 2018962532

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In the memory of my father whose wisdom will always guide me.
—George Jabboure Netto

With love and thanks to Tony and our children Nick, Emilia and Alexis.
—Karen L. Kaul

Preface

The pathologist has an increasingly central role in the management of cancer patients in the era of personalized oncology. Molecular diagnostic and genomic applications are rapidly penetrating the daily practice of the pathologist as the list of actionable genetic alterations in solid and hematologic malignancies continues to expand. At the same time, a paradigm shift in the diagnostic approach for inherited genetic diseases, infectious diseases, and pharmacogenetics is unfolding. As a result, a plethora of clinical genomic applications is being rapidly implemented in diagnostic molecular pathology laboratories as we move closer to the anticipated reality of "precision medicine."

This textbook provides a comprehensive resource of genomic applications to practicing molecular pathologists and hematopathologists, general and subspecialized practicing pathologists, as well as pathology trainees. The target audience also includes oncologists, geneticists, and other medical and surgical clinicians. The 33 chapters encompass a state-of-the-art review of the scientific principles underlying current and emerging genomic technologies and the bioinformatics approaches required to effectively analyze the daunting amount of data generated by next-generation sequencing. Implementation roadmaps for various clinical assays including single gene, gene panel, whole exome, and whole genome assays are addressed. Topics related to reporting and the pathologist's and laboratorian's role in the interpretation and clinical integration of genomic test results are discussed. Practice-related considerations including the regulatory framework, reimbursement, and legal and ethical issues as related to genomic testing are also included. Importantly, chapters on genomic applications for site-specific solid tumors and hematologic and lymphoid neoplasms provide a review with practical and actionable information regarding the latest advances. Finally, genomic applications in pharmacogenomics, inherited genetic diseases, and infectious diseases are also highlighted.

As this most exciting field continues to evolve rapidly, the information in this textbook provides an up-to-date framework for the transition of next-generation sequencing applications from bench to bedside, for genomic assay development, and for responsible implementation of genome-scale testing. We hope that you will enjoy the keen insights from our 62 expert authors and that this text will prove to be a valuable tool in your practice, as it is to ours.

Birmingham, AL, USA — George Jabboure Netto
Palo Alto, CA, USA — Iris Schrijver

Preface

In the few short years since the first edition of our volume on *Genomic Applications in Pathology*, the list of actionable genetic alterations in solid and hematologic malignancies has continued to expand in an unparalleled pace. Equally, momentous advances of the molecular approaches to the study of inherited genetic diseases and pharmacogenetics are unfolding. "Precision pathology" is now an integral part of the practice of "precision medicine."

The current expanded 39 chapters' volume provides the most up-to-date comprehensive discussion of established and emerging genomic technologies and their clinical implementation in molecular diagnostics. As in its first edition, the book places significant emphasis on implementation roadmaps for various clinical assays including single gene, gene panel, transcriptome sequencing, circulating tumor cells and cell-free DNA sequencing, whole exome, and whole genome assays. Detailed guidance on the central role of the pathologist in the interpretation, reporting, and clinical integration of genomic tests is provided. Expert opinions to help navigate growing compliance, reimbursement, and legal and ethical issues are shared in dedicated chapters.

The latest advances in genomic applications in oncologic diseases are addressed in an organ-based format covering the entire spectrum of solid and hematologic neoplasms according to the most current practice guidelines. Dedicated chapters to genomic applications in inherited diseases, sequencing cell-free DNA in maternal circulations, infectious diseases, pharmacogenomics, and the microbiome are also provided.

In a collective effort of 98 expert authors, our textbook will serve as a comprehensive resource for practicing molecular pathologists, general and subspecialized practicing anatomic and clinical pathologists, as well as pathology trainees. The wider target audience continues to include oncologists, geneticists, and other medical and surgical clinicians.

Birmingham, AL, USA — George Jabboure Netto
Evanston, IL, USA — Karen L. Kaul

Contents

Contributors

Nishant Agrawal, MD Department of Surgery, Section of Otolaryngology – Head and Neck Surgery, University of Chicago Medicine, Chicago, IL, USA

Aleodor A. Andea, MD, MBA Departments of Pathology and Dermatology, University of Michigan, Ann Arbor, MI, USA

Matthew W. Anderson, MD, PhD Diagnostic Laboratories, BloodCenter of Wisconsin, part of Versiti, Milwaukee, WI, USA

Arunkanth Ankala, PhD Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

Wyatt M. Arnold, BS, BA Department of Surgery, University of Chicago, Chicago, IL, USA

Biosciences Division (BIO), Argonne National Laboratory, Lemont, IL, USA

Marija Balic, MD PhD Division of Oncology, Department of Internal Medicine, Medical University of Graz, Graz, Austria

Niaz Banaei, MD Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Department of Pathology, Stanford Health Care, Stanford University School of Medicine, Stanford, CA, USA

Julia A. Beaver, MD The Sidney Kimmel Comprehensive Cancer, Center at Johns Hopkins, Baltimore, MD, USA

Joseph A. Bellairs, MD Department of Otolaryngology – Head and Neck Surgery, University of Washington Affiliated Hospitals, Seattle, WA, USA

Ioana Berindan-Neagoe, PhD Research Center for Functional Genomics, Biomedicine and Translational Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Medfuture Research Center for Advanced Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Department of Functional Genomics and Experimental Pathology, The Oncology Institute "Prof. Dr. Ion Chiricuta", Cluj-Napoca, Romania

D. Hunter Best, PhD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Robert G. Best, PhD Biomedical Sciences, University of South Carolina School of Medicine/ Greenville Health System, Greenville, SC, USA

Justin A. Bishop, MD Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA

Cornelia Braicu, PhD Research Center for Functional Genomics, Biomedicine and Translational Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

James D. Brenton, PhD FRCP Functional Genomics of Ovarian Cancer Laboratory, Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

Alan F. Brown, MD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Lynn Bry, MD, PhD Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Reinhard Büttner, MD Center for Integrated Oncology, Institute of Pathology, University Hospital Cologne, Cologne, Germany

George A. Calin, MD, PhD Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Fátima Carneiro, MD, PhD Ipatimup/i3S, Faculty of Medicine of the University of Porto and Centro Hospitalar Sao Joao, Porto, Portugal

Samya Chakravorty, PhD Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

Sarah Chiang, MD Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Richard J. Cote, MD, FRCPath, FCAP Department of Pathology, University of Miami Miller School of Medicine, Miami, FL, USA

Lauren E. Cralle Department of Surgery, University of Chicago, Chicago, IL, USA

Biosciences Division (BIO), Argonne National Laboratory, Lemont, IL, USA

David K. Crockett, PhD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Leticia De Mattos-Arruda, MD, PhD Department of Medical Oncology, Vall d'Hebron Institute of Oncology, Vall d'Hebron University Hospital, Barcelona, Spain

Universitat Autònoma de Barcelona, Barcelona, Spain

Beatriz de Melo Maia, MSc, PhD Research and Development, Phd Laboratory – Surgical and Molecular Pathology, São Paulo, SP, Brazil

Joshua L. Deignan, PhD Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Michael O. Dorschner, PhD Department of Pathology, UW Medicine Center for Precision Diagnostics, Northwest Clinical Genomics Laboratory, Seattle, WA, USA

Dorraya El-Ashry, PhD Department of Laboratory Medicine and Pathology, Masonic Comprehensive Cancer Center, University of Minnesota, Minneapolis, MN, USA

Na Fei, PhD Department of Surgery, University of Chicago, Chicago, IL, USA

Surgery, University of Chicago Medical Center, Chicago, IL, USA

Kevin E. Fisher, MD, PhD Pathology and Immunology, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA

Jack A. Gilbert, PhD Department of Surgery, University of Chicago, Chicago, IL, USA
Surgery, University of Chicago Medical Center, Chicago, IL, USA
Biosciences Division (BIO), Argonne National Laboratory, Lemont, IL, USA
Marine Biological Laboratory, Woods Hole, MA, USA

Thomas J. Giordano, MD, PhD Divisions of Anatomic Pathology and Molecular & Genomic Pathology, Departments of Pathology and Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, MI, USA

Christopher D. Gocke, MD Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Henry T. Greely, JD Center for Law and the Biosciences, Stanford University, Stanford, CA, USA

Wayne W. Grody, MD, PhD Departments of Pathology and Laboratory Medicine, Pediatrics, and Human Genetics, UCLA School of Medicine, Molecular Diagnostic Laboratories and Clinical Genomics Center, UCLA Medical Center, Los Angeles, CA, USA

Diana Gulei, MSc Medfuture Research Center for Advanced Medicine, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania
Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Tina M. Hambuch, PhD Genetics, Invitae, San Francisco, CA, USA

Wolfgang Hartmann, MD Gerhard-Domagk-Institute of Pathology, University Hospital Münster, Münster, Germany

Richard L. Haspel, MD, PhD Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

Madhuri R. Hegde, PhD Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

Carina Heydt, PhD Center for Integrated Oncology, Institute of Pathology, University Hospital Cologne, Cologne, Germany

Charles E. Hill, MD, PhD Pathology and Laboratory Medicine, Emory University Hospital, Atlanta, GA, USA

Elle Simone Hill, BS Department of Surgery, University of Chicago, Chicago, IL, USA
Surgery, University of Chicago Medical Center, Chicago, IL, USA

Hao Ho, MD, PhD Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Ralph H. Hruban, MD The Sol Goldman Pancreatic Cancer Research Center, Departments of Pathology and Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Natasha B. Hunter, MD The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

Mohammad Omar Hussaini, MD Department of Hematopathology and Laboratory Medicine, Moffitt Cancer Center, Tampa, FL, USA

Tamara Jamaspishvili, MD, PhD Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada
Division of Cancer Biology and Genetics, Queen's University, Kingston, ON, Canada

Karen L. Kaul, MD, PhD Department of Pathology and Laboratory Medicine, NorthShore University HealthSystem, Clinical Professor of Pathology, University of Chicago Pritzker School of Medicine, Evanston, IL, USA

Roger D. Klein, MD, JD Faculty Fellow, Center for Law, Science and Innovation, State University, Sandra Day O'Connor College of Law, Arizona State University, Tempe, AZ, USA

Martin Köbel, MD Pathology and Laboratory Medicine, Foothills Medical Centre, Calgary, AB, Canada

Bryan L. Krock, PhD The Children's Hospital of Philadelphia, Division of Genomic Diagnostics, Perelman School of Medicine at the University of Pennslyvania, Department of Pathology and Laboratory Medicine, Philadelphia, PA, USA

Jamie Ahn Ku, MD Head and Neck Institute, Head and Neck Surgery and Oncology, Cleveland Clinic Foundation and Lerner College of Medicine at Case Western Reserve University, Cleveland, OH, USA

Alexander J. Lazar, MD, PhD Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Departments of Genomic Medicine & Dermatology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Martina I. Lefterova, MD, PhD Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Guardant Health, Redwood City, CA, USA

Janina A. Longtine, MD Departments of Pathology and Laboratory Medicine, Yale University School of Medicine, Pathology and Laboratory Medicine, Yale New Haven Hospital and Smilow Cancer Hospital, New Haven, CT, USA

Elaine Lyon, PhD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Howard L. McLeod, PharmD, FCCP Department of Cancer Epidemiology, Individualized Cancer Medicine, Moffitt Cancer Center, Tampa, FL, USA

Department of Individualized Cancer Medicine, Moffitt Cancer Center, Tampa, FL, USA

Martin J. Magers, MD Department of Pathology, University of Michigan Medical School, Ann Arbor, MI, USA

Department of Pathology, Indiana University Health, Indiana University School of Medicine, Indianapolis, IN, USA

Rong Mao, MD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Luciano G. Martelotto, PhD Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

The University of Melbourne Centre for Cancer Research, University of Melbourne, Victorian Comprehensive Cancer Center, Melbourne, VIC, Australia

Mariana Salas Garcia, BA Department of Surgery, University of Chicago, Chicago, IL, USA

Surgery, University of Chicago Medical Center, Chicago, IL, USA

Carri-Lyn Rebecca Mead, PhD Emerging Technologies, Illumina, San Diego, CA, USA

Sabine Merkelbach-Bruse, PhD Center for Integrated Oncology, Institute of Pathology, University Hospital Cologne, Cologne, Germany

Federico A. Monzon, MD Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA

Castle Biosciences, Friendswood, TX, USA

Jan A. Nowak, PhD, MD Department of Pathology and Laboratory Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

Jonathan Nowak, MD, PhD Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Keith Nykamp, PhD Genetics, Invitae, San Francisco, CA, USA

Glenn E. Palomaki, PhD Division of Medical Screening and Special Testing, Department of Pathology and Laboratory Medicine, Women and Infants Hospital/Alpert Medical School of Brown University, Providence, RI, USA

Savjani Institute for Health Research, Windham, ME, USA

Fresia Pareja, MD, PhD Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Ben Ho Park, MD, PhD The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

Benjamin A. Pinsky, MD, PhD Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Guardant Health, Redwood City, CA, USA

Carlos N. Prieto-Granada, MD Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA

Jorge S. Reis-Filho, MD, PhD, FRCPath Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Jacquelyn Reuther, PhD Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA

Lauren L. Ritterhouse, MD, PhD Department of Pathology, University of Chicago, Chicago, IL, USA

Angshumoy Roy, MD, PhD Department of Pathology and Immunology and Pediatrics, Baylor College of Medicine, Houston, TX, USA

Wade S. Samowitz, MD Department of Pathology, University of Utah, Salt Lake City, UT, USA

Kabeer K. Shah, DO Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Anthony Sireci, MD Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA

Matija Snuderl, MD Department of Pathology, NYU Langone Medical Center and Medical School, New York, NY, USA

Jeremy A. Squire, PhD, MSc Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada

Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil

Todd M. Stevens, MD Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA

Rachel L. Stewart, DO, PhD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Department of Pathology and Laboratory Medicine, University of Kentucky College of Medicine, Lexington, KY, USA

Carlos J. Suarez, MD Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

John R. ten Bosch, PhD TPMG Regional Genetics Laboratory, Department of Genetics, Kaiser Permanente Medical Center, San Jose, CA, USA

Scott A. Tomlins, MD, PhD Departments of Pathology and Urology, University of Michigan Medical School, Ann Arbor, MI, USA

Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, MI, USA

Rogel Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA

Tatiana Tvrdik, PhD, LCGC ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

John Van Arnam, MD Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX, USA

Karl V. Voelkerding, MD ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA

Eva Wardelmann, MD Gerhard-Domagk-Institute of Pathology, University Hospital Münster, Münster, Germany

Joshua I. Warrick, MD Departments of Pathology and Surgery, Penn State College of Medicine, Hershey, PA, USA

Britta Weigelt, PhD Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Alyson L. Yee, BS Department of Surgery, University of Chicago, Chicago, IL, USA

Surgery, University of Chicago Medical Center, Chicago, IL, USA

Jessica Yesensky, MD Department of Surgery, Section of Otolaryngology – Head and Neck Surgery, University of Chicago Medicine, Chicago, IL, USA

Linsheng Zhang, MD, PhD Pathology and Laboratory Medicine, Emory University Hospital, Atlanta, GA, USA

Part I

Genomic Technologies

Practicing Pathology in the Post-genomic Era: Challenges and Opportunities

1

Karen L. Kaul

Introduction

Medicine and the field of pathology are both rapidly changing. In this era following sequencing of the human genome, rapid advances in knowledge applicable to patient care occur constantly, as the molecular basis for constitutional and somatic disease is elucidated, along with targets for more effective treatment. This is the era of precision medicine. Considerable technologic advances coupled with our rapidly expanding knowledge allow clinical laboratories to provide more information than ever before, within a time frame that allows efficient patient care and improved outcomes. The combination of new knowledge regarding the molecular basis of disease, the novel technologies, and the growing ability to use archival formalin-fixed paraffin-embedded (FFPE) samples and those that are obtainable by noninvasive methods will facilitate further growth in this area. This genomic information, along with treatments targeting the specific molecular defect(s) causing the disease, makes up the new discipline known as precision medicine.

Many challenges come along with the tremendous potential offered by precision medicine. The rapid evolution of knowledge means that clinical laboratories must often update assays to be consistent with standard of care as defined by consensus guidelines and routine practice. The pathology community, along with clinical practitioners, must become aware of these new advances and fluent in their applications to patient care. Regulatory oversight and limitations to reimbursement continue to present challenges to clinical laboratories performing genomic testing. Outcome studies illustrating the advantages of incorporating genomic data into the care of patients are sorely needed.

K. L. Kaul (✉)
Department of Pathology and Laboratory Medicine, NorthShore University HealthSystem, Clinical Professor of Pathology, University of Chicago Pritzker School of Medicine, Evanston, IL, USA
e-mail: kkaul@northshore.org

This chapter will outline the current state and major issues facing genomic pathology, including these opportunities and challenges, as a preface to the more detailed discussion in the subsequent chapters.

Molecular Targets

Precision medicine entails measurement of personalized information for each patient using a host of new diagnostic tools made possible by recent advances in analytic methods. DNA, RNA, and protein targets have been employed thus far. The speed and cost of analysis, coupled with the ability to accurately analyze formalin-fixed paraffin-embedded tissue samples, have led to the predominant use of DNA analysis, fluorescent in situ hybridization (FISH), and immunohistochemical (IHC) stains, though mRNA and miRNA targets are also utilized.

Genomic or DNA Targets

DNA remains the primary analytic target used in precision diagnostics. DNA offers the advantages of stability, relative ease of recovery, and extensive and growing knowledge regarding mutations and alterations that are clinically relevant. These alterations include single nucleotide substitutions or mutations, copy number variants, deletions, translocations, and other chromosomal rearrangements [1]. General analytic approaches and specific platforms differ in their ability to detect these abnormalities, so assay selection will require understanding of the data needed as well as the clinical application.

Specific genomic alterations are known to have a causal role in the development of many tumors [1]. Data from The Cancer Genome Atlas (TCGA) project and other efforts has been instrumental in reshaping the classification of many cancers [2]. In 2016, for example, new World Health Organization (WHO) classifications for brain tumors and

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_1

lymphoma were published, demonstrating that knowledge of genomic changes is fundamental to the diagnosis and prognosis of a growing number of tumor types [3, 4].

Beyond diagnosis, specific knowledge of genomic alterations is often necessary for appropriate treatment planning. For an increasing number of malignancies, a standard and required part of the pathology report is the notation of important gene alterations that make the tumor susceptible to targeted therapy. The prototypic example in solid tumors is KRAS, a GTPase critical in signal transduction, which is mutated in a wide range of tumor types [5]. A landmark study presented at the American Society of Clinical Oncology (ASCO) meeting over a decade ago reported that patients with metastatic colorectal cancer harboring a mutated KRAS failed to respond to targeted therapy with cetuximab [6, 7]. Molecular pathology labs worked quickly to develop and validate reliable clinical KRAS assays [8, 9]. In 2009, the National Comprehensive Cancer Network (NCCN) and ASCO together recommended mutational profiling of KRAS exons 12 and13 before institution of anti-EGFR therapy for patients with metastatic colorectal cancer. It is since the standard of care to assess formalin-fixed paraffin-embedded tumor tissues from patients with metastatic colon cancer for *KRAS* mutation status [10, 11]. A few years later, new data demonstrated that mutation analysis of other RAS genes was also needed [12, 13]. With the constant evolution of the molecular knowledge base, molecular pathology diagnostic laboratories must be prepared to adjust laboratory assays and clinical practice accordingly and ensure quality assay performance as clinical needs evolve [14].

RNA

While DNA is perhaps most frequently analyzed molecular target, mRNA can also be extremely useful. mRNA offers the potential advantage of smaller analytic size since it is generated post-splicing of exons. It thus can allow a more simplified approach to analysis of large genes with many exons or detection of translocations or deletions. However, mRNA is labile and will degrade at variable rates following sample collection or tissue devascularization. Pre-analytic issues such as ischemic and transport time may thus have a significant impact on the quality of the results [15]. Recently, smaller noncoding RNA molecules known as miRNA have been studied for potential clinical use; miRNA offers the advantage of high stability in tissue and other biologic samples, along with effective recovery from FFPE samples. miRNA is also variably expressed in different tissues and tumors, allowing expression profiling as a research and clinical tool; this approach has been used to generate risk scores predictive of tumor progression [16, 17]. Transcriptome analysis by next-generation sequencing (NGS), such as RNA-Seq, is an important approach to the study of complex gene expression and is increasingly playing important role in biomarker analysis and discovery.

Protein

Pathologists have been using proteins and protein expression as an adjunct to histology for decades. The use of single target immunostains for cellular proteins became routine in the 1980s, with multiplexed detection developed thereafter [18]. More advanced proteomic approaches using mass spectrometry on tissue sections are under development and may impact the practice of anatomic pathology in the future [19]. Protein-directed approaches may be quite complementary to genomic studies as genomic and transcriptomic alterations ultimately impact protein structure and expression.

Analytic Methods

Laboratories use a variety of methods to detect mutations and other genomic abnormalities. Analytic approaches have advanced significantly in recent years. Sanger sequencing, pyrosequencing, allele-specific amplification, PCR-melt curve analysis, and multiplexed methods such as SNaPshot, dHPLC, and SNP arrays have all been utilized to characterize tumors for mutations [20]. The development of NGS platforms that address the speed, cost, and throughput needs of the clinical laboratory settings has paved the way for NGS to become a routine approach used for analysis of tumors and germline samples. Simultaneous interrogation of multiple genes, whether by targeted gene panels or broader analysis, is a more efficient and cost-effective way to profile tumors and other samples and make optimal use of the small-sized samples often available.

At present, two commercial platforms have captured the majority of the clinical market: most labs are utilizing either Illumina or ThermoFisher Ion Torrent technology for NGS. These platforms differ with respect to chemistry, DNA input requirements, time for analysis, and sample throughout. They have differing strengths and weaknesses for clinical analysis [21–23]. Each manufacturer offers a range of platforms suitable for small, targeted analyses to larger, high-throughput, and whole genome analysis. Novel, more rapid, and single molecule approaches may reach mainstream clinical utilization in the future. A summary of the features of various NGS platforms is provided in Table 1.1 [24].

Table 1.1 Summary of currently available instruments for clinical next generation sequencing

Manufacturer	Model	Sequencing chemistry	Analytic capacity	Analytic time	Instrument cost
Illumina					
	MiniSeq	Synthesis	7.5 Gb	4–24 h	49K
	MiSeq	Synthesis	15 Gb	1–2 days	100K
	NextSeq	Synthesis	120 Gb	1–3 days	
	HiSeq	Synthesis	1500 Gb	3 days	125K
	Firefly	Single channel Semiconductor	1 Gb	3–13 days	30K
ThermoFisher Ion Torrent					
	PGM	Semiconductor	30 MB–2 Gb	2–7 h	80K
	S5	Semiconductor	0.5–15 Gb	2–4 h	65K
	Proton	Semiconductor	10–15 GB	2 h	
Pacific Biosciences		Long-read sequencing			
	Sequel		0.5–1 Gb		350K
	RSII		5–10 Gb	20 min	700K
Oxford Nanopore		Single molecule Nanopore sensing			
	PromethION		128 Gb	1–48 h	Minimal
	Minion		21 Gb	1–48 h	Minimal

Adapted by permission from Springer Nature, Kaul [24], and from Perkel and Fung [39], and corporate websites

Interpretation of Data/Informatics Pipeline

The immense amount of complex data generated by next-generation sequencing instrumentation is a challenge to laboratories and healthcare institutions, requiring concomitant investments and advances in informatics and laboratory systems able to handle this data. Interpretation of the data is also a challenge, requiring multiple steps and informatics tools used in tandem to generate a final result from the raw data. Collectively, these software tools are called the "bioinformatics pipeline" and are a critical part of sample analysis and generation of results [25]. The goal for clinical laboratories performing NGS is that the correct result be generated in each analysis and in each laboratory. The complexity of the informatics pipeline for NGS, and the fact that labs may use different components and settings in the pipeline, makes this a potential source of variation. Thus, clinical validation of an NGS assay requires validation of the bioinformatics tools in addition to the "wet lab" portions of the assay that generate the raw sequence data.

Regardless of the sequencing platform used, primary data interpretation begins with raw data requiring signal/noise determination and production of sequence reads, which leads to generation of a FASTQ file. Quality scores can also be generated at this step and can be used for filtering of poor-quality results. Sequence reads are next assembled and aligned, and variant calls are performed, resulting in the VCF file, with generation of additional quality indicators such a depth of coverage. Lastly, variants are annotated, and associated gene variants and other alterations are done via database searches [20]. For clinical reporting, links to information on clinical utility and clinical trials are often useful and may be included in reports.

While the sequences generated will remain static, the interpretation may change over time as new information is gathered regarding variants, adding complexity to the management of genomic data. Similarly, the clinical significance of recognized mutations and variants may evolve as new treatments and trials develop. Thus, unlike other laboratory reports, the interpretive component of NGS studies may be subject to reinterpretation, in which case careful tracking of versions and dates will be important for clinical and potentially liability purposes. While germline sequencing results will remain constant (though the clinical importance of variants will change), the genomes of tumors are known to evolve, particularly under the selective pressure of treatment. No current guidelines exist that recommend schedules for re-biopsy and NGS analysis.

Scope of Analysis

Studies suggest that the human genome includes 30,000 genes, of which 20,000 might be involved in carcinogenesis [3]. These genes sort into 1 of 12 classes such as signal transduction, cell cycle control, and other functions. One of the evolving questions is how broad an analysis should be performed for somatic or constitutional disorders. Genomic analysis can target a single locus, a gene panel, an exome, or a whole genome. The current clinical needs for analysis of

multiple genes have driven the move to NGS in place of individual gene assays, but there remains discussion as to what is "actionable" and how many genes to include. Consensus treatment guidelines generally include multiple gene targets for which strong evidence of clinical utility has accumulated [26, 27]. Beyond that, there are many genes that might be included but which lack definitive evidence such as that obtained from a prospective clinical trial. Analysis beyond a targeted gene panel might include these genes implicated in cancer or perhaps the entire coding or transcribed portion of the genome. This whole exome sequencing (WES) would encompass all coding regions of the genome, similar to transcriptome analysis (which starts with reverse transcription and amplification of mRNA). Both would yield information on mutations, while transcriptome analysis also yields information on gene expression [28]. These analyses would not include regulatory regions, sequences that might alter splicing, or other areas within the so called dark matter, which constitutes the majority of the genome and which to date we know relatively little about. The most expansive analysis would of course be sequencing the whole genome (WGS), which at this time is generally viewed as a research tool. These approaches are summarized in Table 1.2 [24].

Most clinical sequence analysis currently covers a panel of 50–500 genes inclusive of known hotspots and genes implicated in cancer. The more broad panels include genes for which less is known and may thus generate variants of undetermined significance (VUS), which present interpretive challenges for clinical labs. Beyond clinical utility, cost and reimbursement issues may influence the chosen size of the NGS panel. Conversely, broader analysis such as WGS offers the opportunity to gain discovery data on tumor samples that may add to our knowledge and future clinical use. As our knowledge expands and the cost of analysis declines, it is anticipated that broad genomic analysis will become more routine.

Other approaches may be needed to fully characterize genomic abnormalities in tumors, providing complementary information to NGS. Complete analysis may require additional approaches such as fluorescent in situ hybridization, arrays, or other approaches for assessment of structural abnormalities and copy number variations. Novel sequence variations may require extensive in silico and even functional studies to determine significance.

An additional consideration is the sequencing of paired tumor and normal DNA. While the allele frequency in NGS can provide evidence regarding whether a nucleotide variant is somatic or germline, certainty may require the paired analysis of both tumor and non-tumor samples.

Genomics Education

Training Students

As we enter an era in which incorporation of genomic information becomes a routine part of medical care, all physicians will need a basic education in the molecular basis of disease, diagnosis, treatment, and disease monitoring. Medical schools are beginning to incorporate this foundation [29].

Training Clinicians

These advances are broadly revolutionizing the practice of medicine, especially oncology. The pace of change, the rapid accumulation of new knowledge, and the fact that few of our clinical colleagues have been adequately trained in genomic medicine pose challenges for the field. Ongoing professional education, the development of consensus diagnostic and treatment guidelines, and the generation of

Table 1.2 Genomic analysis in tumors

	Targeted analysis	Whole exome	Whole genome	Transcriptome
Includes	Known hotspots 50–200 genes	Coding regions of 20 K genes	Coding and noncoding	Transcribed genes (mRNA)
Detects	Miss new mutations	Translocations Splice variants	Regulatory regions Structural variants	Expression patterns Pathway analysis Splice variants
Utility	Established clinical utility	Clinical utility evolving Discovery?	Clinical utility? Discovery Variants of unknown significance	Clinical Utility evolving Discovery
Analysis	Complex Known targets	More complex Map to genes/pathways	Very complex Map to genes/pathways Determine relevance of variants	Complex Map to pathways
Cost	$	$$	$$$	$$
Clinical QC	Clinical QC, PT available	?	??	??

Adapted from [24]

complete and detailed reports will be key to the success of molecular oncology.

Application of genomics will also impact primary care as we move forward.

There will be significant needs for pathologist consultation regarding the choice of tests, results interpretation, and appropriate test utilization, and education of the pathology and medical workforce will be critical.

Training Pathologists

Molecular characterization of tumors has become part of the routine practice of pathology. As such, residents, fellows, and other trainees must become familiar and competent in the use of this information and perhaps for some in the performance and interpretation of these assays. Determination of what knowledge and skills are needed at each level of training will require broad discussion and planning and is likely to evolve over time. Consensus curricula have been developed to facilitate such training [30].

At present, AP and AP/CP residents should acquire a general knowledge of methods for molecular assessment of tumors, with emphasis on pre-analytic issues such as sample selection and post analytic topics including reporting and consultation at the level of a tumor board presentation. Fellows in anatomic pathology subspecialties should have deeper knowledge for those genes and procedures most relevant to their focus. Molecular genetic pathology fellows should be able to choose the best method for assessment of the sample, given the sample type and information needed, be able to interpret these data, and have an understanding of assay validation and quality assessment.

Reimbursement

Reimbursement remains a challenge for many genomic tests. In general, reimbursement requires demonstration of clinical utility and recognition of such by third-party payers. In the United States, this often equates to lack of reimbursement of a test that is used routinely or is standard of care. US labs also experience difficulties with reimbursement of testing done on samples that are collected during an inpatient stay and for testing for which the CPT codes are not assigned valuation or for which assigned values do not adequately describe the test and its worth [31]. Detailed outcome studies which illustrate the impact of molecular and genomic testing on the avoidance of treatments that will not be effective, on the more rapid use of effective targeted treatments, and on patient outcome Reimbursement overall are very much needed.

Quality and Regulation

Laboratories bear the responsibility to offer testing that is safe, efficacious, and of high quality for patient care. In the rapidly changing field of genomic medicine, constantly evolving information and practices are the reality, presenting challenges for laboratories, manufacturers, and regulatory agencies alike. In the United States, the data and studies needed to support an FDA submission are significant, and the approval process is sometimes slow and costly, yielding test kits that are behind standard practice. Laboratories in the United States and Europe offer tests that have been optimized and validated by the labs themselves according to consensus quality guidelines; this is a more nimble process that is better suited to the rapid advances seen in genomic medicine and demonstrates excellent results [14, 32]. In Europe, interlaboratory standardization, quality, and training programs such as UK NEQAS facilitate high-quality laboratory testing [33]. The laboratory community recognizes the need for standardization of these tests and has generally worked diligently to ensure that tests offered are of the highest quality and that results generated in one lab match those in another. This has required extensive cooperation and sharing of data and samples and has led to peer-reviewed publications that delineate best practices and serve as models for other labs as they set up testing [34–36]. Collectively, these efforts raise the quality in the field as a whole and help to identify the most effective loci, probes, and primers to utilize, the best quality indicators and controls, and the performance parameters such as limits of detection, sensitivity, and specificity. These performance targets are key in designing laboratory assays that meet the needs for patient care.

Additionally, more high-quality reference material would be useful to laboratories in demonstrating the quality of their assays. These materials would be valuable during the validation studies as labs seek to demonstrate performance of either lab-developed or commercial assays. A pilot project is currently underway in the United States, called the Diagnostic Quality Assurance Pilot, which aims to design and develop new reference sample materials [37, 38]. While the College of American Pathologists and UK NEQAS both provide and evaluate proficiency testing materials, additional sources of standardized material could also be useful for expansion of proficiency testing programs and are needed for laboratories to demonstrate ongoing quality results and practices.

Conclusion

Genomic analysis has become a standard procedure in the pathologic assessment of tumors. While the methods and scope of analysis are still evolving along with technical and knowledge advances in the field, standardization of labo-

ratory and quality practices is well underway. Consensus guidelines for the clinical use of genomic information are rapidly being developed and include diagnostic, prognostic, and treatment purposes. In addition to involvement in development, utilization, and quality efforts, the laboratory community has responded by developing programs to train the next generation of pathologists and medical professionals to use these techniques.

References

1. Vogelstein B, et al. Cancer genome landscapes. Science. 2013;339(6127):1546–58.
2. The cancer genome atlas. Available at cancergenome.nih.gov.
3. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Weistler OD, Kleihues P, Ellison DW. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. 2016;131:803–20.
4. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127:2375–90. https://doi.org/10.1182/blood-2016-01-643569.
5. Wang HL, Lopategui J, Amin MB, et al. KRAS mutation testing in human cancers: the pathologist's role in the era of personalized medicine. Adv Anat Pathol. 2010;17:23–32.
6. Lièvre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, Côté JF, Tomasic G, Penna C, Ducreux M, Rougier P, Penault-Llorca F, Laurent-Puig P. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. Cancer Res. 2006;66(8):3992–5. https://doi.org/10.1158/0008-5472.CAN-06-0191. PMID16618717.
7. Di Fiore F, Le Pessot F, Lamy A, et al. *KRAS* mutation is highly predictive of cetuximab resistance in metastatic colorectal cancer. J Clin Oncol. 2007 ASCO Annual Meeting Proc. 2007;25(18S (June 20 Supplement)):10502.
8. Weichart W, Schewe C, Lehmann A, et al. KRAS genotyping of paraffin-embedded colorectal cancer tissue in routine diagnostics: comparison of methods and impact of histology. J Mol Diagn. 2010;12:35–42.
9. Kamel-Reid S, Zhang T, Persons DL, et al. (Molecular Oncology Resource Committee of the College of American Pathologists). Validation of KRAS testing for anti-EGFR therapeutic decisions for patients with metastatic colorectal carcinoma. Arch Pathol Lab Med. 2012;136:26–32. https://doi.org/10.5858/arpa.2011-0220-OA. http://www.ncbi.nlm.nih.gov/pubmed/22208484.
10. Allegra CJ, Jessup JM, Somerfield MR, et al. American Society of Clinical Oncology Provisional Clinical Opinion: testing for *KRAS* mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. J Clin Oncol. 2009;27:2091–5.
11. National Comprehensive Cancer Network Guidelines on Colon and Rectal Cancers. Practice guidelines established for *KRAS* mutation testing in colorectal cancer. NCCN; 2008.
12. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can testing of tumor tissue for mutations in EGFR pathway downstream effector genes in patients with metastatic colorectal cancer improve health outcomes by guiding decisions regarding anti-EGFR therapy? Genet Med. 2013;15(7):517–27. https://doi.org/10.1038/gim.2012.184.
13. Allegra CJ, Rumble RB, Hamilton SR, Roach N, Hantel A, Schilsky RL. Extended *RAS* gene mutation testing in metastatic colorectal carcinoma to predict response to anti–epidermal growth factor receptor monoclonal antibody therapy: ASCO provisional clinical opinion update 2015. J Oncol Pract. 2015; https://doi.org/10.1200/JCO.2015.63.9674.
14. Kaul KL, Sabatini LM, Tsongalis GJ, et al. The case for laboratory developed procedures: quality and positive impact on patient care. Acad Pathol. 2017;4:1–21.
15. Tang W, Hu Z, Muallem H, Gulley ML. Quality assurance of RNA expression profiling in clinical laboratories. J Mol Diagn. 2012;14:1–11. https://doi.org/10.1016/j.jmoldx.2011.09.003.
16. Mortensen MM, Høyer S, Lynnerup A-S, Ørntoft TF, Sørensen KD, Borre M, Dyrskjøt L. Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy. Nat Sci Rep. 2015;5:16018. https://doi.org/10.1038/srep16018.
17. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, Barry G, Dowidar N, Maysuria M, Storhoff J. Analytic validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin embedded breast tumor specimens. BMC Cancer. 2014;14:177. http://www.biomedcentral.com/1471-2407/14/177
18. Peck AR, Girondo MA, Liu C, et al. Validation of tumor protein marker quantification by two independent automated immunofluorescence image analysis platforms. Mod Pathol. 2016;29:1143–54. https://doi.org/10.1038/modpathol.2016.112; published online 17 June 2016.
19. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J-H, Bantscheff M, Gerstmair A, Faerber F, Kuster B. Mass-spectrometry-based draft of the human proteome. Nature. 2014;509:582–7. https://doi.org/10.1038/nature13319.
20. Mycancergenome.org.
21. Quail MA, Smith M, Copeland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific BioScidences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341–54.
22. Corless CL. Next-generation sequencing in cancer diagnostics. J Mol Diagn. 2016;18:813–6.
23. Misura M, Zhang T, Sukhai MA, Thomas M, Garg S, Kamel-Reid S, Stockley TL. Comparison of next generation sequencing panels and platforms for detection and verification of somatic tumor variants for clinical diagnostics. J Mol Diagn. 2016;18:842–50.
24. Kaul KL. Preparing pathology for precision medicine: challenges and opportunities. Virchows Arch. 2017;471(2):141. https://doi.org/10.1007/s00428-017-2141-z.
25. Gargis AG, Kalman L, Bick DP, da Silva C, Dimmock DP, Funke BH, Gowrisankar S, Hegde MR, Kulkarni S, Mason CE, Nagarajan R, Voelkerding KV, Worthey EA, Aziz N, Barnes J, Bennett SF, Bisht H, Church DM, Dimitrova Z, Gargis SR, Hafez N, Hambuch T, Hyland FCL, Luna RA, MacCannell D. Good laboratory practice for clinical next-generation sequencing informatics pipelines. Nat Biotechnol. 2015;33:689–93. https://doi.org/10.1038/nbt.3237.
26. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors. Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013;15:415–53.
27. NCCN clinical practice guidelines in oncology: non-small cell lung cancer v. 3.2017. www.nccn.org.
28. Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. Nature. 2013;12:358–69.

29. Plunkett-Rondeau J, Hyland K, Dasgupta S. Training future physicians in the era of genomic medicine: trends in undergraduate medical genetics education. Genetics. 2015;17:927–34. https://doi.org/10.1038/gim.2014.208.
30. Training residents in genomics. 2012. www.pathologylearning.org/trig/resources.
31. Sireci AN, Aggarwal VS, Turk AT, et al. Clinical genomic profiling of a diverse array of oncology specimens at a large academic center: identification of targetable variants and experience with reimbursement. J Mol Diagn. 2017;19:277–87. https://doi.org/10.1016/j.jmoldx.2016.10.008.
32. Lih C-J, Harrington RD, Sims DJ, et al. Analytical validation of the next-generation sequencing assay for a nationwide signal-finding clinical trial: molecular analysis for therapy choice clinical trial. J Mol Diagn. 2017;19:313–27. https://doi.org/10.1016/j.jmoldx.2016.10.007.
33. https://www.ukneqas-molgen.org.uk/introduction.
34. Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feensgra I, Race V, Sistermans E, Sturm M, Weiss M, Yntem H, Bakker E, Scheffer H, Bauer P. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet. 2016;24:2–5.
35. Deans Z, Watson CM, Charlton R, et al. Practice guidelines for targeted next generation sequencing analysis and interpretation. http://www.acgs.uk.com/media/983872/bpg_for_targeted_next_generation_sequencing_-_approved_dec_2015.pdf.
36. Joseph L, Cankovic M, Caughron S, et al. The spectrum of clinical utilities in molecular pathology testing procedures for inherited conditions and cancer. J Mol Pathol. 2016;5:605–19.
37. Zehnbauer B, Lofton-Day C, Pfeifer J, Shaughnessy E, Goh L. Diagnostic quality assurance pilot: a model to demonstrate comparative laboratory test performance with an oncology companion device assay. J Mol Diagn. 2017;19:1–3. https://doi.org/10.1016/j.jmoldx.2016.10.001.
38. http://www.tapestrynetworks.com/initiatives/healthcare/oncology-therapeutics-and-diagnostics/diagnostic-quality-assurance-pilot.cfm.
39. Perkel JM, Fung PA. Next-gen sequencing 2016 update, www.biocompare.com.

Current Massively Parallel Sequencing Technologies: Platforms and Reporting Considerations

2

John R. ten Bosch and Wayne W. Grody

Introduction

The Human Genome Project was officially completed in 2003 with the publication of the (near) complete sequence of 3.3 billion nucleotides in the haploid genome. Launched in 1990, the Project took 13 years and a budget of about $3 billion to sequence the first human genome. Well, at least that was the timeframe of the publicly funded Human Genome Project. A later entrant, the so-called "private" genome project pursued by the biotechnology company Celera, actually accomplished the same goal in a much shorter time, namely, in about 5 years. This was accomplished by a contrasting sequencing strategy; instead of the organized, targeted, chromosome-by-chromosome, BAC-by-BAC (bacterial artificial chromosome) approach, the Celera group utilized "shotgun" sequencing, a more global sequencing of countless random DNA fragments which were only reassembled at the end into the complete, ordered human genome.

In a similar way, DNA sequencing as performed by both diagnostic and research laboratories all over the world has recently undergone a dramatic transformation in speed and throughput. Instead of the traditional approach of sequencing one small (several hundred bp) DNA region at a time, using a specific pair of complementary primers targeted to just that area of interest, the new sequencing platforms utilize a shotgun approach, randomly shearing the entire genome into over 300 million small fragments, sequencing each of them repeatedly in parallel, and then reconstructing the resulting sequences, using sophisticated computer software, into the complete genome. Just as the entry of the Celera project spurred the total genome sequencing effort to an earlier completion, the advent of this "next-generation" or "massively parallel" DNA sequencing (NGS) technology has truly been a "game changer", allowing for practical and timely sequencing of large panels of genes, of all the coding regions of the genome (the exome), or of the whole genome itself in individual research subjects, patients, or nonhuman samples. And because the approach requires no preexisting knowledge of the target regions (only random/universal primers are used), the technology has opened the way to much new gene discovery and new organism identification (e.g., the microbiome). For the clinical molecular diagnostic laboratory, it has fueled a transition from traditional single-gene testing to a new world of genome-wide sequence analysis in the clinical setting [1].

Platform Chemistry

Early commercial adopters of NGS technology were 454 Life Sciences (since acquired by Roche) and Solexa (since acquired by Illumina). The 454 platform was a pyrosequencing-based system that produced long sequencing reads up to 1 kb in length [2]. As the first commercially available NGS instrument, 454 was a marked improvement to traditional sequencing methods. The reads were long and were able to be sequenced in parallel, the hallmark of all NGS technologies. However, as other NGS competitors such as Illumina and the Ion Torrent (since acquired by ThermoFisher) entered the market, 454 struggled to keep pace with the advances in sequencing throughput that led to a sharp decline in sequencing costs over several years (Fig. 2.1). Nevertheless, 454 remained a viable alternative in the niche market of groups requiring extra-long reads for the sequencing of complex genomic regions such as the HLA genes [3]. Eventually, Illumina and Ion Torrent improved their technologies so that they too could produce long reads. In 2013, Roche shut down operations of the 454 NGS instrument.

J. R. t. Bosch
TPMG Regional Genetics Laboratory, Department of Genetics, Kaiser Permanente Medical Center, San Jose, CA, USA

W. W. Grody (✉)
Departments of Pathology and Laboratory Medicine, Pediatrics, and Human Genetics, UCLA School of Medicine, Molecular Diagnostic Laboratories and Clinical Genomics Center, UCLA Medical Center, Los Angeles, CA, USA
e-mail: WGrody@mednet.ucla.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_2

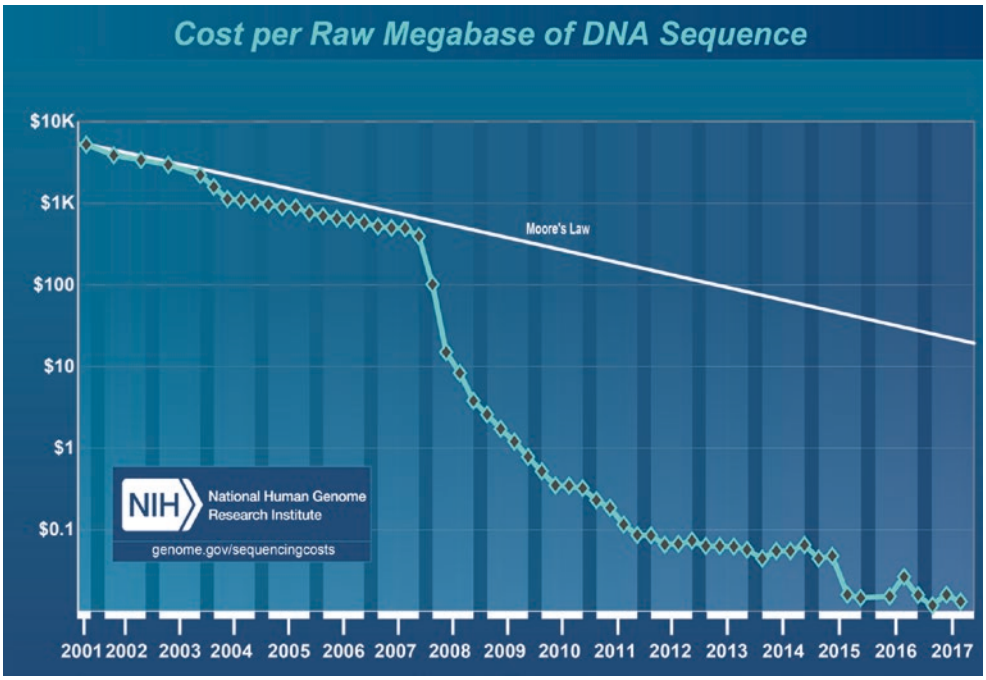


Fig. 2.1 Cost of DNA sequencing over time. The cost of a raw megabase of DNA sequence over time is compared to Moore's law. Wikipedia defines Moore's Law as the "observation that over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years." Since the advent of next-generation sequencing, the cost of sequencing has outpaced Moore's law by a wide margin. (Data and chart provided by the National Human Genome Research Institute (NHGRI) at https://www.genome.gov/sequencingcostsdata/ courtesy of Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata)

The two NGS platforms most common in diagnostic laboratories today are Illumina and Ion Torrent. Illumina technology is based on "sequencing-by-synthesis" chemistry, which is enabled by nucleotides containing reversible dye-terminators in the reaction mixture [4]. The dye terminators halt the extension of growing fragment chains at each of the hundreds of millions of fragment colonies on the flow cell. The incorporated, dye-conjugated nucleotides are detected via laser excitation and image capture with a high-resolution CCD (charge-coupled device) camera. The dye terminators are chemically removed before the next sequencing cycle, which begins with the addition of fresh, reversible dye-terminator nucleotides to the flow cell.

Detection of the fluorescence emitted during each sequencing cycle is facilitated by the creation of clonal fragment clusters prior to sequencing. These clonal clusters emit amplified fluorescent signals capable of detection by the instrument's CCD camera. The clusters are formed by first immobilizing a library of fragments onto the flow cell. The fragments are added to the flow cell at a dilute concentration so that, once formed, the clusters seldom overlap. The surface of the flow cell is coated with primers that enable bridge amplification of the immobilized fragments and the formation of the clonal clusters in the presence of polymerase chain reaction (PCR) reagents. Formation of the clonal clusters is the final step before sequencing.

Ion Torrent developed a novel NGS method based on the change in pH that results from the release of a hydrogen ion during nucleotide incorporation [5]. Ion Torrent chips contain millions of microwells with ion-sensitive field-effect transistor (ISFET) sensors. Beads containing emulsion PCR-amplified clonal DNA fragments are deposited onto the chips and loaded into the sensor-containing microwells via centrifugation. Each microwell accommodates a single bead. Unmodified nucleotides are then added to the chip in a stepwise fashion, and the ISFET sensor detects the pH change that results from the addition of a nucleotide to the growing DNA chain. If more than one nucleotide is incorporated, the change in ion concentration should be proportional. Long homopolymer stretches, however, may not adhere to this rule.

The Ion Torrent instrument completes sequencing runs in a relatively short period of time by industry standards, in part because it does not rely on optics. Because Ion Torrent sequencing chips incorporate the same technology used in the semiconductor industry, any advancements in complementary metal-oxide-semiconductor (CMOS) technology will likely improve the ion chips as well.

As with the Illumina platform, sensitivity in the detection of incorporated nucleotides is facilitated by the amplification of template fragments. The Ion Torrent method uses emulsion PCR to sequester clonal fragments onto beads that are then deposited into individual microwells on the Ion chip.

Prior to emulsion PCR, beads are conjugated with unique DNA fragments. The beads are then added to a water-oil mixture containing emulsion droplets infused with PCR reagents. The droplets envelop and isolate individual beads, thereby enabling the clonal amplification of the attached DNA fragments.

Rapid Sequencing Revolution

Once NGS technology was commercialized, subsequent iterations of the original technologies were designed to lower the cost per base of sequencing by generating longer reads and increasing read capacity (Fig. 2.1). Many in the field thought so-called "third-generation" sequencing technologies, such as single-molecule NGS or nanopore sequencing, would be the next wave of innovations to take hold in the laboratory. Although these third-generation technologies, such as single molecular and nanopore-based sequencing, still hold much promise (see the next chapter in this book), improvements in existing NGS run times have made the greatest impact in laboratories that apply this technology. This is especially true in clinical laboratories, where short turnaround time is often paramount [6].

Benchtop sequencers, which derive their name from relatively small laboratory footprints, were the first NGS instruments to offer significant advancements in sequencing run times [7, 8]. These fast run times are, in part, a result of their smaller sequencing capacity, though instrument and chemistry enhancements certainly contributed as well. Benchtop sequencers were, and continue to be, much more affordable and easier to maintain than their large-capacity predecessors and counterparts. This aspect is why benchtop sequencers remain extremely popular in clinical laboratories with fixed-gene NGS panels. Indeed, the first FDA approval for an in vitro diagnostic application using NGS was awarded to Illumina's benchtop instrument (MiSeq) for mutation detection in the *CFTR* gene (cystic fibrosis) [9]. In addition to their affordability, benchtop sequencers typically produce adequate sequence coverage of most gene panels, despite multiplexing, with the added benefit of a shorter turnaround time.

The two most successful NGS benchtop instruments have been the Illumina MiSeq and the Ion Torrent Personal Genome Machine (PGM). The MiSeq instrument leverages the same sequencing-by-synthesis chemistry used by the large-capacity Illumina sequencers, but with shorter run times. Its rapid sequencing runs are the result of a smaller sequencing capacity as well as enhanced fluidics and automated, onboard cluster generation. The popularity of the MiSeq platform prompted Illumina to transfer several of its features to its larger-capacity instruments. These improvements confer flexibility to these instruments, allowing the user to choose between a rapid-run and high-output mode. However, even in the "low-throughput" rapid-run mode, the HiSeq lives up to its name by producing enough sequence to cover an entire human genome at ample coverage. In addition, Illumina now offers benchtop sequencing alternatives to the MiSeq instrument that are both higher (NextSeq) and lower capacity (MiniSeq) alternatives to their predecessor.

The first commercial instrument sold by Ion Torrent was the PGM, a benchtop sequencer. The PGM is scalable and can accept one of the three different capacity ion sequencing chips, the largest of which produces up to 2 Gb of sequence with 400 bp reads. On-instrument sequencing with the PGM is exceptionally fast because the PGM has no moving parts and no optics, both of which have been rate limiting on other sequencing platforms. Ion Torrent later released its larger-capacity Proton instrument, which can generate up to 15 Gb of sequence in a single day. More recently, the Ion Torrent S5 instrument was launched with variable chip sizes that can replace both the lower-throughput PGM and its larger cousin the Proton.

Enrichment Techniques

The adoption of NGS technology in diagnostics would not have occurred so rapidly without the introduction of several easy and efficient techniques for isolating regions of the genome. Sequencing even a small gene panel is extremely labor-intensive with traditional PCR techniques, and sequencing larger panels is virtually impossible given the time constraints of most clinical tests. Over the past few years, several genomic enrichment techniques have been developed to overcome this bottleneck in targeted next-generation sequencing. Most of these techniques fall under three general categories of genomic enrichment: bait hybridization, highly multiplexed PCR, and microfluidic technologies.

Bait hybridization with microarrays was one of the first enrichment techniques developed [10–12]. This method quickly became one of the most popular when a solution-based approach using biotinylated bait probes was developed [13]. Initially offered by Agilent technologies, these solution-based hybridization techniques use long nucleic acids designed to minimize cross-hybridization of undesired sequence as the bait probes. Briefly, genomic DNA is subjected to NGS library preparation. These adapter-ligated DNA fragments are then hybridized to a pool of bait probes complementary to genomic regions of interest. Once hybridized, the target sequences are enriched using streptavidin-coupled magnetic beads and bead washing. The enriched DNA fragments are then amplified in preparation for sequencing.

Traditional PCR has modest multiplexing capability, much less than is required for even the smallest of NGS gene panels. Several different methods have been developed to overcome this limitation of traditional PCR. Circularization of library

fragments is one technique that can be used to isolate and enrich for thousands of targets in multiplex [14]. Commercialized by Agilent Technologies, this technique, deemed the HaloPlex system, is now able to isolate the tens of thousands of targets necessary for an entire human exome. The Illumina TruSeq Amplicon enrichment system uses a technology derived from their SNP GoldenGate genotyping assay to amplify up to 1536 targets in a single reaction. The TruSeq system, like Agilent HaloPlex, isolates regions of interest using oligonucleotide probes with universal priming sites and target-specific ends that facilitate amplification. However, the TruSeq Amplicon system bypasses the need for template circularization by using probes that flank each target. Ion Torrent similarly developed a technique to overcome the barriers of traditional multiplex PCR. Called Ion AmpliSeq, this technology requires only minimal starting DNA material to enrich DNA from 12 to 24,000 targets for sequencing on the Ion Torrent PGM platform. It has been especially useful for targeted mutation detection in tumor samples.

Microfluidic platforms such as RainDance and Fluidigm take advantage of proprietary instruments that compartmentalize PCR templates and reagents into thousands of PCR mini-reactors. Despite somewhat similar concepts, the RainDance platform delivers a large set of target sequences from a single sample, while the Fluidigm platform isolates a smaller number of targets from multiple samples simultaneously.

Different Tests, Different Outcomes

Choosing an NGS platform can be a difficult task. As a result of rapid NGS innovations, laboratories are investing not only in the current capabilities of a platform but also count on future improvements that will allow that platform to keep pace with the rest of the field and not necessitate purchasing another expensive instrument in a short time. Nevertheless, the half-life of even the most successful NGS instruments, much like the computer infrastructure that supports them, tends to be fairly short. Indeed, starting an NGS laboratory should be considered, not a one-time investment but an ongoing obligation. For this reason, some institutions have chosen to concentrate laboratory resources on NGS analysis and interpretation and to outsource the actual sequencing to another clinical testing laboratory. As throughput, accuracy, and price continue to improve, there has been a tendency in the field to begin thinking of NGS as a sort of generic commodity that can be readily outsourced to other vendors while retaining the interpretive component – which is the true locus of interpretive expertise – in-house. Of course, any outsourced sequencing for clinical use must be performed in a CLIA-certified facility.

The principal criterion to consider when deciding on an NGS platform is the purpose for which it will be used. For example, the needs of a laboratory that considers NGS for a carrier screening assay will be much different than the needs of a laboratory that performs WGS – or a laboratory that leverages deep-coverage NGS to identify somatic mutations in a small subset of cancer genes or drug-resistant mutations in subclones of bacterial or viral microorganisms. How many targets must be sequenced per assay and at what read depth? Which kinds of mutations must be detected? What is the expected sample volume and turnaround time for the test? How much DNA will be available for sequencing? These questions highlight some of the most critical parameters to consider before purchasing an NGS instrument.

The requirements of a laboratory performing WGS and a laboratory performing whole-exome sequencing (WES) could be much different given that the protein-coding portions represent only ~1.5% of the genome. However, the throughput requirements of the two laboratories might be quite similar if the WES laboratory used indexed DNA barcodes to combine multiple samples in each run. In such a scenario, both laboratories would require full-capacity runs of a high-throughput instrument to obtain the necessary amount of sequence in a single instrument run.

WES is an assay that targets all of the approximately 23,000 protein-coding genes in the genome. Diagnostic WES tends to be favored when the phenotype of a patient does not suggest a particular disorder or group of genes. This is in contrast to targeted panels comprising genes that, when mutated, contribute to a common or related set of syndromes (e.g., hearing loss, cardiac abnormalities). Fixed-gene panels cost less and typically guarantee minimum sequence coverage for all of the genes in the panel. They have further advantages of often being more likely reimbursed by health insurers and of avoiding “incidental findings” (see below). However, because a limited number of genes are sequenced in these panels and it is challenging for them to incorporate new disease gene discoveries in a timely fashion, WES may still be indicated in event of a negative result.

WES targets more genes than panel assays, but its sensitivity may be lower if certain genes are not captured or sequenced uniformly. This has become less of a problem as overall gene coverage rates in WES have continued to improve, currently approaching 99–100% across most genes so that only a small minority of genes or exons tend to be missed or exhibit suboptimal coverage. These low-coverage regions usually have a skewed GC percentage or other sequence-related issues, such as regions of homology or associated pseudogenes that make enrichment difficult, so coverage problems are often predictable [15].

WGS targets the entire genome, but at increased overall cost (though the differential in expense continues to diminish). WGS sample preparation is much easier than the preparation required for WES or even panel sequencing, given that no enrichment step is required. In addition, sequencing data are obtained from intronic and intergenic regions, and it is

becoming increasingly clear that much intergenic sequence has a biochemical or regulatory function and is not, in fact, "junk DNA" [16–19]. In addition, large copy number (insertion and deletion) analysis can be performed more readily from WGS than WES, although chromosomal microarray remains the gold standard for this type of analysis.

Mutation detection using NGS is even more complex with cancer samples [20, 21]. This is, in part, due to the fact that tumors are heterogeneous, both at the cellular level with tissue being a mixture of tumor and normal cells and at the genetic level with different populations of cells harboring different combinations of mutations. In addition, many tumor types have a high rate of genomic rearrangement. Within chromosomal regions of increased ploidy, nucleotide-level variants are diluted even further. Each of these characteristics of tumor samples makes it difficult to fully characterize this somatic mosaicism and assure identification of the mutations that helped drive cancer transformation (as opposed to less relevant secondary or "passenger" mutations). Furthermore, it may be necessary to sequence normal patient tissue in order to distinguish somatic from inherited variants in the patient sample of interest. Finally, tumor tissue is often scant, and with certain biopsies, there may not be sufficient tissue available for molecular analysis. These challenges have limited most clinical cancer NGS to relatively limited panels of genes that are sequenced to a very high depth [22, 23]. However, continued advancements in technique and software algorithms are enabling ever larger gene panels and even the detection of gene fusions/translocations by NGS [24].

Analysis

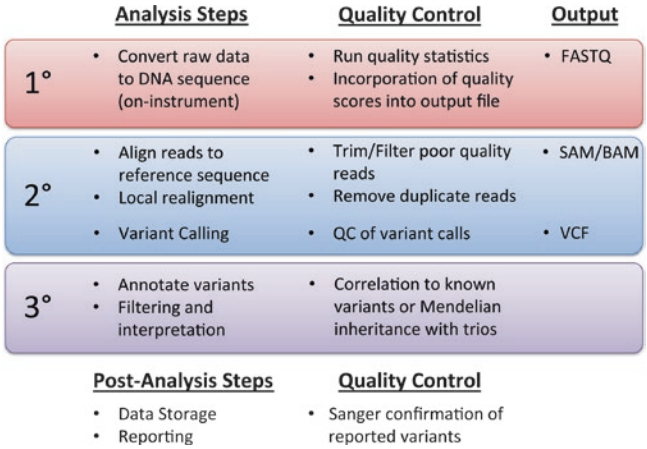
Over the last few years, the College of American Pathologists (CAP) has been updating its Laboratory Accreditation Checklist for Molecular Pathology to include more content specific to both the wet-bench and bioinformatic facets of NGS. While both are equally important, the standards are addressed separately, in part to account for those laboratories who outsource either the sequencing or interpretive component [25]. Included in the analytical wet-bench procedure are library preparation, enrichment, indexing of pooled samples, and the sequencing process itself. Confirmatory testing of reported NGS findings is also included in this section. Bioinformatics includes the pipeline used to support the analysis, interpretation, and reporting of NGS results (Fig. 2.2a). Bioinformatics therefore includes the algorithms used to analyze the results as well as the scripts used to tie together the analysis steps. It also includes any in-house databases used to interpret and store identified variants.

NGS analysis consists of three discrete processes described as primary, secondary, and tertiary stages (see Chap. 15, "Next-Generation Sequencing for Single-Gene Analysis," for detailed discussion). In brief, these stages include the conversion of raw NGS data to DNA sequence, the mapping and alignment that identifies sequence variants, and the annotation and filtering of variants [26]. Each stage of analysis provides an opportunity for the integration of quality control (QC) measures to avoid potential false calls. NGS bioinformatic pipelines incorporate these QC elements into an automated workflow that ties together the distinct steps of sequence analysis. Most laboratory-developed or in-house pipelines use external analysis tools to process the data and internal scripting to facilitate the movement of data through pipeline and file format optimization [27]. Commercial NGS analysis solutions are often composed of proprietary analysis tools that could possibly be strung together to provide an integrated workflow.

Primary NGS data analysis is the process of converting raw data (e.g., images or sensor data) to DNA sequence. Such analysis often occurs on the instrument using vendor-provided software. The most common file format output of primary analysis is the FASTQ file [28]. The FASTQ file format is a variant of the well-established FASTA format. However, FASTQ files contain both sequence and individual base quality scores. Inclusion of the quality scores allows the trimming and/or removal of poor-quality reads prior to mapping. Trimming may be beneficial if a significant decrease in base quality is observed toward the end of a read.

Secondary analysis consists of quality assurance (QA) filtering of raw reads, alignment of reads, and variant calling. NGS platforms operate in a "shotgun" manner, meaning that sequencing reads are obtained by a random sampling of the genomic DNA. For human re-sequencing, individual reads must be mapped to a reference genome in order to determine the locations from which they originated. To permit the mapping of sequence reads that harbor variants, alignment parameters should allow for slight deviations from the reference sequence. Unfortunately, more complex variants, such as indels, are among the most difficult to map precisely because each conflicting nucleotide makes it less likely the read will be recognized as a derivation of the reference sequence [29]. Longer reads help counter this problem by providing additional sequence for comparison to the reference. Longer reads are also useful for proper alignment to repetitive regions of the genome (such as short tandem repeats and trinucleotide repeat expansions) and for accurate phasing (i.e., determining whether two variants in the same gene lie in *cis* or *trans*). Reads from pseudogenes are particularly problematic because they can mimic variant-containing reads from the functional gene relative. Another strategy to help avoid misalignment is to employ a local realignment method to fine-tune the results from the initial mapping. When using two rounds of alignment, mapping parameters can be loosened during the first step so that more computationally intensive algorithms can be focused on the reads that

a



b

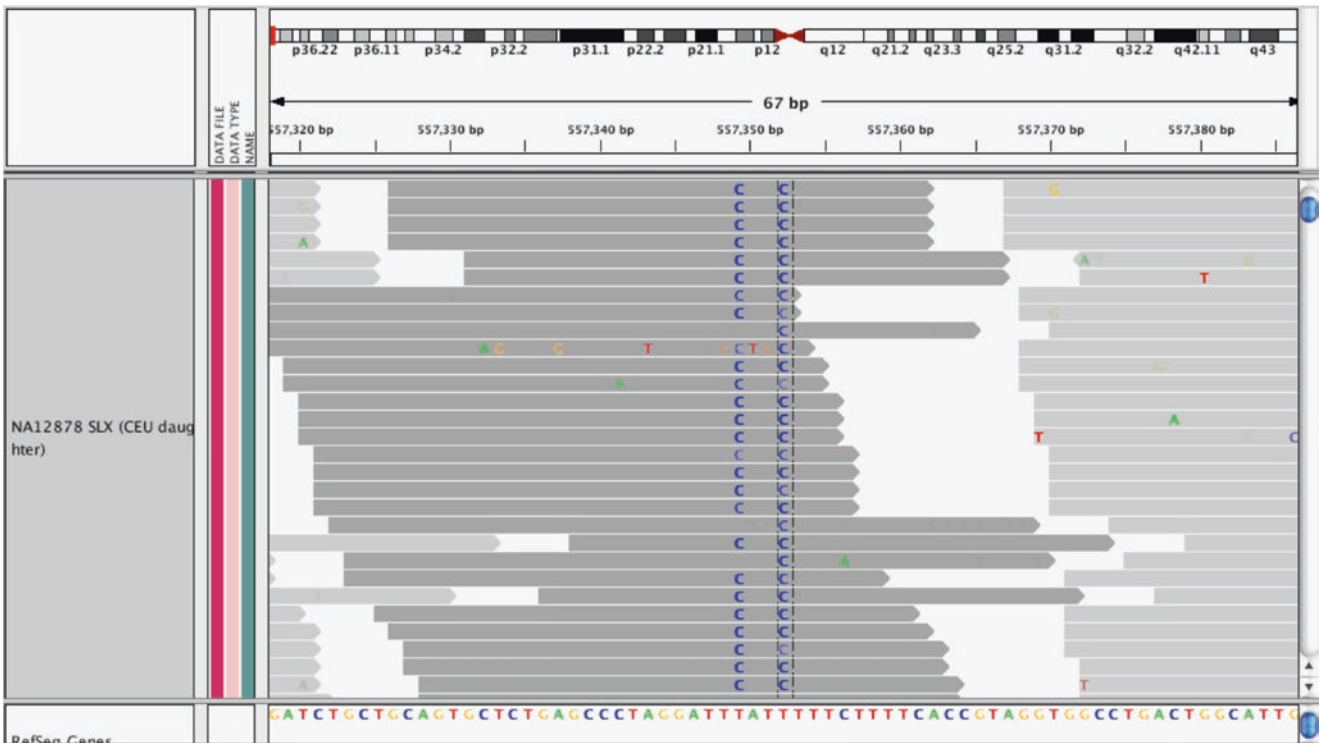


Fig. 2.2 Next-generation sequencing analysis. (**a**) Sample NGS analysis pipeline and common post-analysis steps. Analysis steps are split into primary, secondary, and tertiary stages of analysis. (**b**) Visualization of secondary alignment data using the Integrated Genomics Viewer (IGV). IGV is an open-source data visualization viewer created by the Broad Institute. (**c**) Sample tertiary analysis workflow. In this example, the correlation of a phenotype with that of the patient can be performed either at the beginning of the analysis or closer to the end. Where this filter is applied can have profound implications on the variants seen and ultimately reported

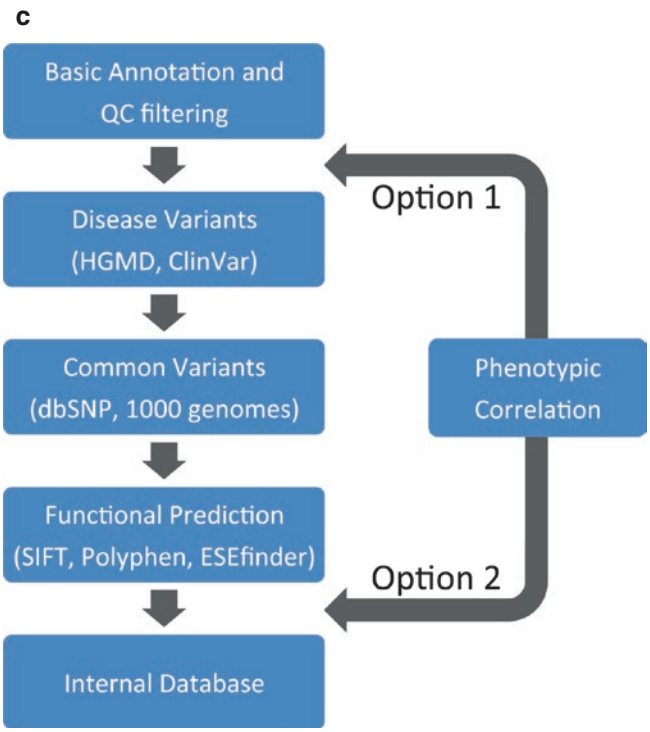


Fig. 2.2 (continued)

require the most attention, i.e., those with mismatched nucleotides.

There are several alignment algorithms for human re-sequencing, many of which are freely available as open-source software [30]. Each of these mapping algorithms varies in terms of speed, memory requirement, and sequencing platforms supported. The standard output file format for sequence alignment data is the Sequence Alignment/Map (SAM) file and its compressed binary twin, the BAM file [31]. Conversions between these file formats, as well as a variety of other manipulations, can be performed by the open-source SAMtools or other similar utilities. In addition, alignment files can be uploaded into applications, such as the open-source Integrated Genomics Viewer (IGV), for the visualization of aligned reads and sequence coverage of NGS data [32] (Fig. 2.2b).

Variant calling is the process of identifying the differences between the aligned reads and the reference sequence. The universal format for the variant file is the Variant Call Format (VCF) file, a tab-delimited text file that includes reference and variant nucleotides, chromosome positions, unique variant identifiers, quality scores, and any number of other potential fields the user wishes to add [33]. Tertiary analysis includes annotation of the variant file and the filtering that follows. Populating the supplementary fields of the VCF file typically involves extracting information from databases, both internal and external, to help elucidate the significance of each variant. These databases include dbSNP and the 1000 Genomes Project [34, 35], both of which are used to identify common single nucleotide polymorphisms (SNPs). These SNPs are typically benign variants found in asymptomatic individuals. The Human Gene Mutation Database (HGMD) and the public archive, ClinVar, are examples of curated databases that gather information about the clinical significance of variants both benign and pathogenic [36, 37]. Inclusion of phenotypic information associated with a particular gene or variant has proven to be an effective tool for exome and genome sequencing laboratories that deal with thousands and sometimes millions of variants per sample. Another common annotation practice is the inclusion of values from online software prediction tools that assign variant scores based on the projected pathogenicity of variants, taking into consideration the nature of the DNA/amino acid change, evolutionary conservation of the sequence at that position, tolerance of the gene to that type of alteration, etc. [38, 39]. Unfortunately, experience to date with these algorithms is that they are of limited predictive value and often reach contradictory conclusions among one another.

Interpretation and Reporting

Whether or not a conclusive, causative mutation is identified when a genomic assay is performed on a patient, there will always be a large number of variants of uncertain pathogenicity revealed. This is not new in genetic testing, as it has been a feature of both routine chromosome analysis and genomic microarray assays. Because exome and genome sequencing are such high-resolution assays capable of detecting an extremely large number of variants, the rates of such variants of uncertain clinical significance (VUS) with these tests are the highest yet observed. Indeed, a typical whole-exome sequence produces >20,000 variants that differ from the reference sequence, while a whole-genome sequence yields >3 million. This phenomenon not only makes genome-wide test interpretations many orders of magnitude more complex than single-gene or gene-panel tests but also introduces an ethical conundrum as to what level of clinical certainty should be required for any particular variant to be reported out to the physician and patient.

Prior to the advent of genome-based diagnostics, the interpretation of molecular assays was, for the most part, performed manually and without the use of software to filter out variants based on a set of assumptions. Alternatively, the testing community may agree, a priori, to the restriction of reportable mutations from a larger set, as is practiced among laboratories that perform cystic fibrosis carrier screening. These targeted mutation panels are useful in that they identify the most common disease-causing mutations while avoiding VUS or known milder variants that would introduce problematic clinical decision-making, especially in the pre-

natal setting. The decision to include or exclude variants from a final list is made only after careful consideration of each candidate mutation [40]. The final mutation panel can then be "locked" so the same loci are accepted or rejected with each run of the assay.

In contrast, the automated filtering of NGS datasets, as is done in exome sequencing, occurs without any preconceived notion of the variants that will pass through the bioinformatic pipeline filters [33]. These filters are designed based on a set of assumptions that are put in place to deal with the thousands of variants obtained from NGS assays. The use of these filters is meant to ensure that the variants most likely to be disease-causing are prioritized. However, any mutations that do not fit the filter assumptions may be misclassified or discarded.

A sample variant annotation and filtering routine are shown in Fig. 2.2c. First, variants are annotated using information from a public genome database. The affected gene is recorded, as is the nature of the mutation, be it an amino acid change, splice site disruption, or one that affects transcript production or stability. Curated clinical-grade disease databases such as HGMD and ClinVar are then queried to determine if any variants have been previously identified as pathogenic. Variants are subsequently checked against dbSNP and 1000 Genome data and now more commonly the large ExAC and gnomAD sequence databases of supposedly healthy individuals [41] (along with any suitably large internal set of "control" genomes/exomes the laboratory may have accrued), to determine allele frequency and whether the observed changes have been identified or may be inferred (based on their background frequency) to be benign variants. The remaining variants are then subjected to algorithms that predict whether the variant is a putative mutation based on an amino acid change or splice site interruption. Finally, variants are checked against an internal mutation database to determine if the laboratory has previously observed them and, if so, how they were characterized. Some laboratories may also choose to sequence parental samples, or samples from other relatives, or companion benign tissue from the cancer patient, to aid in the interpretation of variants found in the proband. Such comparison samples can be of tremendous value in determining clinical significance of novel variants, as well as whether they are inherited or de novo (which can impact recurrence risk) [42].

Pipeline filters that incorporate the phenotypic information associated with the variants detected in these assays will dictate the rate of VUS and incidental variants ultimately reported, as well as the sensitivity of the test. As is shown in Fig. 2.2c, these phenotypic filters are applied either at the beginning of the filtering process or toward the end. When phenotypic filters are applied only after disease-causing mutations are identified, some laboratories will feel obligated to report them even if they are indeed unrelated to the test indications (see "Dealing with Incidental Findings," below). In contrast, when genes are preselected using keywords related to the patient's phenotype, all variants that occur outside of this virtual gene panel are removed from consideration prior to analysis. In such a scenario, incidental findings would likely not be found, but neither would significant mutations that occur in genes yet to be correlated with the disease phenotype – so each of these decisions represents a trade-off of sorts. A somewhat hybrid strategy is to perform multiple rounds of interpretation, increasing the size of the virtual gene panel each time, until a significant variant is identified or the entire exome is unmasked.

Dealing with Incidental Findings

It is inevitable that a laboratory performing clinical WES or WGS is going to be confronted from time to time with incidental (also sometimes referred to as secondary, off-target, or unexpected) findings. At the inception of clinical genomic sequencing, there was intense debate in the genetics community surrounding how such findings should be handled, a debate which is likely to continue for some time [43, 44]. The classic example, often used as a basis for discussion, is the finding of a clearly pathogenic mutation in a gene associated with one of the adult-onset, dominant cancer syndromes (such as familial breast/ovarian cancer or Lynch syndrome) in a young child or baby undergoing genome-wide sequencing for an unrelated condition such as congenital deafness, seizure disorder, or autism. The medical genetic community has long adhered to an ethical policy of not performing predictive/presymptomatic genetic testing for adult-onset disorders in children, unless there is some medical or surgical intervention that would need to be introduced in childhood in order to prevent or minimize the condition. Since *BRCA*-associated breast and ovarian cancers, for example, do not occur in childhood, nor would a baby girl ever be a candidate for prophylactic mastectomy or oophorectomy, testing for *BRCA* mutations would never be sanctioned in such a young patient. But what happens if the laboratory happens to stumble upon one incidentally during sequencing for one of the unrelated indications like congenital hearing loss? Now that it has been seen, should it be reported? Could nondisclosure, per the existing ethical policy, eventually cause harm to the child or her mother (if the mutation were in fact passed down from her)? Could a compromise solution be designed whereby the incidental finding is "flagged" in the electronic medical record to reappear and be reported out when the girl reaches age 18? Or should patients or parents be offered a multitiered consent form prior to testing, in which they get to select which types, if any, of incidental findings they wish or do not wish to receive?

After intensive deliberation, a guideline on this topic was issued by the American College of Medical Genetics and

Genomics (ACMG) [45]. The guideline as written fell quite clearly within the "duty to warn" camp. The major recommendations were:

- Documented mutations in a select list of high-penetrance, potentially lethal, but actionable conditions *must be sought and reported*.
- The same rules apply to sequencing of healthy parents in a "trio" or benign companion tissue when doing tumor sequencing.
- No distinction is made between adult and pediatric patients.
- These results are given to the ordering clinician who has responsibility for deciding which, when, and how to convey to the patient.
- The *patient cannot opt out* from receiving these incidental findings.

The target list of genes consists of high-penetrance familial cancer syndromes, cardiomyopathies, malignant hyperthermia, and other conditions meeting the criteria. Generating the most subsequent controversy has been the proviso that these targets must be actively sought out (as opposed to incidentally "stumbling" upon them) and the elimination of an "opt-out" choice for patients. One of the rationales used to justify these conclusions is the analogy with radiology: a radiologist is obligated to report all abnormal findings seen in a chest X-ray, regardless of the indication or specific suspicion upon which the X-ray had been ordered. However, these points could also be interpreted as infringing on both patient and laboratory autonomies.

Not surprisingly, feelings run strong on both sides of this debate [46, 47]. A subsequent update of the guideline [48] did allow for patient (or parent) "opt out" from receiving incidental findings if they so wished. In actual practice, it must be admitted that much of the concern surrounding this question has been rendered moot by the observation that the vast majority of families choose to receive incidental findings. Many studies are ongoing to ascertain the longer-term impact of this information on family dynamics and medical decision-making.

Gene Patents

The issue of intellectual property and restrictive gene patents has been one with which the molecular diagnostic community has had to contend almost since its inception. All of us have examples of molecular tests we have had to remove from our menus after receiving "cease-and-desist" letters from the exclusive gene patent-holder. Fortunately, this has now become something of a moot point, with the dramatic 2013 Supreme Court ruling in the *Association for Molecular Pathology et al. vs. Myriad Genetics* case, which was brought as a direct challenge to the restrictive intellectual property tied to the *BRCA1* and *BRCA2* genes, allowing only Myriad Genetics to offer the full-gene sequencing test. In brief, the Court ruled in June 2013 that genes represent "products of nature" and therefore cannot be patented, thus invalidating not only the *BRCA* patents but those for all other genes, as well [49]. Therefore, despite all the other challenges in implementing and reporting NGS tests, fear of gene patent infringement should no longer be a hindrance. Of note, Myriad's proprietary and very extensive database of BRCA sequence variants, accrued over 20 years of clinical testing, was not at issue in the Supreme Court case and therefore has not been made accessible to the public, making full interpretation of BRCA variants detected by NGS still somewhat challenging. To address this, there are ongoing efforts to reconstruct such a database from other sources [50].

Follow-Up, Reanalysis, and Duty to Recontact

Unlike the situation for long-established analytes in clinical chemistry or microbiology, our knowledge of the clinical implications of genetic and genomic variants is constantly changing as new discoveries are published and additional mutations and polymorphisms are deposited in DNA databases. Thus, the clinical interpretation of a particular nucleotide variant today may not be the same is it might be next year or even next month. This begs the question, long debated in the genetic and oncology communities, about whether or not there is a "duty to recontact" patients as our knowledge of previously tested targets changes and, if so, whether the responsibility for doing so should fall on the testing laboratory or on the ordering clinician. While some laboratories that focus largely on a particular gene or set of genes, such as Myriad Genetics (see above), keep extensive databases of the variants found in those genes and flag recurrent ones for recontact even years later if the interpretation has changed, consensus has emerged that such a requirement would place an untenable burden on the average clinical molecular diagnostic laboratory dealing with many different disorders on a daily basis [51]. Obviously, this challenge would be multiplied by many orders of magnitude when performing sequencing tests on all the genes in the genome, and no one laboratory or director can conceivably be held responsible for keeping up-to-the-minute on all the literature pertaining to 23,000 genes. Given that reality, and the concern that even an implied duty to recontact would entail an open-ended laboratory-physician-patient relationship that could not be met and could lead to future liability, current thought in the field is that the responsibility for monitoring developments that might result in a revised interpretation of genomic results

should fall to the patient and/or their physician, either of whom can, when appropriate, request reanalysis of archived exome or genome sequence data or submit a new specimen for re-sequencing [52].

Conclusions

An outsider viewing the current state of NGS for clinical purposes is likely to be intimidated by the many challenges and hurdles it presents: astronomically expensive instruments, an infinite and constantly changing knowledge base, potential errors in capture, sequencing and alignment, the huge numbers of VUS produced on every case, the need for multidisciplinary interpretations that could take hundreds of hours, uncertainties about costs and reimbursement, and so on. But just as we recently saw a solution to the gene patent problem, we can be certain that these remaining challenges will be met as well in the coming years. Based on trends thus far, we can be sure that genome-level DNA sequencing will continue to improve in accuracy, user-friendliness, speed, and cost-effectiveness. As lower cost continually expands the market for these tests, it is even likely that many of the ethical questions that we find so difficult to answer at present – such as the return of incidental findings – will sort themselves out. Indeed, there may come a time in the not-so-distant future when we may wonder why our predecessors ever agonized so much over these questions, when society might come to accept routine WGS of every newborn as no more controversial than the standard heel-stick for metabolic disease screening that we accept now. When or whether that ever comes to pass is at this point an open question. But there can be no question that NGS will assume an ever-increasing role in molecular diagnostic testing in the years to come, ultimately usurping or replacing Sanger sequencing and other traditional methods.

References

1. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J Mol Diagn. 2008;10:484–92.
2. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376–80.
3. Trachtenberg EA, Holcomb CL. Next-generation HLA sequencing using the 454 GS FLX system. Methods Mol Biol. 2013;1034:197–219.
4. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456:53–9.
5. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348–52.
6. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, Andraws N, Patterson ML, Krivohlavek LA, Fellis J, Humphray S, Saffrey P, Kingsbury Z, Weir JC, Betley J, Grocock RJ, Margulies EH, Farrow EG, Artman M, Safina NP, Petrikin JE, Hall KP, Kingsmore SF. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med. 2012;4(154):154ra135.
7. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434–9.
8. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4):294–6.
9. Bijwaard K, Dickey JS, Kelm K, Težak Ž. The first FDA marketing authorizations of next-generation sequencing technology and tests: challenges, solutions and impact for future assays. Expert Rev Mol Diagn. 2015;15:33–40.
10. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4:903–5.
11. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007;39:1522–7.
12. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. Nat Methods. 2007;4:907–9.
13. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27:182–9.
14. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J. Multiplex amplification of large sets of human exons. Nat Methods. 2007;4:931–6.
15. Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, Duffy E, Hegde M, Santani A, Lebo M, Funke B. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genet Med. 2016;18:1282–9.
16. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
17. Sheikine Y, Kuo FC, Lindeman NI. Clinical and technical aspects of genomic diagnostics for precision oncology. J Clin Oncol. 2017;35:929–33.
18. Harris G, O'Toole S, George P, Browett P, Print C. Massive parallel sequencing of solid tumours – challenges and opportunities for pathologists. Histopathology. 2017;70:123–33.
19. Kanagal-Shamanna R, Singh RR, Routbort MJ, Patel KP, Medeiros LJ, Luthra R. Principles of analytical validation of next-generation sequencing based mutational analysis for hematologic neoplasms in a CLIA-certified laboratory. Expert Rev Mol Diagn. 2016;16:461–72.
20. Gundry M, Vijg J. Direct mutation analysis by high-throughput sequencing: from germline to low-abundant, somatic variants. Mutat Res. 2012;729:1–15.
21. Ding L, Wendl MC, Koboldt DC, Mardis ER. Analysis of next-generation genomic data in cancer: accomplishments and challenges. Hum Mol Genet. 2010;19(R2):R188–96.
22. Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M, Corless CL. Combining highly multi-

plexed PCR with semiconductor-based sequencing for rapid cancer genotyping. J Mol Diagn. 2013;15:171–6.
23. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, Sanford T, Buddavarapu K, Krosting J, Garmire L, Wylie D, Shinde R, Beaudenon S, Alexander EK, Mambo E, Adai AT, Latham GJ. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. J Mol Diagn. 2013;15:234–47.
24. Vendrell JA, Taviaux S, Béganton B, Godreuil S, Audran P, Grand D, Clermont E, Serre I, Szablewski V, Coopman P, Mazières J, Costes V, Pujol JL, Brousset P, Rouquette I, Solassol J. Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. Sci Rep. 2017;7:12510.
25. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DG, Merker JD, Nagarajan R, Palicki LA, Robetorye RS, Schrijver I, Weck KE, Voelkerding KV. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med. 2015;139:481–93.
26. Altmann A, Weber P, Bader D, Preuss M, Binder EB, Müller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Hum Genet. 2012;131:1541–54.
27. Torri F, Dinov ID, Zamanyan A, Hobel S, Genco A, Petrosyan P, Clark AP, Liu Z, Eggert P, Pierce J, Knowles JA, Ames J, Kesselman C, Toga AW, Potkin SG, Vawter MP, Macciardi F. Next generation sequence analysis and computational genomics using graphical pipeline workflows. Genes (Basel). 2012;3:545–75.
28. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38:1767–71.
29. Ruffalo M, LaFramboise T, Koyutürk M. Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics. 2011;27:2790–6.
30. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11:473–83.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
32. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.
33. Stitziel NO, Kiezun A, Sunyaev S. Computational and statistical approaches to analyzing variants identified by exome sequencing. Genome Biol. 2011;12:227.
34. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
35. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D, Leinonen R, Shumway M, Sherry S, Flicek P, 1000 Genomes Project Consortium. The 1000 genomes project: data management and community access. Nat Methods. 2012;9:459–62.
36. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. The human gene mutation database: 2008 update. Genome Med. 2009;1:13.
37. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2017;46:D1062.
38. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
39. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073–81.
40. Grody WW, Cutting GR, Klinger KW, Richards CS, Watson MS, Desnick RJ. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. Genet Med. 2001;3:149–54.
41. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.
42. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014;312:1880–7.
43. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA. 2006;296:212–5.
44. Green RC, Berg JS, Berry GT, Biesecker LG, Dimmock DP, Evans JP, Grody WW, Hegde MR, Kalia S, Korf BR, Krantz I, McGuire AL, Miller DT, Murray MF, Nussbaum RL, Plon SE, Rehm HL, Jacob HJ. Exploring concordance and discordance for return of incidental findings from clinical sequencing. Genet Med. 2012;14(4):405–10.
45. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.
46. Wolf SM, Annas GJ, Elias S. Point-counterpoint. Patient autonomy and incidental findings in clinical genomics. Science. 2013;340:1049–50.
47. Allyse M, Michie M. Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. Trends Biotechnol. 2013;31:439–41.
48. ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet Med. 2015;17:68–9.
49. Supreme Court of the United States. Association for Molecular Pathology et al. vs. Myriad Genetics, Inc.; 2013. http://www.supremecourt.gov/opinions/12pdf/12-398_1b7d.pdf.
50. Béroud C, Letovsky SI, Braastad CD, et al. BRCA share: a collection of clinical BRCA gene variants. Hum Mutat. 2016;37:1318–28.
51. Hirschhorn K, Fleisher LD, Godmilow L, et al. Duty to re-contact. Genet Med. 1999;1:171–2.
52. Pyeritz RE. The coming explosion in genetic testing – is there a duty to recontact? N Engl J Med. 2011;365:1367–9.

Emerging Next-Generation Sequencing Technologies

3

Matthew W. Anderson

Introduction

What would be the ideal sequencing machine? Can a single sequencing technology replace all the tools of molecular pathology and cytogenetics that we currently utilize to interrogate the genome in both health and disease? An ideal sequencing instrument should detect all types of genomic variation including structural [single nucleotide polymorphisms (SNPs), indels, copy number variation, inversions, chromosomal rearrangements], epigenomic, and transcriptional. Long read lengths are required to enable efficient genomic assembly and accurate phasing, and the detection method must produce highly accurate base calls to minimize errors and reduce costly iterative sequencing. Finally, the system should be inexpensive, be easy to maintain and operate, and require short run times. At the time of writing this chapter, such a sequencing machine simply does not exist. However, engineers, physicists, and biologists in both industry and academia are actively working to solve the major technical challenges facing the development of new sequencing technologies. While this chapter focuses on discussing these "emerging" sequencing technologies, the reader is cautioned that the development of new sequencing technologies occurs at a prodigious pace. In addition, many of the cutting-edge advances in sequencing technology are being developed within a commercial environment, where it is difficult for those outside the company to obtain detailed and vetted information about instrument performance. With those caveats in mind, this chapter endeavors to provide a broad overview of emerging new sequencing technologies and some of the potential applications in nucleic acid analysis which will be enabled by these technological advances.

M. W. Anderson (✉)
Diagnostic Laboratories, BloodCenter of Wisconsin, part of Versiti, Milwaukee, WI, USA
e-mail: matthew.anderson@bcw.edu

Advantages of Single-Molecule Sequencing

Biomolecular detection (whether for sequencing, chemistry, or immunology) typically requires a signal amplification step for robust and reproducible analyte detection. For both Sanger and many next-generation sequencing platforms, signal amplification occurs through PCR amplification of the target DNA, ensuring that the fluorescence or luminescence signals generated during the sequencing reaction are sufficiently strong. Although PCR is the mainstay of molecular biology protocols, it is not without its disadvantages. Sequence artifacts can be generated during the PCR reaction to include nucleotide misincorporation events, amplification bias due to GC content, preferential allele amplification, and formation of chimeric sequences during later PCR cycles, all of which can be reflected in the sequencing results [1–4]. When smaller regions of DNA are amplified, significant contextual information (phase, haplotype, etc.) is also lost.

To solve these issues, many newer sequencing technologies are designed to sequence individual nucleic acid molecules. Single-molecule sequencing offers a number of practical advantages. Sample preparation is greatly simplified, because there is less experimental manipulation required to create sequencing libraries. For example, the ability to directly sequence RNA would eliminate the additional steps typically required to convert RNA into cDNA prior to sequencing. Single-molecule approaches would theoretically reduce the required amount of input DNA, an important consideration for the analysis of rare cellular populations or individual cells. Single-molecule sequencing also enables long templates to be sequenced in phase, preserving long-range structural variation.

Single-Molecule Cycle Sequencing

In 2003, Stephen Quake and colleagues were the first to report single-molecule DNA sequencing through the use of fluorescence microscopy and fluorescence resonance energy transfer (FRET) [5]. Using this technique, the authors were

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_3

able to detect the incorporation of up to five nucleotides on a single DNA template. Although the strategy was promising, the relatively short molecular distance over which FRET can occur limited the theoretical sequencing read length to approximately 15 base pairs (bp). A year later, Quake cofounded Helicos BioSciences (Cambridge, MA), with the goal of developing a commercial sequencing instrument based on single-molecule sequencing. In the Helicos sequencing strategy, FRET-based detection is replaced by a "sequencing by synthesis" approach, in which fluorescently labeled nucleotides are added sequentially during the sequencing reaction. Similar to other high-throughput sequencing technologies, only one type of nucleotide (A, T, G, or C) is added to the reaction during each cycle of sequencing. Therefore, not every template molecule incorporates a nucleotide during each round of sequencing.

The Helicos sequencing protocol is relatively simple. First, sequencing libraries are prepared by randomly fragmenting genomic DNA to produce short (100–200 bp) fragments. Next, multiple adenosine nucleotides are added to the 3′ end of the DNA fragments to allow the template molecules to hybridize to poly-T oligonucleotide anchors on the surface of the flow cell (Fig. 3.1). The terminal adenosine nucleotide is fluorescently labeled so that each template molecule can be spatially localized on the flow cell surface, prior to the start of sequencing. During the sequencing reaction, DNA polymerase and one of four fluorescently labeled nucleotides are sequentially added to the flow cell. Each nucleotide is modified with a terminator moiety to prevent multiple nucleotide additions during each sequencing cycle. After nucleotide incorporation, the array is imaged, and the terminator moiety and fluorescence label are removed to enable subsequent rounds of sequencing.

In 2008, the company shipped its first sequencing instrument and reported the use of the technology to sequence the M13 phage genome [6]. Utilizing data generated from sequencing his own genome with the Helicos instrument [7], Quake and his group reported an average read length of 33 bp, and an error profile composed predominantly of deletions and insertions (approximately 3–5% overall). Although the raw read error rate was relatively high, the overall consensus accuracy was 99% for SNPs. Although the same genomic sequence was later analyzed by a multidisciplinary group at Stanford to create a personal genome-based clinical assessment [8], the high error rate, short read length, and high cost per base made the Helicos technology impractical for whole-genome sequencing. The Helicos single-molecule sequencing approach has since been shown to be advantageous for other applications including direct RNA sequencing (RNA-seq) [9] and chromatin profiling (CHiP-seq) [10]. Although Helicos BioSciences ceased commercial operations in 2012, the technology has been licensed to a company (Direct Genomics) that plans to adapt the Helicos approach for targeted clinical diagnostic sequencing [11].

Real-Time Single-Molecule Sequencing with Polymerase

What if one could directly observe DNA polymerase as it synthesizes DNA? DNA sequence information would be generated in "real time" at a rate equal to nucleotide incorporation catalyzed by DNA polymerase, with read lengths theoretically limited only by the processivity of the polymerase or the size of the DNA template. By eliminating the iterative sequencing cycles required by current high-throughput

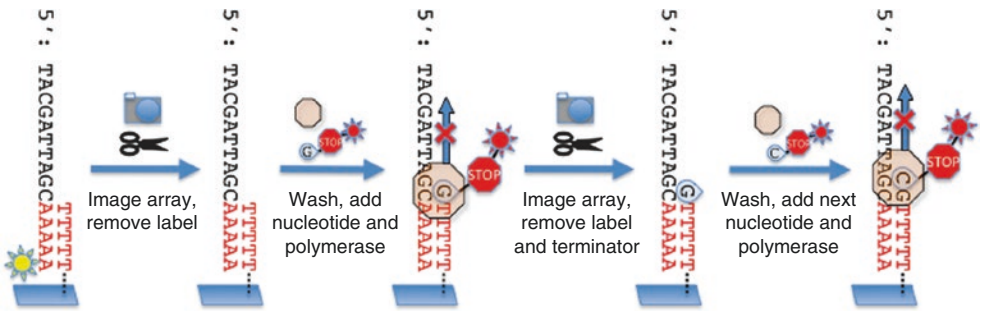


Fig. 3.1 Helicos sequencing chemistryDNA template molecules are modified by the 3′ addition of adenosine nucleotides and hybridized to poly-T oligonucleotides covalently linked to the surface of the flow cell. The terminal adenosine nucleotide is fluorescently labeled to allow the instrument to record the location of each template molecule on the flow cell surface. Prior to the start of sequencing, the fluorescently labeled 3′ adenosine is cleaved and washed away. During each round of sequencing, polymerase and a single fluorescently labeled nucleotide (A, T, G, or C) are added to the flow cell. The labeled nucleotides are modified with a cleavable terminator residue which prevents multiple base incorporations during each cycle. After nucleotide incorporation, the array is imaged, and the fluorescence signal is recorded for each template molecule. Once the images are captured, the fluorescent label and terminator are removed to regenerate a template suitable for the next round of nucleotide addition. (From Anderson and Schrijver [57], Fig. 4b. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/))

sequencing technologies, the overall cost of obtaining a complete genome sequence could also be significantly reduced.

Whereas the potential advantages of real-time single-molecule DNA sequencing are readily apparent, designing a sequencing instrument that can "eavesdrop" on a single DNA polymerase molecule is extremely challenging from an engineering and biophysics perspective. For example, the detection method must be able to accurately detect signals generated from the activity of a single DNA polymerase molecule, and true nucleotide incorporation events must be discerned against a background of high concentrations of unbound labeled nucleotides. Two solutions to this problem have been developed, either through physically confining DNA polymerase to a small observation volume or by the use of FRET to detect when a labeled nucleotide is in close proximity to the polymerase active site.

In 2003, Levene et al. reported the development of "zero-mode waveguide" (ZMW) technology, a technique that utilizes nanoscale holes in a metal film to restrict incident laser light to a small focused detection volume approximating 10^{-21} l (zeptoliter) [12]. By creating an extremely focused region in which laser light can excite a fluorophore, ZMVs enable single-molecule analysis in the presence of high concentrations of fluorescently labeled ligands. As a demonstration of the applicability of ZMW technology for DNA sequencing applications, the authors immobilized DNA polymerase and M13 phage DNA within the detection volume of ZMWs. After the addition of a fluorescently labeled nucleotide, temporally distinct fluorescent signals were detected within the ZMWs consistent with polymerase-catalyzed nucleotide incorporation events.

In 2004, Pacific Biosciences (Menlo Park, CA) was founded to develop a DNA sequencing instrument using ZMW technology. The company published its first proof-of-concept study describing single-molecule real-time (SMRT®) sequencing in 2009 [13] and released its first commercial DNA sequencing instrument (PacBio RS) in 2010. In the SMRT sequencing method, DNA template libraries are prepared by shearing genomic DNA into 250 bp to 10 kilobase (kb) fragments and ligating hairpin adapters to each end of the molecule to create a circular DNA template. Primed DNA templates lacking the hairpin adapters can also be sequenced, but the number of reads generated is reduced considerably [14]. Individual DNA polymerase molecules bound to DNA template are then localized at the bottom of the ZMW through simple diffusion and biotin/streptavidin interactions [15]. A mixture of nucleotides is subsequently added to the chip, with each nucleotide uniquely labeled with a different fluorophore attached to the base via linkage to the phosphate chain. Unbound nucleotides rapidly diffuse in and out of the ZMV detection volume, far too quickly to be registered as a fluorescence signal by the detector (Fig. 3.2). When a nucleotide enters the active site of the polymerase, its motion is dramatically slowed, allowing time for the laser to excite the fluorophore and generate a fluorescent signal. DNA polymerase cleaves the phosphate chain as the nucleotide is incorporated, freeing the fluorophore to rapidly diffuse out of the detection volume of the ZMW. The reaction reconstitutes a free 3′-hydroxyl group, which can then be used for the next round of nucleotide addition.

SMRT sequencing has a high median per-base error rate (~11%) dominated by insertions and deletions, presumably due to non-templated nucleotides binding to the active site of the polymerase and incomplete incorporation events [16]. Interestingly, the error rate profile appears to be random and not context-specific (i.e., homopolymer errors) as opposed to other high-throughput sequencing platforms [17]. Much of the high per-base error rate can be overcome through iterative sequencing of the circular templates, resulting in very high consensus accuracy (>99.9%) [13]. Because the error rate is independent of sequence context, GC-rich and homopolymer regions of the genome can be sequenced and analyzed including the CGG repeat region of the *FMR1* gene implicated in fragile X syndrome [18].

An attractive feature of SMRT sequencing is the ability to directly detect modified bases such as 5-methylcytosine [19]. Compared to standard high-throughput sequencing techniques which rely on bisulfite treatment of the DNA library to characterize methylation [20], SMRT sequencing can directly detect modified bases during the sequencing reaction as changes in the kinetics of DNA polymerase. Because no prior chemical modification of the DNA library is required, multiple different base modifications on the same DNA template molecule can be detected simultaneously [21]. Another key advantage of SMRT technology is long read lengths averaging 10 kb, with maximum read lengths of >60 kb. The long read length of SMRT sequencing not only increases the accuracy of germline structural variant characterization [22] but also has clinical utility for sequencing complex immunogenetic loci such as *KIR* and *HLA* [23] and characterizing tumor-associated gene fusions [24].

In 2015, PacBio released the Sequel, a less expensive yet higher-throughput version of the RS II instrument. The Sequel utilizes a SMRT chip containing approximately one million ZMWs, which generates 365,000 reads per run, generating sufficient sequence output to produce a low-coverage (10X) human genome in a single run. Although PacBio may not currently be the platform of choice for low-cost high-throughput whole-genome sequencing, the technology clearly has unique advantages for the analysis of complex regions of the genome.

An alternative approach for single-molecule polymerase-based sequencing utilizes FRET to detect nucleotide incorporation events. Although there are no published data on this technology, presentations describing the method suggest that the strategy involves generating FRET between quantum dot-labeled DNA polymerase molecules and fluorescently

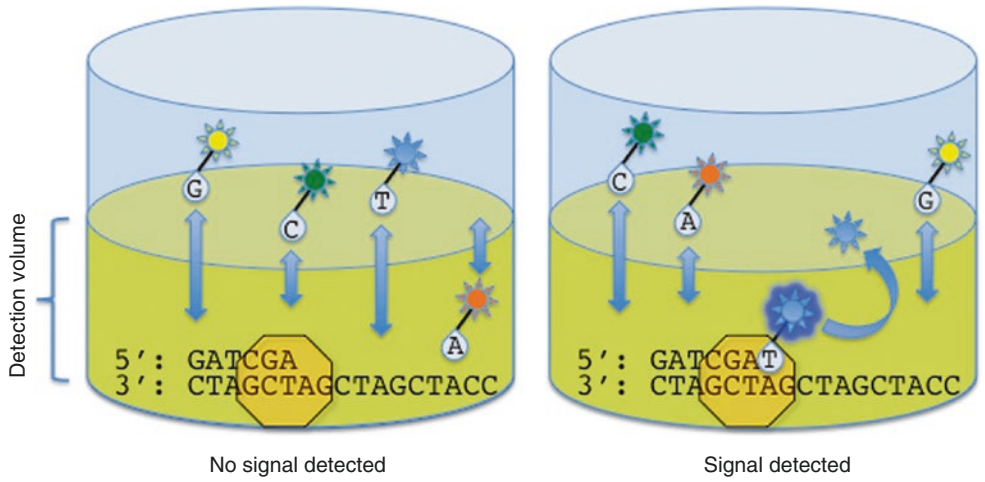


Fig. 3.2 Real-time single-molecule sequencing with polymerase. In SMRT chemistry, DNA polymerase and template molecules are immobilized at the bottom of a zero-mode waveguide (ZMW) sequencing well. The ZMW focuses laser energy to create an extremely small detection volume at the bottom of the ZMW where the polymerase and template molecule are localized. A mixture of all four nucleotides is added to the ZMW, each uniquely labeled with a different fluorescence moiety. Unincorporated nucleotides rapidly diffuse in and out of the detection volume at a rate too fast for a fluorescence signal to be recorded. As a nucleotide is incorporated, the fluorescent moiety is held within the detection volume long enough to be excited by the laser and give off a fluorescence signal which can be recorded. During nucleotide addition, the fluorophore is cleaved away as the phosphodiester bond is formed. The liberated fluorophore rapidly diffuses out of the ZMW detection volume, terminating the fluorescence signal for that particular nucleotide incorporation event. (From Anderson and Schrijver [57], Fig. 5. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/))

labeled nucleotides [25]. When a fluorescently labeled nucleotide enters the active site of the polymerase, two signals are generated during FRET. The signal from the quantum dot decreases, indicating that a nucleotide is bound to the active site, while the signal from the bound fluorescently labeled nucleotide increases. Theoretically, the presence of two distinct but temporally related signals from the same nucleotide incorporation event could result in highly accurate base calls. Despite the potential of this approach, plans for further development or commercialization remain unclear.

Sequencing Through Direct Imaging

Direct visualization of biological macromolecules has long been proposed as an approach to determine nucleic and amino acid sequences. The potential benefits of sequencing through direct imaging include extremely long read lengths, fast analysis, and preservation of large-scale structural variation. Optical-based approaches to mapping long DNA fragments have been developed utilizing either restriction enzymes or fluorescent labeling [26], but the resolution of light or fluorescence microscopy is far too low to allow for single-base identification. Recently, scanning transmission electron microscopy (STEM) has been explored as a potential direct imaging sequencing technology. In fact, the use of electron microscopy to image DNA is not new. In the 1970s, STEM was used to generate low-resolution images of purified genomic DNA from both *Drosophila* [27] and human samples [28]. Secondary structures of DNA (hairpin loops, etc.) were readily observable, and the images provided support for the presence of inverted repeat sequences in the human genome. However, for STEM to become useful for nucleic acid sequencing, it must demonstrate sufficient resolution to accurately image and identify each nucleotide by its unique chemical structure.

In STEM, resolution is directly related to the ability of a target atom to scatter the electron beam. Heavier atoms are more easily visualized as the higher atomic mass results in increased electron scattering. Unfortunately, STEM cannot readily distinguish between each base in a nucleic acid sequence because natural nucleotides differ by only a few atoms with low atomic mass. Therefore, STEM-based approaches to sequence DNA must involve modifications to the target DNA (i.e., heavy atom labeling) to make the nucleotides "visible" to the electron microscope.

In a paper describing the use of STEM to sequence DNA [29], target DNA molecules are labeled by performing PCR in the presence of thymine nucleotides modified with a single mercury atom. Mercury-labeled DNA molecules are then purified, linearized, and deposited onto a carbon substrate. The labeled DNA strands are imaged with the electron microscope, and modified nucleotides are detected by an increase in current as the electron beam is scattered by the heavy atom label (Fig. 3.3). For their initial experiments, the authors used

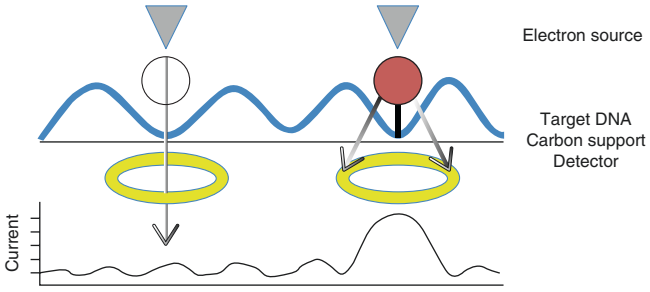


Fig. 3.3 Sequencing with electron microscopy. Fragments of DNA are labeled through the incorporation of nucleotides modified with heavy atoms (*red circle*). Labeled DNA template molecules are then stretched and applied to a carbon substrate overlying a circular detector. When the electron beam encounters an unlabeled nucleotide (*open circle*), the path of the beam is undisturbed, and the electrons pass through the center of the detector. In contrast, heavy atoms scatter the electron beam, resulting in increased current within the detector. (Figure adapted from Bell et al. [29])

M13 phage and a synthetic DNA molecule that contained labeled thymine nucleotides at defined positions in the sequence. Although STEM could detect labeled thymines in the test DNA molecules, only half of the labeled thymine residues predicted by the test sequence were identified due either to inefficient incorporation during PCR or loss during processing of the sequencing templates. In addition, there was partial overlap between the signals generated by background current and the current generated by labeled thymine.

Although promising, several technical hurdles must be overcome before STEM becomes a viable approach to DNA sequencing. In particular, improved methods must be devised to uniquely label each nucleotide and ensure a high degree of label incorporation into the sequencing templates. One potential solution to this problem would be the use of low-energy electron microscopy to image and sequence unlabeled DNA molecules [30]. The commercial potential of electron microscopy for DNA sequencing is unclear, but mapping of long DNA fragments by either STEM or other optical-based approaches may ultimately provide complementary structural information to guide genome assembly from short read high-throughput sequencing data [31, 32].

Sequencing with Protein Nanopores

To date, all single-molecule sequencing approaches require labeling (fluorescence, heavy atoms, etc.) of the template molecule or nucleotides. Labeling adds complex preparation steps to the sequencing workflow, increases reagent costs, and can have adverse effects on the sequencing reaction such as inhibiting the action of polymerase. An alternative approach to avoid the use of labels would be through electrochemical detection of nucleotides in the DNA sequence. Electrochemical detection would eliminate the use of expensive enzymes and labels and could represent the fastest and simplest option for DNA sequencing.

In the early 1990s, David Deamer, Daniel Branton, and George Church filed a patent application for the use of protein membrane channels (nanopores) as a method for electrochemical nucleic acid sequencing. The authors envisioned that an ionic gradient could be established across a lipid bilayer containing nanopores. Nucleic acids passing through the nanopores would disrupt the flow of ions, resulting in changes in current that could be used to decode the nucleotide sequence. In an initial proof-of-concept study [33], Deamer and coworkers showed that RNA and DNA molecules could transit through α-hemolysin nanopores derived from *Staphylococcus aureus* and the passage of nucleic acids through the nanopore was reflected by changes in ionic current. Unfortunately, nucleic acid molecules passed through the nanopore far too quickly for each nucleotide in the sequence to generate a unique base-specific change in current. Therefore, new strategies had to be devised to control the transit of nucleic acids through the nanopore in order for each nucleotide to be reliably detected for sequencing.

In subsequent years, significant advances in protein and electrochemical engineering have enabled routine nanopore sequencing [34]. Modifications were made to several naturally occurring nanopores to improve the signals generated as nucleotides or polynucleotides traverse the nanopore [35], and enzymatic approaches were developed to control the movement of nucleic acids through the nanopore [36, 37]. One strategy developed by Hagan Bayley and colleagues involved the use of exonuclease to cleave individual nucleotides from a nucleic acid polymer and then detection of the free nucleotides as they flowed through the nanopore channel [38]. The theoretical advantage of this approach is that the signals are significantly less complex because only four distinct current signals are generated, each unique to a different nucleotide. However, the accuracy of this technique is critically dependent on close alignment between the exonuclease and the nanopore to ensure that each liberated nucleotide is detected exactly in order according to the nucleotide sequence on the template strand. A related approach (Fig. 3.4a) currently under commercial development by Genia Technologies (Santa Clara, CA) utilizes DNA polymerase covalently linked to a α-hemolysin pore [39]. Template DNA is extended in the presence of modified nucleotides to which variably sized oligonucleotide polymer tags are attached to the terminal phosphate moiety. Alteration of current through the nanopore occurs, while the modified nucleotide remains within the active site of the DNA polymerase and before the tag is removed during nucleotide incorporation. Changes in current through the nanopore are proportional to the different molecular sizes of the tag molecules. These characteristic patterns of current changes can then be analyzed and converted into sequence reads.

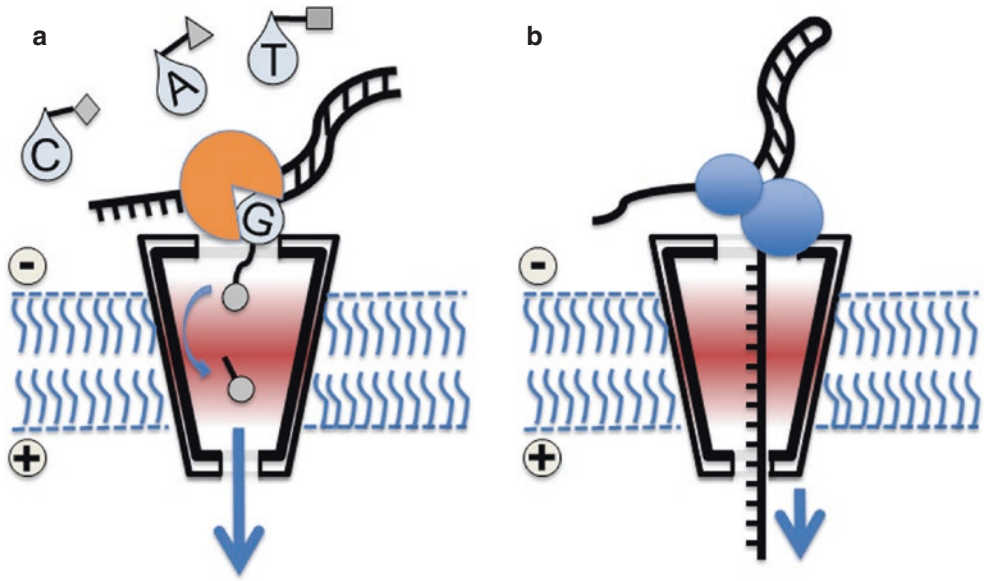


Fig. 3.4 Nanopore sequencing. (**a**) Nanopore sequencing with tagged nucleotides. Nucleotides are synthesized with differentially sized polymer "tags." As the tagged nucleotides bind the active site of the DNA polymerase, the tag polymer enters the pore causing a current blockade which is dependent on the relative molecular size of the tag. As the nucleotide is incorporated into the template strand, the tag is liberated and passes through the pore completing the sequencing cycle. (**b**) Nanopore sequencing with helicase. Double-stranded DNA is modified with a hairpin adapter and an adapter sequence which facilitates binding of the DNA to a helicase enzyme localized to the entrance of the pore. As the helicase "unzips" the double-stranded DNA molecule, single-stranded DNA enters the pore altering the current in a manner dependent on the nucleotide sequence. The hairpin adapter enables sequencing of the complementary DNA strand to improve the accuracy of base calling

An alternative approach is to utilize molecular motors to control the movement of single-stranded nucleic acid in the nanopore. In 2012, two groups led by Jens Gundlach and Mark Akeson showed that DNA polymerase could control the rate of nucleic acid translocation through a protein nanopore to generate well-defined ionic currents [36, 37]. These initial nanopore experiments led to the development of the Oxford Nanopore Technologies MinION (Oxford, UK), which became commercially available in 2014 through an early access program [40]. The MinION device itself is small and lightweight, requiring only a USB3 port and a standard desktop computer for operation. The MinION sequencing approach takes advantage of helicase enzymes to control translocation of single-stranded DNA through the nanopore (Fig. 3.4b). First, sequencing libraries are prepared by modifying double-stranded DNA molecules with a hairpin and a leader adapter, respectively. The leader adapter enhances binding of the DNA templates to helicase molecules localized to the entrance of the nanopore. As the helicase enzyme unwinds the double-stranded DNA molecule, single-stranded DNA enters the pore, altering the current flowing through the nanopore in a sequence-dependent manner. The complex changes in ionic current produced by DNA translocation through the nanopore are analyzed and converted into short 5–6-nucleotide-long sequences termed "kmers." Sequence reads derived from the template strand ("1D" reads) have a reported accuracy of 65–75%, which can improve to 80–88% by sequencing the complementary strand of the DNA molecule ("2D" reads) through the hairpin adapter [41]. The sequencing chemistry continues to improve rapidly, with newer chemistries achieving single-read accuracy of 92% [42].

One of the major advantages of MinION sequencing is the extremely long sequencing read lengths (>45 kb) [42] which may be advantageous for applications such as resolving complex repeat regions in the genome [43], structural variants in cancer [44], and RNA splice variants [45]. One of the drawbacks of the MinION device is its relatively low throughput, limiting the applications to small genomes and targeted sequencing. However, the simplicity and portability of the MinION device can be very useful for rapid point-of-care diagnostics as demonstrated by the use of the MinION to rapidly sequence Ebola virus isolates during an outbreak [46]. Oxford Nanopore Technologies is currently addressing the needs of genomic researchers by developing a larger throughput instrument (PromethION) which is currently in an early access program.

Solid-State Nanopore Sequencing

An alternative approach to the use of engineered protein pores for sequencing would be to construct nanopores from inorganic materials (solid-state nanopores). Leveraging advances in materials science, solid-state nanopores have

been fabricated from silicon [47], graphene [48], and carbon nanotubes [49]. Theoretically, solid-state nanopores would be highly stable and could be manufactured utilizing existing infrastructure built by the semiconductor industry.

Various approaches have been developed to detect nucleic acids as they pass through a solid-state nanopore. Nabsys (Providence, RI) is developing a solid-state nanopore sequencing system that relies on hybridization of labeled probe oligonucleotides to single-stranded DNA [50]. After hybridization, the remaining template molecule is converted to double-stranded DNA and coated with a DNA-binding protein to increase the resistance as the template moves through the nanopore [51]. The template sequence can then be decoded by determining the relative positions of the labeled probes. Although contiguous stretches of DNA can be sequenced using this method, an added benefit is that the density of the probes can be reduced to enable mapping of large genomes [52]. However, for this strategy to be effective, the nanopore must be able to precisely determine the nucleotide distance between the probes.

An added advantage of solid-state nanopores is that they can be modified with electrochemical sensors that can detect and identify nucleotides as they pass through the nanopore. For example, solid-state nanopores have been fitted with nanoelectrodes that transfer tunneling current through passing nucleotides to generate unique current signatures [53]. Chemically modified probes have also been designed that facilitate tunneling current by forming complementary interactions with each base as it passes through the nanopore [54]. In an elegant solution, graphene ribbons have been used to identify nucleotides in a near-planar orientation by taking advantage of natural pi-stacking interactions contributed by the aromatic rings of the nucleobases [55]. Although solid-state nanopore sequencing is still in the research and development stage, the first commercial application of solid-state nanopore technology may be for targeted molecular diagnostics. Recently, investigators at Two Pore Guys (Santa Cruz, CA) reported that the ΔF508 *CFTR* mutation could be reliably detected using solid-state nanopores and synthetic peptide nucleic acid probes [56]. If other macromolecules (antibodies, etc.) can be identified using this approach, solid-state nanopore technology may eventually prove useful for point-of-care diagnostics.

Conclusions

This chapter provides an overview of the evolution of single-molecule sequencing technologies, from fluorescence-based approaches to direct electrochemical detection via solid-state nanopores. Although it is difficult to predict which sequencing technology will eventually find widespread adoption, there appears to be an inexorable progression toward the goal of sequencing individual nucleic acid molecules with virtually no sample preparation. Given the complexity of the human genome, it is likely that no one single technology will provide a complete solution for genomic analysis. However, emerging single-molecule sequencing approaches appear poised to revolutionize clinical molecular diagnostics if they can deliver on the promise of fast, cost-effective, and accurate high-throughput sequencing.

References

1. Walsh PS, Erlich HA, Higuchi R. Preferential PCR amplification of alleles: mechanisms and solutions. PCR Methods Appl. 1992;1:241–50.
2. Cha RS, Thilly WG. Specificity, efficiency, and fidelity of PCR. PCR Methods Appl. 1993;3:S18–29.
3. Shuldiner AR, Nirula A, Roth J. Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res. 1989;17:4409.
4. Mutter GL, Boynton KA. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. Nucleic Acids Res. 1995;23:1411–8.
5. Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A. 2003;100:3960–4.
6. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. Single-molecule DNA sequencing of a viral genome. Science. 2008;320:106–9.
7. Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. Nat Biotechnol. 2009;27:847–52.
8. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. Clinical assessment incorporating a personal genome. Lancet. 2010;375:1525–35.
9. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. Nature. 2009;461:814–8.
10. Goren A, Ozsolak F, Shoresh N, Ku M, Adli M, Hart C, Gymrek M, Zuk O, Regev A, Milos PM, Bernstein BE. Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. Nat Methods. 2010;7:47–9.
11. Krol A. Direct genomics' new clinical sequencer revives a forgotten DNA technology. Bio-IT World. Oct 29, 2015. (http://www.bio-itworld.com/2015/10/29/direct-genomics-new-clinical-sequencer-revives-forgotten-dna-technology.html). Last accessed 22 March 2017.
12. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;299:682–6.
13. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.

14. Coupland P, Chandra T, Quail M, Reik W, Swerdlow H. Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. BioTechniques. 2012;53:365–72.
15. Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc Natl Acad Sci U S A. 2008;105:1176–81.
16. Korlach J. White paper: understanding accuracy in SMRT® sequencing. 2013. (www.pacb.com/wp-content/uploads/2015/09/Perspective_UnderstandingAccuracySMRTSequencing.pdf). Last accessed 5/10/17.
17. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012;13:375.
18. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, McCalmon S, Hagerman RJ, Tassone F, Hagerman PJ. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res. 2013;23:121–8.
19. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7:461–5.
20. Lister R, Ecker JR. Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res. 2009;19:959–66.
21. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat Biotechnol. 2012;30:1232–9.
22. English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, Ritter DI, Beck CR, Davis CF, Dahdouli M, Ma S, Carroll A, Veeraraghavan N, Bruestle J, Drees B, Hastie A, Lam ET, White S, Mishra P, Wang M, Han Y, Zhang F, Stankiewicz P, Wheeler DA, Reid JG, Muzny DM, Rogers J, Sabo A, Worley KC, Lupski JR, Boerwinkle E, Gibbs RA. Assessing structural variation in a personal genome – towards a human reference diploid genome. BMC Genomics. 2015;16:286–301.
23. Mayor NP, Robinson J, McWhinnie AJM, Ranande S, Eng K, Midwinter W, Bultitude WP, Chin CS, Bowman B, Marks P, Braund H, Madrigal JA, Latham K, Marsh SGE. HLA typing for the next generation. PLoS One. 2015;10:e0127153.
24. Cavelier L, Ameur A, Häggqvist S, Höijer I, Cahill N, Olsson-Strömberg U, Hermanson M. Clonal distribution of BCR-ABL1 mutations and splice isoforms by single-molecule long-read RNA sequencing. BMC Cancer. 2015;15:45–57.
25. Vander Horn PB. Single molecule real-time sequencing on the surface of a quantum-dot nanocrystal. J Biomol Tech. 2011;22(Suppl):S9.
26. Neely RK, Deen J, Hofkens J. Optical mapping of DNA: single-molecule-based methods for mapping genomes. Biopolymers. 2011;95:298–311.
27. Schmid CW, Manning JE, Davidson N. Inverted repeat sequences in the Drosophila genome. Cell. 1975;5:159–72.
28. Deininger PL, Schmid CW. An electron microscope study of the DNA sequence organization of the human genome. J Mol Biol. 1976;106:773–90.
29. Bell DC, Thomas WK, Murtagh KM, Dionne CA, Graham AC, Anderson JE, Glover WR. DNA base identification by electron microscopy. Microsc Microanal. 2012;18:1049–53.
30. Mankos M, Shadman K, Persson HHJ, N'Diaye AT, Schmid AK, Davis RW. A novel low energy electron microscope for DNA sequencing and surface analysis. Ultramicroscopy. 2014;145:36–49.
31. Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnheim R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC. High-resolution human genome structure by single-molecule analysis. Proc Natl Acad Sci U S A. 2010;107:10848–53.
32. Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, Goldstein S, Pape L, Mehan MR, Churas C, Pasternak S, Forrest DK, Wise R, Ware D, Wing RA, Waterman MS, Livny M, Schwartz DC. A single molecule scaffold for the maize genome. PLoS Genet. 2009;5:e1000711.
33. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci U S A. 1996;93:13770–3.
34. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. Nat Biotechnol. 2016;34:518–24.
35. Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, Gundlach JH. Nanopore DNA sequencing with MspA. Proc Natl Acad Sci U S A. 2010;107:16060–5.
36. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat Biotechnol. 2012;30:349–53.
37. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nat Biotechnol. 2012;30:344–8.
38. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4:265–70.
39. Fuller CW, Kumar S, Porel M, Chien M, Bibillo A, Benjamin Stranges P, Dorwart M, Tao C, Li Z, Guo W, Shi S, Korenblum D, Trans A, Aguirre A, Liu E, Harada ET, Pollard J, Bhat A, Cech C, Yang A, Arnold C, Palla M, Hovis J, Chen R, Morozova I, Kalachikov S, Russo JJ, Kasianowicz JJ, Davis R, Roever S, Church GM, Ju J. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. Proc Natl Acad Sci U S A. 2016;113:5233–8.
40. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17:239–50.
41. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif. 2015;3:1–8.
42. Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, Leggett RM, Eccles DA, Zalunin V, Urban JM, Piazza P, Bowden RJ, Paten B, Mwaigwisya S, Batty EM, Simpson JT, Snutch TP, Birney E, Buck D, Goodwin S, Jansen HJ, O'Grady J, Olsen HE. MinION analysis and reference consortium: phase 1 data release and analysis. F1000Res. 2015;4:1075–110.
43. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nat Methods. 2015;12:351–6.
44. Norris AL, Workman RE, Fan Y, Eshleman JR, Timp W. Nanopore sequencing detects structural variants in cancer. Cancer Biol Ther. 2016;17:246–53.
45. Bolisetty MT, Rajadinakaran G, Graveley BR. Determining exon connectivity in complex mRNAs by nanopore sequencing. Genome Biol. 2015;16:204–16.
46. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sánchez Á, Carter LL, Doerrbecker J, Enkirch T, García-Dorival I, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallasch E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Racine T, Bello A, Sall AA, Faye O, Faye O, Magassouba N, Williams CV, Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil

V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N, Roger I, Williams CJ, Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner DJ, Pollakis G, Hiscox JA, Matthews DA, O'Shea MK, Johnston AM, Wilson D, Hutley E, Smit E, Di Caro A, Wölfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B, Keïta S, Rambaut A, Formenty P, Günther S, Carroll MW. Real-time portable genome sequencing for Ebola surveillance. Nature. 2016;530:228–32.

47. Rosenstein JK, Wanunu M, Merchant CA, Drndic M, Shepard KL. Integrated nanopore sensing platform with sub-microsecond temporal resolution. Nat Methods. 2012;9:487–92.
48. Heerema SJ, Dekker C. Graphene nanodevices for DNA sequencing. Nat Nanotech. 2016;11:127–36.
49. Liu H, He J, Tang J, Liu H, Pang P, Cao D, Krstic P, Joseph S, Lindsay S, Nuckolls C. Translocation of single-stranded DNA through single-walled carbon nanotubes. Science. 2010;327:64–7.
50. Balagurusamy VS, Weinger P, Ling XS. Detection of DNA hybridizations using solid-state nanopores. Nanotechnology. 2010;21:335102.
51. Kowalczyk SW, Hall AR, Dekker C. Detection of local protein structures along DNA using solid-state nanopores. Nano Lett. 2010;10:324–8.
52. Thompson JF, Oliver JS. Mapping and sequencing DNA using nanopores and nanodetectors. Electrophoresis. 2012;33:3429–36.
53. Ohshiro T, Matsubara K, Tsutsui M, Furuhashi M, Taniguchi M, Kawai T. Single-molecule electrical random resequencing of DNA and RNA. Sci Rep. 2012;2:501.
54. Ohshiro T, Umezawa Y. Complementary base-pair-facilitated electron tunneling for electrically pinpointing complementary nucleobases. Proc Natl Acad Sci U S A. 2006;103:10–4.
55. Min SK, Kim WY, Cho Y, Kim KS. Fast DNA sequencing with a graphene-based nanochannel device. Nat Nanotechnol. 2011;6:162–5.
56. Morin TJ, Shropshire T, Liu X, Briggs K, Huynh C, Tabard-Cossa V, Wang H, Dunbar WB. Nanopore-based target sequence detection. PLoS One. 2016;11:e0154426.
57. Anderson MW, Schrijver I. Next generation DNA sequencing and the future of genomic medicine. Genes. 2010;1:38–69.

Transcriptome Sequencing (RNA-Seq)

4

Jacquelyn Reuther, Angshumoy Roy,
and Federico A. Monzon

Introduction

The transcriptome is the entire assembly of RNA transcripts in a given cell type, including protein-coding RNA such as messenger RNA (mRNA) and nonprotein-coding transcripts like ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA), and other noncoding RNA (ncRNA) [1]. As opposed to the genome, which is shared by all cells in a given organism, the transcriptome is specific to a given tissue or cell type or even specific to the single-cell level. Transcriptome sequencing (RNA-Seq) is a recently developed technology that uses high-throughput sequencing approaches to determine the sequence of all RNA transcripts in a given specimen. Because it is not based on hybridization to existing probes, RNA-Seq is regarded as an unbiased technique to assess differential gene expression and thus allows the identification of novel transcripts as well as detecting splicing and/or allelic usage patterns that are present in specific cells or situations. This technology has rapidly deepened our understanding of gene expression profiles of various tissues and cells, including a better understanding of the use of alternative splicing in normal and disease processes, the role of functional elements of the genome, and the role of noncoding RNAs, and has enabled us to discover a large number of fusion transcripts in cancer [2, 3]. The use of this technology for transcriptome research has allowed new insights in development and in the study of benign and malignant disease processes and has now migrated into the clinical arena with diagnostic tests aimed at detecting oncogenic fusion transcripts with a role in the diagnosis and management of cancer patients [3–5].

J. Reuther
Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA
e-mail: Jacquelyn.reuther@bcm.edu

A. Roy
Department of Pathology and Immunology and Pediatrics, Baylor College of Medicine, Houston, TX, USA
e-mail: aroy@bcm.edu

F. A. Monzon (✉)
Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA

Castle Biosciences, Friendswood, TX, USA
e-mail: fmonzon@castlebiosciences.com

Technologies in Transcriptomics

Multiple technologies exist to interrogate the transcriptome, each with their own advantages and disadvantages (Table 4.1). In the following section, three of the main technologies are discussed, including methodologies, their advantages, associated biases, and limitations.

Microarrays

Much of our current knowledge regarding clinically relevant gene expression signatures have come from the use of microarrays, with the technology aiding in areas such as patient risk stratification and identification of tissue origin for poorly differentiated or undifferentiated tumor specimens [6]. Microarrays are able to determine expression levels of known, highly annotated transcripts via hybridization of purified RNA to complementary oligos fixed on a solid substrate. Microarrays can be used with total RNA extracts from both frozen and formalin-fixed paraffin-embedded (FFPE) tissue, with transcript measurements showing 80–97% consistency between the two sample types [7, 8]. Total RNA is extracted from the specimens of interest followed by complementary DNA (cDNA) generation by reverse transcription (RT). Reverse-transcribed transcripts are then fluorescently labeled and hybridized to the array which can have probes generated from cDNA, oligonucleotides, polymerase chain reaction (PCR) fragments, restriction enzyme-digested fragments, oligomers, or expressed sequence tags (ESTs) [9]. Once the transcripts are hybridized, the array is scanned with a laser to measure the light intensity of the

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_4

Table 4.1 Comparison of transcriptome interrogation methods

Method	Throughput abilities	Time required	Type of information	Advantages	Disadvantages
Microarray	Up to 20,000 genes	<24 h	Qualitative	Relatively cheaper, high throughput, quick turnaround, good for FFPE	Limited to gene expression detection, limited dynamic range, and relies on existing transcriptome definitions
NanoString	Up to 800 genes	<24 h	Quantitative	Automated, quick turnaround, able to detect gene fusions and miRNA, good for FFPE	Limited dynamic range and relies on existing transcriptome definitions
RNA sequencing	Targeted RNA-Seq to whole transcriptome	3–7 days	Quantitative	High dynamic range, agnostic to transcriptome definitions	Expensive, substantial computational support needed for data analysis. Use with FFPE still limited

fluorescently labeled, hybridized transcripts that act as a measure of relative or absolute transcript abundance in the sample [10]. With most microarrays, gene expression is a cumulative measure of all transcripts related to a gene, as microarrays tend to have difficulty in differentiating between transcript isoforms; however, specialized microarrays have been developed to capture these isoforms either by high-resolution genomic tiling microarrays or with probes that are complementary to the exon junctions [9]. In general, there are two approaches to gene expression using microarrays: one is to directly compare a sample to a control specimen by labeling each of them with a different fluorescent label and hybridizing them to the same microarray and then measuring the relative transcript abundance in the test sample as compared to that of the control one [9]. The second approach is to hybridize test and control samples separately, which requires normalization of gene expression, across all arrays used in an experiment. In general, microarrays have difficulty in detecting low-abundance transcripts and exhibit limited quantitation abilities (dynamic range) [9, 11, 12].

Regardless of the approach to gene expression analysis with microarrays used, the transcript abundance measurements will undergo bioinformatic analysis by classifier models to create a gene expression signatures which can be used to aid in the prediction of clinical end points such as tumor subtyping or treatment response. These microarray-based classifiers and associated gene signatures have been found to be highly robust with gene signatures showing 80–90% consistency across platforms [13]. Two gene expression microarray assays and their associated classifiers have been currently FDA approved, including the Tissue of Origin Test and MammaPrint [14, 15]. In addition to these two assays, there are a vast number of other commercially available, non-FDA-approved, microarray-based gene expression profilers. The development of microarrays and their classifiers has expanded our knowledge regarding intrinsic variation of expression between tissue types, during neoplastic transformation, as well as has aided in identification of molecular tumor subtypes [16, 17].

Extensive research and efforts from the Food and Drug Administration's (FDA) MicroArray Quality Control (MAQC) project and other groups have led to detailed descriptions of biases and variabilities associated with microarrays and their analysis [18, 19]. Their results have led to vast improvements in experimental design, best practice analysis strategies, and reproducible and accurate quality of microarray data [9, 18]. Preprocessing including removal of areas with poor signaling, image adjustment for size/shape of array grids, and normalization for differences in labeling efficiencies and RNA quality, are vital for accurate results [10]. Known sources of variability include dye-associated biases, differential hybridization efficiencies, labeling and amplification methods and biases related to sources, and preparation of array probes [10, 20]. Many of these variabilities, however, can be overcome through the use of proper control samples, technical and biological replicates, as well as dye-swapping experiments [10, 21]. However, several limitations still exist despite new methods to improve data quality. These limitations include reliance on existing transcriptome definitions, cross-hybridization effects leading to relatively high background levels, and a limited dynamic range for evaluating transcript abundance due to background levels and saturation of the reporter signals [22]. Despite these limitations, microarrays have been found to be generally precise, with cross-platform, and inter-/intra-site reproducibility, and represent a relatively inexpensive way to interrogate clinically relevant transcriptome profiles [18].

Digital Transcript Profiling

Technological advancements led to the development of digital analysis of transcripts, which allows for detecting each RNA transcript molecule individually, rather than detection of a combined signal from all unique transcripts hybridized to the same spot in a microarray. This approach allows for targeted, digital quantification of RNA expression in a high-throughput manner, assessing the expression of up to 800 transcripts at one time [23]. This approach is commercially available using the NanoString's nCounter® system, which has comparable sensitivity to microarrays and real-time PCR [24–26]. The system can use extracted total RNA as well as

crude lysates from whole blood and tissue from a variety of specimen types including FFPE to interrogate gene fusions, gene expression, as well as lncRNA and miRNA expression. The platform has been shown to work efficiently with as little as 100 ng input, even from specimens with highly degraded RNA such as FFPE [27].

Digital detection is possible due to a dual-probe system that consists of a reporter probe that uniquely tags each transcript molecule using a transcript-specific fluorophore barcode and a capture probe that allows the targeted transcript and hybridized molecular barcode to be immobilized and detected via a CCD camera and image analysis software. Overall, this platform has been shown to generate highly reproducible data, with low intra-sample technical and biological variability [27]; however, like microarrays, it is hampered by a limited dynamic range and relies on existing transcriptome definitions, limiting its ability to identify novel transcripts, alternative splicing isoforms, and fusion genes with novel partners [28]. This platform can be highly automated and has lower costs compared to other transcriptomic profiling methods; these characteristics as well as its ability to interrogate gene expression from poor-quality specimens in a multiplexed manner without amplification and with high sensitivity make this platform an attractive option for development of targeted clinical panels such as the FDA-cleared Prosigna® Breast Cancer Prognostic Gene Signature Assay [23, 24, 27, 29].

RNA Sequencing

With the development of high-throughput massively parallel next-generation sequencing (NGS), sequencing-based transcriptomics has now become routine and standard in biological and medical research and is rapidly gaining widespread acceptance in clinical practice. RNA sequencing (RNA-Seq) allows for digital quantification of whole or targeted transcriptome expression in a high-throughput manner from a variety of specimen types. Of the several advantages of RNA-Seq over existing microarray-based methods, the three most important relate to its unbiased nature with no a priori knowledge necessary of genome or transcript sequence, an extremely high dynamic range allowing quantification of both low- and high-expressed transcripts, and relatively low noise levels from cross-hybridization with genome sequences [22, 30]. With regard to applications, unlike microarrays, RNA-Seq-based methodologies provide not only digital quantitation of absolute and differential transcript expression but also sequence data readout at base-pair resolution allowing for identification of novel transcripts (including alternative and aberrantly spliced transcripts and noncoding RNA), expressed somatic variants in cancer, allele-specific expression, novel RNA editing events, posttranscriptional sequence changes (mutations or editing), as well as gene fusions [12, 31]. Together with the rapidly diminishing costs of NGS in general and the development of highly evolved tools for transcriptome analysis, RNA-Seq has firmly become the tool of choice in transcriptome research and clinical development.

RNA-Seq Library Preparation Protocols

"RNA-Seq" is a generic technical term used to refer to a variety of RNA and cDNA sequencing technologies, the two most commonly deployed techniques being mRNA sequencing on poly-A+ fraction of total cellular RNA and cDNA sequencing of total RNA fractions enriched by hybridization capture or amplification (Fig. 4.1). While an exhaustive discussion of all technical aspects is beyond the scope of this chapter, the key steps in RNA-Seq are described below. Presently, direct sequencing of RNA molecules is not routinely performed due to the high error rate of single-molecule sequencing technologies; therefore, we limit this section to the discussion of sequencing of cDNA libraries typically prepared from total RNA [22].

RNA Quality Assessment

Due to the inherent instability of RNA as a biological molecule, the quality and accuracy of RNA-Seq data are highly dependent on the quality of input RNA [22]. Assessment of

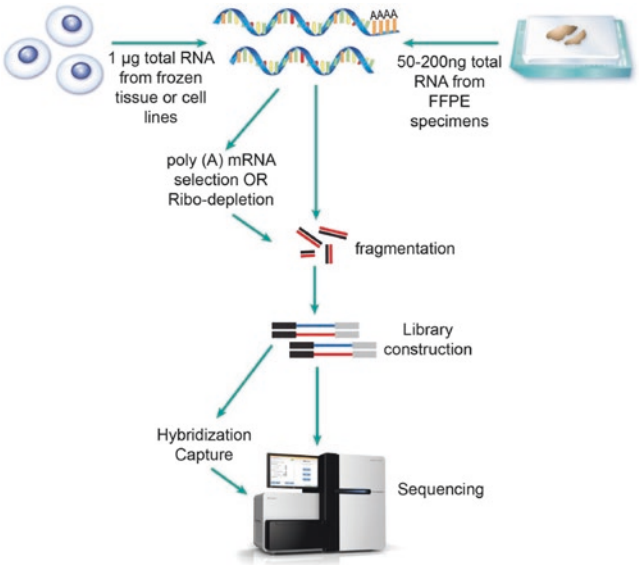


Fig. 4.1 Overview of RNA-Seq library preparation. RNA-Seq libraries are generated from RNA extracted from cell lines, frozen tissue, or FFPE specimens that have undergone RNA selection and fragmentation. Following library preparation, the transcriptome libraries are then sequenced using NGS platforms

RNA quality for RNA-Seq experiments is yet to be standardized, with several pre-analytic metrics of RNA integrity like the RNA integrity number (RIN), DV_{200} (% RNA fragments >200 nt) metric, quantitative PCR-based methods, as well as bioinformatic algorithms for assessment of mRNA quality from the RNA-Seq data itself, which have been described [32, 33]. While no one metric from these techniques has been shown to be highly accurate in the predicting success for generating a high-complexity RNA-Seq library, RNA with RIN scores >7 and DV_{200} values >50–60% is usually considered of "high" quality for RNA-Seq applications. The choice of a suitable RNA-Seq protocol is influenced by the quality of RNA; e.g., poly-A+ mRNA-Seq is impractical for low-quality and heavily fragmented RNA, as is often the case with FFPE and even frozen samples in clinical practice [32].

RNA input requirements for NGS libraries also vary by protocol; amplification-based targeted RNA-Seq libraries can be prepared from as little as 10–20 ng of total RNA, whereas whole transcriptome or coding transcriptome analysis requires higher inputs ranging from 40 ng to 1 ug of total RNA. RNA input levels are typically higher for selection methods that employ rRNA depletion or for positive selection of RNA fragments with intact poly-A tails, while hybridization capture-based selection methods generally require less RNA input (see Selection Method section for discussion).

RNA-Seq Selection Methods

Total RNA from biological specimens contains multiple RNA species including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal (rRNA), and other noncoding RNA (ncRNA), with rRNA accounting for approximately 85% of all RNAs in a cell [34]. Due to the overrepresentation of rRNA, it is necessary to deplete rRNA from extracted total RNA, as this is critical for the sensitivity of RNA-Seq in detecting biologically or clinically important RNA transcripts as well as alterations within these transcripts. Ribodepletion methods commonly used to effectively remove rRNA transcripts include polyadenylated (poly-A) enrichment and ribosomal RNA (rRNA) depletion [32]. Depending on the type of application, further strategies to selectively enrich for certain mRNA fractions, e.g., non-poly-A+ RNA fraction (small nucleolar RNAs, histone mRNAs, etc.) or the protein-coding transcriptome, usually utilize hybridization capture of cDNA generated by reverse transcription of RNA or amplification-based enrichment strategies (e.g., Anchored Multiplex PCR) [32, 35].

When considering the selection method, benefits and drawbacks should be carefully considered to determine the best approach for the sample type and experimental goal. As stated above, poly-A selected and rRNA-depleted libraries are the most common approaches to mRNA selection for RNA-Seq, both effectively excluding most rRNA transcripts. Poly-A selection uses oligo-dT beads, enriching for mature mRNA with intact 3′ poly-A tails, whereas rRNA depletion methods use hybridization capture probes specific for rRNA to remove these transcripts by bead-based subtraction, with both selection types having highly similar mRNA transcript sensitivity [35–37]. While poly-A selection is highly efficient at characterizing mRNA transcripts, due to the reliance on an intact 3′ poly-A tail, fragmented and degraded mRNA molecules with loss of poly-A tails show underrepresentation or 3′ bias in mRNA coverage by RNA-Seq [32, 35, 38]; this method is also not well-suited for non-polyadenylated RNA species [34, 35]. rRNA depletion, on the other hand, has more consistent coverage across transcripts and can be used for non-polyadenylated RNA species. However, rRNA depletion has been shown to be slightly less efficient regarding exonic coverage than poly-A selected libraries [35, 36, 38].

For a more targeted selection of the transcriptome, targeted DNA re-sequencing techniques to capture select regions for RNA-Seq have been developed. Two of the most common targeted selection methods include hybridization-based capture selection and amplicon-based transcriptome sequencing. Size ranges of targeted transcriptome sequencing panels can vary greatly, from a few genes to the entire known coding transcriptome [39]. Targeted RNA-Seq has been shown to be very efficient for highly degraded specimens such as FFPE, with hybridization capture performing better on degraded specimens than poly-A selected and rRNA-depleted libraries [39]. Hybridization-based capture and amplicon-based transcriptome sequencing have been shown to have similar genomic coverage and normalized gene expression measurements compared to poly-A selected transcriptomes yet often require less RNA input levels than poly-A selection [39, 40]. Further, these methods tend to also have higher sensitivity regarding variant calling in comparison to poly-A transcriptomes [39]. One specific targeted RNA-Seq enrichment method that has shown early promise in clinical diagnostics is based on Anchored Multiplex PCR (AMP™) methods. AMP™-based RNA-Seq has shown high sensitivity for gene fusion detection agnostic to fusion partner, in addition to simultaneous detection of single-nucleotide variants, insertions, deletions, and copy number changes [41].

Targeted sequencing allows for significant savings in cost, time, and computing power, reduced incidental findings, and has been used for the development of clinically useful panels [42, 43]. However, targeted sequencing, especially amplification-based enrichment methods, can introduce biases that compromise the complexity of the library and can result in underrepresentation of low-level transcripts. The clinical utility of whole transcriptome sequencing versus

targeted RNA-Seq is a current area of research being addressed by multiple groups, including the National Institute of Health's Clinical Sequencing Evidence-Generating Research (CSER2) program [44].

Fragmentation and Library Preparation

Fragmentation of RNA (or cDNA or dsDNA) is required prior to end repair and adaptor ligation for NGS sequencing library preparation. Specific protocols have been developed to retain mRNA strand information through the library preparation process, including the incorporation of dUTP during second-strand synthesis followed by uracil-DNA glycosylase (UDG) digestion. The stranded cDNA molecules then undergo typical NGS library preparation methods such as the end repair, A-tailing, and ligation of library adapters that include sample barcodes. After RNA selection and NGS library prep, the adapter-ligated cDNA fragments are then amplified and sequenced on an NGS platform in a high-throughput manner [22]. Given that complexity and diversity of cellular transcriptomes are highly variable and dependent among several factors, including tissue of origin and disease state, generating a complex RNA library with adequate representation of all RNA transcripts of interest, including low novel and unannotated transcripts, is a major challenge [9]. Library preparation and selection methods as well as the quality and quantity of input RNA are the principal contributors to library complexity, which can be assessed bioinformatically using duplication percentage as a measure of "uniqueness" of sequence reads. Low library complexity (high duplication rates) can be due to limiting amounts of RNA, heavily fragmented RNA, and use of excessive PCR cycles during library preparation. Several RNA-Seq library preparation protocols now incorporate molecular barcodes (or unique molecular identifiers) that tag RNA or cDNA molecules prior to the PCR steps and can be used for accurately assessing library complexity [45].

RNA-Seq Data Analysis

Due to the broad applications of RNA-Seq experiments, data analysis steps are multiple, often computationally intensive, and no one suite of software programs offers all of the available analytical tools (Fig. 4.2 and Table 4.2). The key processes common to all pipelines are quality control and read alignment. Depending on the nature of the experiment, specific programs are then incorporated into the pipeline to perform expression profiling and differential expression, somatic mutation and fusion detection, and alternative splicing, to name a few. A comprehensive assessment of RNA-Seq data analysis and best practice guidelines can be found elsewhere [46]; in the following section, some key steps and applications are discussed.

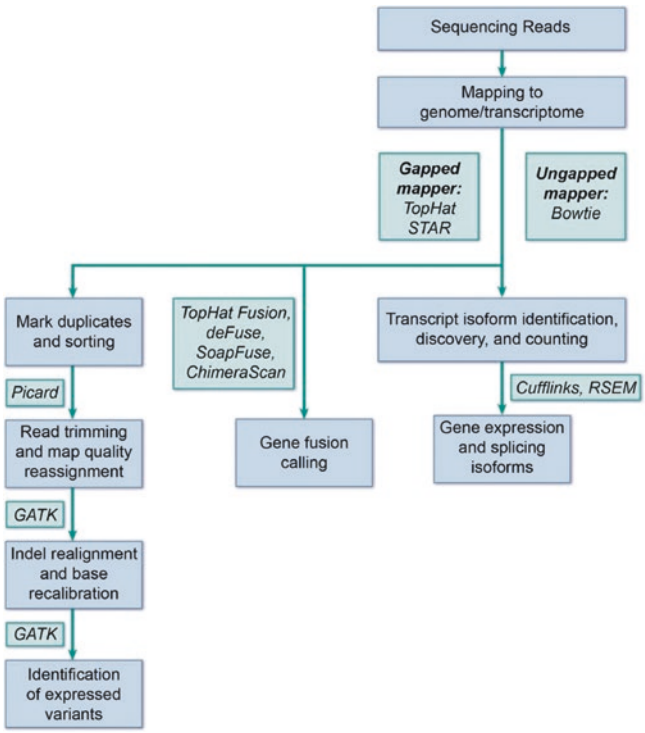


Fig. 4.2 RNA-Seq data analysis. Transcriptome sequencing informatics pipelines are highly dependent on the nature of the experiment. After transcriptome sequencing data is aligned to the human genome or transcriptome reference sequence, specific programs are integrated into the pipeline to perform specialized analysis for expression profiling and detection of somatic mutation, gene fusions, and alternative splicing events

Transcriptome Assembly

Sequence reads obtained from the most commonly used NGS platforms are often short and therefore need to be reconstructed into full-length transcripts, with the exception of short sequence length RNA classes such as miRNA [47]. Before assembly, raw sequencing reads are preprocessed to remove low-quality reads, PCR duplicates, adaptor sequences, sequencing errors, and other artifacts by tools such as FastQC, FASTX-Toolkit, and Trimmomatic [47–50]. In particular, sequencing errors are removed or corrected by using the quality score for each read, a probability function that a specific base in the sequence is correct, and/or the k-mer frequency, which is the number of times a short oligonucleotide of length k appears in a set of DNA sequences. Very low-frequency k-mers usually originate from sequencing errors, and reads containing these errors can be removed. However, this could remove real but very rare genuine transcripts [47].

Table 4.2 Bioinformatic software for transcriptome sequencing analysis

Analysis type		Software application
Quality control and preprocessing		FastQC [48]
		FASTX-Toolkit [49]
		Trimmomatic [50]
		Picard [61]
		RSeQC [63]
Alignment	Reference based	STAR [52]
		Bowtie2/TopHat [51, 177]
	De novo	Trinity [57]
		SOAPdenovo-Trans [58]
		Rnnotator [56]
		Trans-ABySS [59]
Transcript identification and quantification		Cufflinks [54]
		RSEM [55]
Differential gene expression		DESeq2 [77]
		edgeR [76]
		NOISeq [78]
Alternative splicing		CuffDiff2 [79]
		rMATS [81]
		DESeq2 [77]
Variant discovery		GATK [86]
Gene fusion discovery		deFuse [89]
		SOAPfuse [93]
		Chimera [94]
		TopHat-Fusion [90]
		PRADA [92]
		Fusion Hunter [91]

After preprocessing, transcriptome assembly can be achieved by either reference genome-based assembly or "de novo" assembly (Fig. 4.3) [46]. The reference-based method comprises two parts. First, reads are aligned to a reference genome or transcriptome with a "splice-aware aligner" such as TopHat or STAR [51–53]. Second, overlapping reads from the same locus are clustered in a graph to arrive at all possible isoforms, followed by analysis using programs such as Cufflinks or RSEM for transcript isoform resolution, discovery, and quantification [54, 55]. Reference-based methods require less computing power, eliminate some artifacts and errors as these would not align to the reference genome, and are very sensitive to rare transcripts. It is important to note that errors caused by the short-read aligners can carry over into assembly and that spliced reads spanning longer introns can be missed. "Multi-reads," where a sequence aligns equally well to several loci in the genome, can be excluded; however, this will leave gaps in the final sequence assembly [47].

"De novo" assembly does not use a reference genome (Fig. 4.3). Instead, it leverages the redundancy in the short reads and uses the overlaps to assemble the transcriptome with software tools such as Rnnotator, SOAPdenovo-Trans, Trans-ABySS, and Trinity [56–59]. Trinity was developed specifically for RNA-Seq data and prevents overlapping genes on the same strand being erroneously interpreted as fusion transcripts. It also groups related linear sequences that represent alternative isoforms or paralogous gene families into nonlinear structures containing "bubbles" and alternative ends. In general "de novo" strategies demand lots of computing power [57, 60]. Cloud computing is an alternative, and cloud-based genome assemblers have been developed [47].

Transcriptome Sequencing Quality Assessment

After assembly, it is essential to assess the quality of the resultant sequence and alignment [46]. Multiple programs exist for assessing sequencing quality including Picard (Broad Institute), RNA-SeQC (Broad Institute), FastQC (Babraham Institute), and RSeQC (Baylor College of Medicine) [48, 61–63]. Sequencing QC metrics should be used to regularly assess sequencing quality for each run as well as for determining precision rates for each sequencing platform and library preparation kit to aid in the detection of instrumentation issues and lot-to-lot variability in library preparation and sequencing kits. The Genetic European Variation in Disease (GEUVADIS) consortium has recommended several parameters to be assessed as quality control including distribution of base quality scores and GC content, mapping rate, standard deviation of insert size, and coverage distribution across the transcript [64]. Of these, one of the most important parameters is the percentage of mapped reads, which tends to be a general indicator of the overall sequencing accuracy. Low mapping rates, below 70% when mapping to the human transcriptome, may be an indication of DNA contamination [52]. Another important parameter, outside of those recommended by GEUVADIS, is the percentage of uniqueness with low levels possibly indicating low library complexity which can be resultant from poor RNA quality. Sensitivity of RNA-Seq, like all next-generation sequencing techniques, is highly dependent on library complexity and the depth of sequencing. Library complexity, as stated above, can be quantified by quality control (QC) software, such as RNA-SeQC, which details the unique rate of mapped transcripts based upon unique read start sites [62]. In our experience, a unique RNA rate of approximately 40% or higher tends to indicate high-quality NGS libraries. Unfortunately, determining the optimal depth of sequencing needed for RNA-Seq is not well defined. As RNA expression is variable and can span several orders of magnitude, the depth of sequencing needed to detect the transcripts of interest inversely correlates to the natural levels of transcript expression in the sample [37, 65]. Highly expressed transcripts will be more easily detected than transcripts with moderate- to low-level expression and thus require less sequencing depth. As low to moderate expressed transcripts

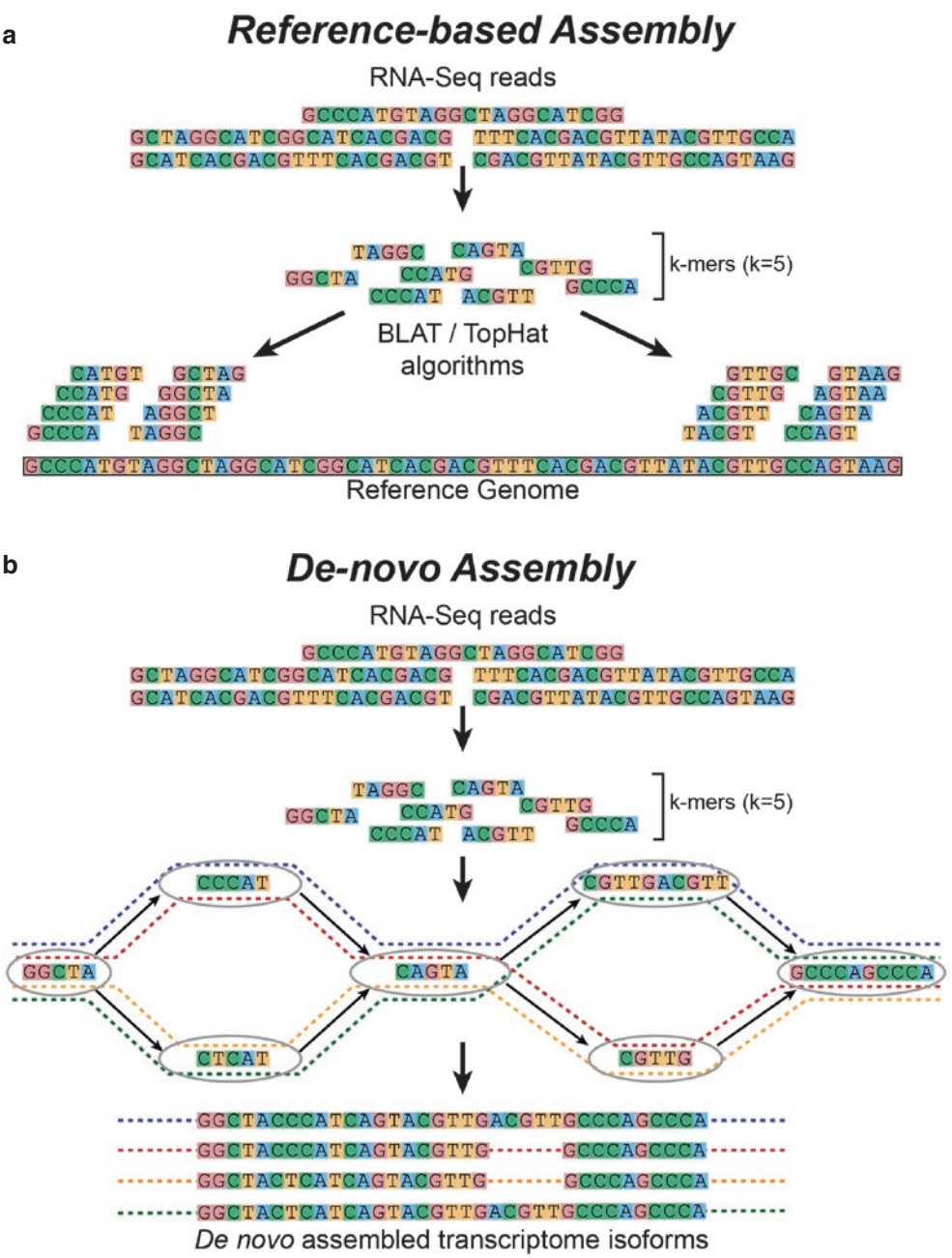


Fig. 4.3 Transcriptome assembly. RNA sequencing reads can be assembled by either reference-based (**a**) or de novo assembly methods (**b**). Reference-based assembly methods divide RNA-Seq reads into substrings (k-mers), mapping these smaller fragments to a reference genome to create an assembled transcriptome. De novo assembly methods, similar to reference-based assembly, divide the RNA-Seq reads into k-mers; however, following generation of the k-mers, these substrings are then organized based on the presence of at least one overlapping base in adjacent k-mers and assembled into a de Bruijn graph to determine all possible combinations of isoform sequences

represent a smaller proportion of the full population of transcripts, detection of the lower expressed transcripts will require higher sequencing depths [37, 66]. Insufficient depth of sequencing can lead to erroneous gene expression and variant allele frequency metrics as well as an inability to detect low expressed transcripts, gene fusions, and variants. Currently, there are no standards or guidelines regarding the depth of sequencing required for high confidence in detecting the full representation of the transcriptome or for targeted applications. The use of cell-mixing studies and synthetic spike-ins using true-positive variants and gene fusions with a wide range of expression rates will help to inform on optimal coverage needed for each assay.

Reference Material and Quality Control

Given that RNA-Seq is an evolving technology with improvements constantly in flux regarding advances in instrumentation, sequencing chemistries, new library construction methodologies, and computational analysis, in order for RNA-Seq to be clinically viable, there is an imminent need for the development of standards, "best practice" guidelines, and reference materials to ensure maintenance of quality between sequencing runs. While reference standards for DNA have become more widely established, standards for RNA have lagged behind due to the complexity and diversity of the transcriptome as well as due to the increased variation in RNA

sample quality, different RNA selection and library preparation methods, and requirement for more highly complex bioinformatic analysis [67]. One of the first RNA reference materials developed for microarrays and transcriptome sequencing was the universal human reference RNA samples comprised of RNA derived from an equimolar mixture of multiple cell lines. These reference RNA samples have been developed and tested by many consortiums including SeQC and the Association of Biomolecular Resource Facilities (ABRF), with resultant data available for download from the Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) repository for use as quality control [31, 37, 68, 69]. However, one major limitation of using biological human genetic material such as the universal reference RNA samples is the inability for these controls to be directly combined with patient samples without contaminating downstream analyses [67]. Being able to combine a reference standard directly with patient samples provides the added benefit of having the control concurrently undergo all library preparation and sequencing steps as the clinical specimen, thereby acting as an internal control that accurately reflects all processes and technical variabilities associated with each patient sample. Exogenous spike-in controls serve as a good alternative to the human reference RNA samples as these controls often comprise of non-human or artificial sequences allowing the derivative reads to be distinguished from reads derived from the patient sample, thus allowing these controls to be spiked into each patient's RNA specimen [67]. Further, the use of exogenous RNA standards rather than endogenous transcripts such as "housekeeping genes" provides a more reliable standard due to the identical nature and constant expression across samples [70].

The External RNA Controls Consortium (ERCC) was formed to develop external RNA controls for use in the evaluation of technical performance of gene expression assays including microarray and RNA-Seq. These external controls were developed to aid in determining the accuracy, reproducibility, and limits of detection of transcriptomic profiling assays, as well as to serve as a standard measurement for sequencing error rates, coverage biases, and transcript quantification [70]. The effort by ERCC has led to the development of external RNA standards known as the ERCC spike-ins that consist of two pre-formulated sets of 92 polyadenylated transcripts from the ERCC plasmid reference library generated from the genomes of *Bacillus subtilis* and *Methanocaldococcus jannaschii*, as well as from synthetic DNA [70, 71]. Large studies including the SeQC and ABRF next-generation sequencing projects have utilized the ERCC spike-ins to assess sensitivity and precision of RNA-Seq, and these have been ultimately recommended by ABRF as a useful sample-based quality metric [31, 37, 70]. Outside of these efforts, multiple labs have generated custom synthetic RNA spike-in controls to assess sensitivity of RNA-Seq in regard to the detection of transcript isoforms, small RNA classes, and clinically relevant alterations including gene fusions [70, 72–74]. These synthetic spike-ins generated to model human genetic variation can be important in assessing the sensitivity of RNA-Seq in regions which are often hard to characterize by NGS technologies, such as in GC-rich areas. Many genomic alteration types can be also multiplexed within a single set of RNA spike-ins enabling a variety of alteration types to be appraised at once. Despite all the benefits related to external spike-in controls, there is a constant challenge for these reference controls to maintain commutability between different library preparation methods and sequencing platforms as the synthetic constructs do not always perform similarly to that of native RNA transcripts, as demonstrated in the ABRF study in which ERCC spike-ins performed better in ribo-depleted libraries than in poly-A libraries [31, 37]. It is important to note that the development of RNA-Seq reference standards is an ongoing process, as evidenced by the recently launched ERCC 2.0 project, to begin the development of updated RNA spike-in controls including improved mRNA mimics, as well as novel cancer gene fusion and small RNA controls [75].

Gene Expression and Alternative Splicing Detection

After sequencing quality assessment, downstream analysis for transcript quantification and identification of splicing isoforms, variants, and gene fusions can be initiated. One of the most common applications of transcriptome sequencing is related to the estimation of transcript expression levels. Transcript expression from RNA-Seq is quantified by aggregating the total number of reads that align to each transcript followed by normalization based on library size and feature lengths, with transcript expression expressed in terms of "expected fragments per kilobase of transcript per million fragments mapped" (FPKM) or "reads per kilobase per million mapped reads" (RPKM) [9]. Specialized programs such as the Cufflinks suite and RSEM have developed sophisticated algorithms that aid in accurate assignment of reads mapping to more than one transcript by the use of an expectation-maximization approach which accounts for particular biases associated with nonuniform read distribution along the length of the gene [54, 55]. Once the transcripts have been quantified and normalized, data can then be used to assess differential gene expression signatures among different samples, which are typically computed using discrete probability distributions such as the Poisson or negative binomial. Popular methods for differential expression analysis from RNA-Seq data include edgeR, DESeq2, and NOISeq [76–78]. Transcript-level differential analysis can also detect alternative splicing events by comparing the expression of different transcript isoforms. Programs developed for the detection of alternative splicing events include CuffDiff2, DSGSeq, and rMATS, all of which use different algorithms to

identify alternative isoforms [79–81]. CuffDiff2, part of the Cufflinks suite, integrates estimation of transcript abundance and differential isoform expression analysis, with built-in methods for sequence bias correction and use of a beta-negative binomial distribution to assign fragments to each isoform [79]. DSGSeq compares read counts at exons and their junctions for the detection of significant differences in isoform abundance, while rMATS identifies exon junctions present within the reads to detect differential exon splicing events [80, 81]. Exon- or junction-based methods tend to have higher accuracy in identifying specific alternative splicing events; however, all of these approaches tend to be generally hampered by the intrinsic limitations of short-read sequencing [82].

Multiple studies have shown strong concordance of normalized transcript expression between RNA-Seq and microarrays, with some studies attributing RNA-Seq with higher sensitivity [9, 83]. However, it should be noted that efforts by the FDA and their SeQC project as well as by the ABRF have shown that absolute gene expression measurements from RNA-Seq are unreliable, while the relative gene expression measurements are accurate and reproducible across sites and platforms [31, 37]. Further, RNA-Seq is susceptible to biases common to NGS sequencing, such as GC content, positional biases, batch effects, background noise, and other sequence biases, in addition to biases associated with transcript selection methods [31, 84]. Many preprocessing filters and analytical methods have been developed to reduce background noise, though it's important to note these often improve accuracy at the expense of precision [31]. R packages such as NOISeq R have been developed to aid in identification of biases specific to individual datasets and assist in normalization of the data in accordance [78]. The use of multiple biological replicates in all gene expression and isoform detection experiments is highly encouraged as often data from small sample sets can be noisy. Further, studies have shown that the choice of method and even the software version can markedly affect the analysis results and no single methodology may serve as a best fit for all datasets [46]. Therefore, it is important to thoroughly document the software, its version, and settings used for future analysis use.

Variant and Gene Fusion Detection

Utilizing the advancements of next-generation sequencing, RNA-Seq allows for nucleotide-level resolution in the detection and expression quantification of small to large disease-associated alterations such as single-nucleotide variants, insertions, deletions, and gene fusion events and determines allele-specific expression [46]. Notably, RNA-Seq can also capture variants that are sometimes difficult to detect at the genome level, such as translocations or splicing events. Detection of the presence and/or expression of clinically relevant single-nucleotide variants and gene fusions represents one of the most readily translatable avenues for clinical transcriptome sequencing, especially as treatment strategies become ever more dependent on molecular data and the presence of specific genetic alterations in the diseased tissue. In fact, targeted RNA-based NGS panels designed specifically for the detection of variants and gene fusions have already begun to be integrated into current clinical practice. Due to the complexity of RNA-Seq analysis, different "best practice" analysis workflows have been developed for each variant type [85].

Detection of single-nucleotide variants (SNVs) and small indels in RNA-Seq data has similar workflows to those used in DNA sequencing, characterized by alignment to the human genome/transcriptome and pre-preprocessing steps such as read trimming, followed by variant calling using variant callers typically used for DNA sequences. Of note, some variant callers, such as the HaplotypeCaller from GATK, have developed RNA-Seq-specific modes in order to increase specificity by combating erroneous variant calls associated with difficulty in resolving splice junction sites [86]. Detection of large deletions in RNA-Seq data is currently much more challenging due to difficulty in local alignment in RNA-Seq data as well as complications due to exon splicing events which can lead to a large degree of false-positive calls. Despite some of the bioinformatic challenges associated with variant calling in RNA-Seq data, RNA pipelines can exhibit high sensitivity in the detection of variants yet often require highly stringent, specialized variant filtering in order to increase specificity [85, 87].

Another variant class that can be detected by RNA-Seq is gene fusion events. Gene fusions are well-known drivers in cancer and can often serve as diagnostic or predictive biomarkers for the disease [88]. RNA-Seq has been shown to be highly sensitive for the detection of fusion events. Discovery of gene fusions by RNA-Seq, while analogous to novel isoform discovery, is complicated by the fact that transcript segments often map to multiple chromosomes or are separated by large distances on the same chromosome. As such, specific programs have been developed for the detection of gene fusions in RNA-Seq data, including deFuse, TopHat-Fusion, Fusion Hunter, PRADA, Chimera, and SOAPfuse [89–94]. Each algorithm has its own benefits, with some tools outperforming others in regard to sensitivity, positive prediction value, time consumption, and memory usage [95]. Further, the performance of each program can also depend on the sample quality, read length, quality of the reads, and total number of sequencing reads achieved [95, 96]. Artifacts are also highly common, resulting from misalignment of reads due to homology, sequencing errors, and polymorphisms [46]. Therefore, similar to RNA-Seq SNV calling pipelines, post-processing methodologies using heuristic filters are required for maximizing specificity.

One major benefit of RNA-Seq for the detection of gene fusions over more targeted clinical assays such as reverse transcription polymerase chain reaction (RT-PCR) is the added benefit of agnostic detection of alterations in a high-throughput manner. RT-PCR assays are highly sensitive yet specific for a single alteration with multiple individual reactions required when testing for several fusion partners or various possible breakpoints, which can be a large drawback for genes that are known to have a highly diverse set of secondary gene partners that generate multiple different fusion transcripts, such as ALK, KIT, and ROS1 in lung cancer [97]. Moreover, fluorescent in situ hybridization (FISH), an alternative method to RT-PCR for detection of translocations, is able to detect rearrangements featuring the targeted gene irrespective of the secondary gene yet is unable to identify the specific secondary gene partner. RNA-Seq represents an advancement over these two methodologies through its high-throughput, agnostic nature while still maintaining a similar degree of sensitivity [97].

Currently, RNA-Seq analysis is associated with complex bioinformatic analyses, often limiting the degree of adopters to the technology. To reduce some of the bioinformatic burden associated with RNA-Seq, "plug and play" user-friendly analysis programs have been developed that can make it more accessible for the typical clinical lab. For example, Illumina and Ion Torrent have integrated a wide range of plug-ins in their BaseSpace and Torrent Browser applications, respectively, for the detection of variants, gene fusions, and differential gene expression in a user-friendly manner requiring very little informatics skills. Other companies, such as ArcherDx and Asuragen, have developed easy-to-use bioinformatic analysis suites that utilize graphical user interfaces as companions to their targeted RNA NGS panels in order to make the informatics analysis more accessible to labs. This trend is expected to continue as more and more companies begin to develop NGS panels for clinical use. However, for clinical applications, it is important to understand the basic concepts and assumptions from each of these tools, as they can have a significant impact in the sensitivity and specificity of each clinical test. Clinical laboratories that implement these out-of-the-box tools, as well as those that develop their own pipelines, should adhere to the guidelines from the Association for Molecular Pathology (AMP) and the College of American Pathologists (CAP) on the development and validation of bioinformatic pipelines for NGS [98].

Clinical Applications of RNA-Seq

Clinical gene expression profiling was developed in the last decade, with applications for cancer diagnosis, cancer prognosis, and transplant rejection detection, some of which have obtained the US Food and Drug Administration (FDA) clearance [99–103]. However, despite great promise, widespread use of transcriptome profiling for chronic inflammatory, neurological, and infectious diseases, for example, has not become reality [104–106]. Aided by the rapidly decreasing cost of data generation, and the ongoing technical advances of NGS, RNA-Seq has the potential to become a powerful tool in the management and treatment of human disease, although this technology is not yet in widespread clinical use [107, 108]. In the paragraphs below, we discuss examples of RNA-Seq applications for different clinical scenarios.

Inherited Conditions

Since the completion of the human genome, there has been a steady increase in the identification of genes responsible for monogenic conditions [109–111]. NGS has been successfully employed for DNA-based diagnostic assays such as disease-specific panels and whole-exome/whole-genome approaches [112–114]. RNA-Seq is being used to study phenotypic variation among individuals affected with genetic diseases and for diagnosis in cases where DNA-based sequencing and deletion/duplication analyses are unsuccessful. In a recent example of the possible contribution of RNA-Seq to clinical diagnostics, Chandrasekharappa and collaborators showed that the addition of RNA-Seq to NGS-based DNA sequencing and array comparative genomic hybridization (aCGH) allows the detection of more disease alleles in patients with Fanconi anemia (FA) [115]. The use of RNA-Seq allowed the identification of exon skipping associated with synonymous, missense, and nonsense mutations, as well as intronic pathogenic mutations in FA genes. RNA-Seq has also proved to be an important tool to improve our understanding of complex phenotypes in multigenic disorders, as seen in Down syndrome, for example [116]. In a recent study by Costa et al., RNA-Seq was performed in human trisomic endothelial progenitor cells, revealing differential expression of genes expressed at low levels, novel regions of active transcription outside known loci, identification of non-polyadenylated long and short noncoding RNAs, and identification of novel splice isoforms and novel extended untranslated regions for known genes which could represent novel miRNA targets or regulatory sites for gene transcription [116]. This approach could help better understand the mechanisms involved in the generation of Down syndrome phenotypes and the observed individual variability.

Complex Conditions

Genome-wide association studies (GWAS) linking SNPs with specific phenotypes of complex traits and common diseases have shown that only a small fraction of associated

SNPs falls within coding regions and that most are intronic or intergenic [117]. This suggests that these nucleotide variants affect gene expression rather than protein function [118]. These variants, therefore, are known as expression quantitative trait loci (eQTL). Given that RNA-Seq allows for the integrative analysis of variation in allele-specific transcript sequence, regulatory sites, and expression levels, it is expected that its use will help better understand regulatory variation at single-base resolution and this could translate into better insight into the molecular pathogenesis of complex traits [119]. A recent example of success with this approach, although not done with RNA-Seq, is the demonstration that a common noncoding polymorphism at the 1p13 locus, which was found to be associated with changes in plasma low-density lipoprotein cholesterol isoforms and risk for myocardial infarction, affects the expression of a gene that regulates lipoprotein production by the liver [120].

Precision Oncology

The impact of RNA-Seq has arguably been most significant in the field of cancer biology and clinical oncology. From the discovery of several novel gene fusions as oncogenic drivers in different tumor types to the identification of expression signatures predictive of response of immunotherapy, RNA-Seq-based studies have had a transformative influence in our understanding of cancer [121–126]. Increasingly this has led to attempts to harness the power of RNA-Seq for clinical use in precision oncology. While molecular diagnostics in cancer remains heavily reliant on DNA sequencing-based approaches, RNA-Seq offers several complementary advantages in what we now call personalized medicine, namely, the ability to detect structural variations leading to targetable gene fusions that are mostly undetected by standard targeted DNA-sequencing-based approaches (including whole-exome sequencing) and the ability to assess the tumor microenvironment and predict a response to immunotherapy and assess expression levels of oncogenic mutations. By virtue of providing both a structural snapshot of genomic aberrations and a dynamic picture of cellular processes through expression analysis, RNA-Seq offers an unparalleled functional overview of the cancer genome. Multiple recent large genomic sequencing trials in cancer have incorporated RNA-Seq into DNA-based diagnostics with promising results such as the iCat and MI-ONCOSEQ trials, and several others are underway, e.g., NCI-MATCH and Texas KidsCanSeq [127–131].

Of the several uses of RNA-Seq in cancer diagnostics, none is more readily apparent than the ability to detect fusion genes in an unbiased manner. Because RNA-Seq provides full transcript sequences and is capable of assembling transcripts without relying on preexisting reference sequences, it has been used for the detection of fusion transcripts, both for those previously known to be associated with specific tumors and for previously unidentified fusions [107]. Although recurrent gene fusions have been well documented in hematologic malignancies and sarcomas for several decades, the discovery of recurrent gene fusions in epithelial solid tumors is relatively recent. In 2005, Tomlins et al. reported the discovery of fusion transcripts between the TMPRSS2 and the ETS transcription factor genes in prostate cancer [132]. This discovery transformed our understanding of solid tumors and opened the door to an avalanche of studies reporting recurrent fusions in a variety of tumor types [133–139]. Some of these discoveries, such as the identification of ALK rearrangements in lung cancer, have already been incorporated into the diagnostic algorithms and management strategies of cancer patients [140, 141]. RNA-Seq-based assays, both whole transcriptome and targeted RNA-Seq panels, are being used to detect these fusion transcripts in solid tumors and hematologic malignancies alike. It is likely that the number of reported fusion transcripts in human cancers will continue to rise as the use of RNA-Seq becomes more prevalent. As in the case of lung and thyroid cancer, some of these fusion transcripts might be useful for diagnosis, prognosis, or selection of targeted therapies. Detection of fusion transcripts has rapidly become routine in the practice of pathology [97, 142].

Another use of transcriptome profiling is the identification of tissue origin in cancers of unknown primary site. Several commercial platforms are clinically available for this purpose [99, 143, 144]. Although this application has not yet been migrated to RNA-Seq platforms, it is foreseeable that this will occur in the near future, because this technology enables not only the identification of the site of origin based on the expression pattern but also the detection of expressed mutations and alternative splicing events that could be of utility for therapy selection [145].

Breast cancer, lung cancer, thyroid cancer, melanoma, and prostate cancer management are other areas where transcriptome profiling has been incorporated into routine clinical management, with the use of gene expression profiles that determine the likelihood of a cancer diagnosis, are prognostic of tumor recurrence, or are used for therapeutic management in early-stage patients [100, 142, 146–153]. While most of these tests use either quantitative reverse transcriptase PCR (RT-PCR) or microarrays, efforts to translate some of these panels into RNA-Seq platforms are already underway [154]. This has the potential to expand the content of these panels and to incorporate therapeutic biomarkers in these prognostic tests [101, 154].

The utility of RNA-Seq is best appreciated when integrated with multimodal genomic profiling. When performed with whole-exome sequencing or targeted sequencing, RNA-Seq offers the ability to evaluate the functional effect of splice variants, exon skipping, and aberrant splicing. When combined with copy number assessment, RNA-Seq allows

for the correlation between gain-of-function alterations such as high-level amplifications and upregulation of expression and uncovering loss-of-function alterations such as monoallelic expression in the context of promoter silencing.

In summary, basic and clinical research with RNA-Seq in oncology is providing us with a treasure trove of information that should allow us to better understand tumor initiation, progression, and resistance to therapy. The complexity of cancer has been made evident by recent research efforts, and it is now clear that understanding cancer biology and our ability to personalize treatment and to impact outcomes will require the use of all available "omics" technologies, because not all molecular alterations that drive tumor behavior are detected by a single approach [155]. Examples of this are pediatric tumors, in which mutations are less frequent than in adults and where transcriptome analyses have recently identified dysregulated genes that might uncover new targeted therapeutic approaches [156].

Clinical Microbiology Applications

The advent of NGS has made it possible to study and identify a large number of microbial populations in humans and to start defining the normal microbiome as well as microbiome changes associated with abnormal states [157, 158]. The rapid development of microbiome studies mostly relies on the use of 16S rRNA gene sequencing, which is based on DNA sequencing and has become a transforming force in clinical microbiology [159]. However, RNA-Seq technology has created new opportunities for the study of bacterial gene expression [160, 161]. One of the advantages of the RNA-Seq approach is the possibility to study unculturable bacteria or bacteria that cannot be isolated [162]. As such, transcriptome analysis by RNA-Seq has been applied to various clinically relevant microorganisms including *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Mycoplasma pneumoniae*, *Listeria monocytogenes*, *Helicobacter pylori*, *Salmonella typhi*, *Vibrio cholerae*, *Chlamydia trachomatis*, and *Bacillus anthracis*, among others [161]. Whereas RNA-Seq offers several advantages over prior technologies, prokaryotic RNA biology poses specific new challenges for this technology [163]. These include the absence of a poly-A tail (which allows for easy retrieval of coding RNA in eukaryotes), the highly unstable nature of bacterial RNA, and the fact that up to 50–80% of bacterial RNA preparations are composed of ribosomal rRNA and tRNA [164–166]. However, sequencing-based microbial transcriptome studies have been made possible by removing, at least partially, the rRNA and/or tRNA through a variety of extraction methods including r/tRNA depletion through hybridization with magnetic bead-linked complementary oligonucleotides or the use of terminator exonucleases involved in specific degradation of transcripts with a 5′ monophosphate group [163, 166–168]. These bacterial RNA-Seq studies have contributed to a more refined understanding of bacterial gene expression and its impact on microbial ecology and physiology and ultimately its potential use in clinical settings [162, 163, 169].

Perhaps one of the most important aspects in the clinical practice of microbiology is the ability to predict or assess microbial virulence and pathogenicity [161, 162]. One of the most important findings in this area is the identification of a larger number of untranslated regions (UTRs) in bacterial transcripts [170]. These UTRs contain riboswitches and binding sites of regulatory small RNAs (sRNAs) and are likely involved in the regulation of gene expression in bacteria, including the expression of genes related to pathogenicity [170]. RNA-Seq experiments have discovered that sRNAs account for up to 20% of bacterial RNA, including antisense RNAs, and these sRNAs appear to have regulatory roles [162, 169, 171]. In a study by Perkins et al. focused on *Salmonella typhi*, strand-specific cDNA sequencing (ssRNA-Seq) was used to identify transcriptionally active genes, revealing a large number of previously unknown transcribed regions, including novel noncoding RNAs, some of which might impact the expression of virulence genes [167]. Sharma et al. obtained similar results in transcriptome profiles of *Helicobacter pylori* and were able to establish a correlation between the size of 5′ UTRs and cellular function, concluding that UTR size correlated with pathogenicity [172]. Interestingly, the use of "dual RNA-Seq," in which both pathogen and host RNA are sequenced together, has revealed important insights about gene expression in infected tissues that were not possible by using experimental models [173, 174].

As detailed above, RNA-Seq is starting to provide an in-depth view of pathogen transcriptomes, and this research is predicted to have a direct impact not only on clinical diagnostics and epidemiology but also in the future progress of the field of pathogenomics. Nevertheless, many challenges remain. Single-molecule technology could allow sequencing of full-length polycistronic transcripts, which are commonly found in bacteria, and uncover how alternative transcription origins are utilized and regulated [162]. Probably the biggest hurdle for clinical implementation of RNA-Seq in the infectious diseases arena is the lack of reliable clinically based genotype-phenotype correlations that will enable clinical decision-making based on bacterial expression profiles [169].

Conclusions and Future Directions

RNA sequencing is making it possible to study transcriptomes at unprecedented resolution and with the ability to detect previously unknown noncoding and fusion tran-

scripts. This technology is currently being applied to the study of inherited, neoplastic, and infectious disorders. Results from these transcriptome analyses are increasing our understanding of normal and disease processes, and it is expected that this new knowledge will translate into clinical applications in the near future. Certainly, this technology has moved to real clinical utility in oncology, with the development of RNA-Seq assays to detect common fusion transcripts in hematologic malignancies and lung and thyroid cancer which inform clinical management [97, 142, 175]. Ongoing research is likely to increase the development of new applications as RNA sequencing now has been incorporated in several ongoing clinical trials in patients with advanced cancers [127–129]. The future use of RNA sequencing in precision oncology can be illustrated by its use in a patient with acute lymphoblastic leukemia followed by success in the use of specific targeted therapy, based on the transcriptome profile [176]. In addition, RNA-Seq has the potential to improve the diagnosis of genetic diseases and to reveal important clues about bacterial pathogenicity and genes important for pathogen/host interactions. It is safe to predict that the use of RNA-Seq will continue to increase in the practice of clinical and anatomic pathology.

Acknowledgments The authors would like to thank Karen Prince of Texas Children's Hospital for her help with the design of the figures for this chapter.

References

1. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77–94.
2. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87–98.
3. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet. 2016;17(5):257–71.
4. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA Cancer J Clin. 2016;66(1):75–88.
5. Rodriguez SA, Impey SD, Pelz C, Enestvedt B, Bakis G, Owens M, et al. RNA sequencing distinguishes benign from malignant pancreatic lesions sampled by EUS-guided FNA. Gastrointest Endosc. 2016;84(2):252–8.
6. Li X, Quigg RJ, Zhou J, Gu W, Nagesh Rao P, Reed EF. Clinical utility of microarrays: current status, existing challenges and future outlook. Curr Genomics. 2008;9(7):466–74.
7. Fedorowicz G, Guerrero S, Wu TD, Modrusan Z. Microarray analysis of RNA extracted from formalin-fixed, paraffin-embedded and matched fresh-frozen ovarian adenocarcinomas. BMC Med Genet. 2009;2:23.
8. Coudry RA, Meireles SI, Stoyanova R, Cooper HS, Carpino A, Wang X, et al. Successful application of microarray technology to microdissected formalin-fixed, paraffin-embedded tissue. J Mol Diagn. 2007;9(1):70–9.
9. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 2011;9:34.
10. Jaluria P, Konstantopoulos K, Betenbaugh M, Shiloach J. A perspective on microarrays: current applications, pitfalls, and potential uses. Microb Cell Factories. 2007;6:4.
11. Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, Terry PA, et al. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR -- how well do they correlate? BMC Genomics. 2005;6:59.
12. Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucleic Acids Res. 2002;30(10):e48.
13. Fan X, Lobenhofer EK, Chen M, Shi W, Huang J, Luo J, et al. Consistency of predictive signature genes and classifiers generated using different microarray platforms. Pharmacogenomics J. 2010;10(4):247–57.
14. Dumur CI, Fuller CE, Blevins TL, Schaum JC, Wilkinson DS, Garrett CT, et al. Clinical verification of the performance of the pathwork tissue of origin test: utility and limitations. Am J Clin Pathol. 2011;136(6):924–33.
15. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med. 2016;375(8):717–29.
16. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503–11.
17. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–52.
18. Consortium M, Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006;24(9):1151–61.
19. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001;29(4):365–71.
20. Ma C, Lyons-Weiler M, Liang W, LaFramboise W, Gilbertson JR, Becich MJ, et al. In vitro transcription amplification and labeling methods contribute to the variability of gene expression profiling with DNA microarrays. J Mol Diagn. 2006;8(2):183–92.
21. Schulze A, Downward J. Navigating gene expression using microarrays–a technology review. Nat Cell Biol. 2001;3(8):E190–5.
22. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
23. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. Curr Protoc Mol Biol. 2011;94;Chapter 25:Unit25B 10.
24. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26(3):317–25.
25. Richard AC, Lyons PA, Peters JE, Biasci D, Flint SM, Lee JC, et al. Comparison of gene expression microarray data with count-based RNA measurements informs microarray interpretation. BMC Genomics. 2014;15:649.
26. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hossein A, et al. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. BMC Res Notes. 2009;2:80.
27. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer. 2014;14:177.
28. Prokopec SD, Watson JD, Waggott DM, Smith AB, Wu AH, Okey AB, et al. Systematic evaluation of medium-throughput mRNA abundance platforms. RNA. 2013;19(1):51–62.

29. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordonez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC Biotechnol. 2011;11:46.
30. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621–8.
31. Consortium SM-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol. 2014;32(9):903–14.
32. Cieslik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2018;19(2):93–109.
33. Feng H, Zhang X, Zhang C. mRIN for direct assessment of genome-wide and gene-specific mRNA integrity from large-scale RNA-sequencing data. Nat Commun. 2015;6:7816.
34. Morlan JD, Qu K, Sinicropi DV. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. PLoS One. 2012;7(8):e42882.
35. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics. 2014;15:419.
36. O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. Curr Protoc Mol Biol. 2013;103;Chapter 4:Unit 4 19.
37. Li S, Tighe SW, Nicolet CM, Grove D, Levy S, Farmerie W, et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. Nat Biotechnol. 2014;32(9):915–25.
38. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dokel S, Ralser M, et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics. 2014;15:675.
39. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome Res. 2015;25(9):1372–81.
40. Li W, Turner A, Aggarwal P, Matter A, Storvick E, Arnett DK, et al. Comprehensive evaluation of AmpliSeq transcriptome, a novel targeted whole transcriptome RNA sequencing methodology for global gene expression analysis. BMC Genomics. 2015;16:1069.
41. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med. 2014;20(12):1479–84.
42. Levin JZ, Berger MF, Adiconis X, Rogov P, Melnikov A, Fennell T, et al. Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. Genome Biol. 2009;10(10):R115.
43. Nikiforov YE, Yip L, Nikiforova MN. New strategies in diagnosing cancer in thyroid nodules: impact of molecular markers. Clin Cancer Res. 2013;19(9):2283–8.
44. National Human Genome Research Institute. Clinical Sequencing Evidence-Generating Research (CSER2) 2017 [updated 08/08/2017]. Available from: https://www.genome.gov/27546194/clinical-sequencing-exploratory-research-cser/.
45. Kivioja T, Vaharautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9(1):72–4.
46. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13.
47. Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011;12(10):671–82.
48. Babraham Bioinformatics. FastQC 2010 [updated 01/10/2018]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
49. Hannon GJ. FASTX-Toolkit 2009 [updated 02/02/2010]. Available from: http://hannonlab.cshl.edu/fastx_toolkit/index.html.
50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
51. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
53. Engstrom PG, Steijger T, Sipos B, Grant GR, Kahles A, Ratsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods. 2013;10(12):1185–91.
54. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
55. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
56. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, et al. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics. 2010;11:663.
57. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
58. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;30(12):1660–6.
59. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7(11):909–12.
60. Iyer MK, Chinnaiyan AM. RNA-Seq unleashed. Nat Biotechnol. 2011;29(7):599–600.
61. Broad Institute. Picard [Available from: http://broadinstitute.github.io/picard].
62. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012;28(11):1530–2.
63. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28(16):2184–5.
64. t Hoen PA, Friedlander MR, Almlof J, Sammeth M, Pulyakhina I, Anvar SY, et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol. 2013;31(11):1015–22.
65. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121–32.
66. Nazarov PV, Muller A, Kaoma T, Nicot N, Maximo C, Birembaut P, et al. RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. BMC Genomics. 2017;18(1):443.
67. Hardwick SA, Deveson IW, Mercer TR. Reference standards for next-generation sequencing. Nat Rev Genet. 2017;18(8):473–84.
68. National Center for Biotechnology Information. Sequence read archive [Available from: https://www.ncbi.nlm.nih.gov/sra].
69. National Center for Biotechnology Information. Gene expression omnibus [Available from: https://www.ncbi.nlm.nih.gov/geo/].
70. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 2011;21(9):1543–51.
71. Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, et al. The External RNA Controls Consortium: a progress report. Nat Methods. 2005;2(10):731–4.
72. Hardwick SA, Chen WY, Wong T, Deveson IW, Blackburn J, Andersen SB, et al. Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods. 2016;13(9):792–8.
73. Lutzmayer S, Enugutti B, Nodine MD. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. Sci Rep. 2017;7(1):5913.

74. Lukas Paul PK, Horner G, Ante M, Hollaender I, Alexander S, Reda T. SIRVs: Spike-In RNA Variants as external isoform controls in RNA-sequencing. bioRxiv. Posted October 13, 2016. https://doi.org/10.1101/080747.
75. National Institute of Standards and Technology. ERCC 2.0: developing a new suite of RNA controls 2017 [updated 02/17/2017]. Available from: https://www.nist.gov/programs-projects/ercc-20-developing-new-suite-rna-controls.
76. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
77. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
78. Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 2015;43(21):e140.
79. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31(1):46–53.
80. Wang W, Qin Z, Feng Z, Wang X, Zhang X. Identifying differentially spliced genes from two groups of RNA-seq samples. Gene. 2013;518(1):164–70.
81. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014;111(51):E5593–601.
82. Steijger T, Abril JF, Engstrom PG, Kokocinski F, Consortium R, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013;10(12):1177–84.
83. Bradford JR, Hey Y, Yates T, Li Y, Pepper SD, Miller CJ. A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics. 2010;11:282.
84. Rehrauer H, Opitz L, Tan G, Sieverling L, Schlapbach R. Blind spots of quantitative RNA-seq: the limits for assessing abundance, differential expression, and isoform switching. BMC Bioinformatics. 2013;14:370.
85. Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. Nat Commun. 2017;8(1):59.
86. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. Posted November 14, 2017. https://doi.org/10.1101/201178.
87. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet. 2013;93(4):641–51.
88. Schram AM, Chang MT, Jonsson P, Drilon A. Fusions in solid tumours: diagnostic strategies, targeted therapy, and acquired resistance. Nat Rev Clin Oncol. 2017;14(12):735–48.
89. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol. 2011;7(5):e1001138.
90. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011;12(8):R72.
91. Li Y, Chien J, Smith DI, Ma J. FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. Bioinformatics. 2011;27(12):1708–10.
92. Torres-Garcia W, Zheng S, Sivachenko A, Vegesna R, Wang Q, Yao R, et al. PRADA: pipeline for RNA sequencing data analysis. Bioinformatics. 2014;30(15):2224–6.
93. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 2013;14(2):R12.
94. Beccuti M, Carrara M, Cordero F, Lazzarato F, Donatelli S, Nadalin F, et al. Chimera: a Bioconductor package for secondary analysis of fusion products. Bioinformatics. 2014;30(24):3556–7.
95. Kumar S, Vo AD, Qin F, Li H. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. Sci Rep. 2016;6:21597.
96. Davila JI, Fadra NM, Wang X, McDonald AM, Nair AA, Crusan BR, et al. Impact of RNA degradation on fusion detection by RNA-seq. BMC Genomics. 2016;17(1):814.
97. Rogers TM, Arnau GM, Ryland GL, Huang S, Lira ME, Emmanuel Y, et al. Multiplexed transcriptome analysis to detect ALK, ROS1 and RET rearrangements in lung cancer. Sci Rep. 2017;7:42259.
98. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. J Mol Diagn. 2018;20(1):4–27.
99. Monzon FA, Lyons-Weiler M, Buturovic LJ, Rigl CT, Henner WD, Sciulli C, et al. Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. J Clin Oncol. 2009;27(15):2503–8.
100. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. Nat Rev Clin Oncol. 2011;9(1):48–57.
101. Baehner FL, Lee M, Demeure MJ, Bussey KJ, Kiefer JA, Barrett MT. Genomic signatures of cancer: basis for individualized risk assessment, selective staging and therapy. J Surg Oncol. 2011;103(6):563–73.
102. Chibon F. Cancer gene expression signatures – the rise and fall? Eur J Cancer. 2013;49(8):2000–9.
103. Pham MX, Teuteberg JJ, Kfoury AG, Starling RC, Deng MC, Cappola TP, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. N Engl J Med. 2010;362(20):1890–900.
104. Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ. Gene expression profiling in human neurodegenerative disease. Nat Rev Neurol. 2012;8(9):518–30.
105. Drew JE. Cellular defense system gene expression profiling of human whole blood: opportunities to predict health benefits in response to diet. Adv Nutr. 2012;3(4):499–505.
106. Shih B, Watson S, Bayat A. Whole genome and global expression profiling of Dupuytren's disease: systematic review of current findings and future perspectives. Ann Rheum Dis. 2012;71(9):1440–7.
107. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177–95.
108. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med. 2012;4(154):154ra35.
109. Adam M, Ardinger H, Pagon R, Wallace S, Bean L, Stephens K, et al. GeneReviews®. Seattle: University of Washington; 1993–2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1116/.
110. Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet. 2012;91(6):1022–32.
111. Naidoo N, Pawitan Y, Soong R, Cooper DN, Ku CS. Human genetics and genomics a decade after the release of the draft sequence of the human genome. Hum Genomics. 2011;5(6):577–622.
112. Teekakirikul P, Kelly MA, Rehm HL, Lakdawala NK, Funke BH. Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era. J Mol Diagn. 2013;15(2):158–70.

113. Mayer AN, Dimmock DP, Arca MJ, Bick DP, Verbsky JW, Worthey EA, et al. A timely arrival for genomic medicine. Genet Med. 2011;13(3):195–6.
114. Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Greiner TC, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. J Mol Diagn. 2012;14(6):525–40.
115. Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, et al. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. Blood. 2013;121(22):e138–48.
116. Costa V, Angelini C, D'Apice L, Mutarelli M, Casamassimi A, Sommese L, et al. Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21. PLoS One. 2011;6(4):e18493.
117. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011;43(6):513–8.
118. Costa V, Aprile M, Esposito R, Ciccodicola A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet. 2013;21(2):134–42.
119. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends Genet. 2011;27(2):72–9.
120. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714–9.
121. Peters TL, Kumar V, Polikepahad S, Lin FY, Sarabia SF, Liang Y, et al. BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children. Mod Pathol. 2015;28(4):575–86.
122. Vendrell JA, Taviaux S, Beganton B, Godreuil S, Audran P, Grand D, et al. Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. Sci Rep. 2017;7(1):12510.
123. Kim HP, Cho GA, Han SW, Shin JY, Jeong EG, Song SH, et al. Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. Oncogene. 2014;33(47):5434–41.
124. Chen PL, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA, et al. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. Cancer Discov. 2016;6(8):827–37.
125. Jamieson NB, Maker AV. Gene-expression profiling to predict responsiveness to immunotherapy. Cancer Gene Ther. 2017;24(3):134–40.
126. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell. 2016;165(1):35–44.
127. Harris MH, DuBois SG, Glade Bender JL, Kim A, Crompton BD, Parker E, et al. Multicenter feasibility study of tumor molecular profiling to inform therapeutic decisions in advanced pediatric solid tumors: the individualized cancer therapy (iCat) study. JAMA Oncol. 2016;2:608.
128. Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. JAMA. 2015;314(9):913–25.
129. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548(7667):297–303.
130. National Cancer Institute. NCI-MATCH [Available from: https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match].
131. Clinical Sequencing Evidence-Gathering Research Consortium. KidsCanSeq [Available from: https://cser-consortium.org/projects/27].
132. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310(5748):644–8.
133. Santoro M, Melillo RM, Fusco A. RET/PTC activation in papillary thyroid carcinoma: European Journal of Endocrinology Prize Lecture. Eur J Endocrinol. 2006;155(5):645–53.
134. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561–6.
135. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;458(7234):97–101.
136. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol. 2011;12(1):R6.
137. Robinson DR, Kalyana-Sundaram S, Wu YM, Shankar S, Cao X, Ateeq B, et al. Functionally recurrent rearrangements of the MAST kinase and notch gene families in breast cancer. Nat Med. 2011;17(12):1646–51.
138. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. Nature. 2012;488(7413):660–4.
139. Wu YM, Su F, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao X, et al. Identification of targetable FGFR gene fusions in diverse cancers. Cancer Discov. 2013;3(6):636–47.
140. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013;15(4):415–53.
141. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med. 2010;363(18):1693–703.
142. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, et al. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. Cancer. 2018;124:1682.
143. Monzon FA, Koen TJ. Diagnosis of metastatic neoplasms: molecular approaches for identification of tissue of origin. Arch Pathol Lab Med. 2010;134(2):216–24.
144. Erlander MG, Ma XJ, Kesty NC, Bao L, Salunga R, Schnabel CA. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. J Mol Diagn. 2011;13(5):493–503.
145. Varadhachary G. New strategies for carcinoma of unknown primary: the role of tissue-of-origin molecular profiling. Clin Cancer Res. 2013;19(15):4027–33.
146. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817–26.
147. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med. 2015;373(21):2005–14.
148. Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, Ferguson JS, et al. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. N Engl J Med. 2015;373(3):243–51.
149. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med. 2012;367(8):705–15.
150. Gerami P, Cook RW, Wilkinson J, Russell MC, Dhillon N, Amaria RN, et al. Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. Clin Cancer Res. 2015;21(1):175–83.
151. Berger AC, Davidson RS, Poitras JK, Chabra I, Hope R, Brackeen A, et al. Clinical impact of a 31-gene expression profile test for cutaneous melanoma in 156 prospectively and consecutively tested patients. Curr Med Res Opin. 2016;32(9):1599–604.
152. Zager JS, Gastman BR, Leachman S, Gonzalez RC, Fleming MD, Ferris LK, et al. Performance of a prognostic 31-gene expres-

sion profile in an independent cohort of 523 cutaneous melanoma patients. BMC Cancer. 2018;18:130.
153. Spratt DE, Zhang J, Santiago-Jimenez M, Dess RT, Davis JW, Den RB, et al. Development and validation of a novel integrated clinical-genomic risk group classification for localized prostate cancer. J Clin Oncol. 2017;36:581:JCO2017742940.
154. Sinicropi D, Qu K, Collin F, Crager M, Liu ML, Pelham RJ, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. PLoS One. 2012;7(7):e40092.
155. Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. Nat Rev Gastroenterol Hepatol. 2012;9(8):477–86.
156. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui CH, Ding L, et al. The pediatric cancer genome project. Nat Genet. 2012;44(6):619–22.
157. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14.
158. Li E, Hamm CM, Gulati AS, Sartor RB, Chen H, Wu X, et al. Inflammatory bowel diseases phenotype, C. difficile and NOD2 genotype are associated with shifts in human ileum associated microbial composition. PLoS One. 2012;7(6):e26284.
159. Woo PC, Lau SK, Teng JL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect. 2008;14(10):908–34.
160. Croucher NJ, Thomson NR. Studying bacterial transcriptomes using RNA-seq. Curr Opin Microbiol. 2010;13(5):619–24.
161. Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. Application of RNA-seq to reveal the transcript profile in bacteria. Genet Mol Res. 2011;10(3):1707–18.
162. Guell M, Yus E, Lluch-Senar M, Serrano L. Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nat Rev Microbiol. 2011;9(9):658–69.
163. van Vliet AH. Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett. 2010;302(1):1–7.
164. Condon C. Maturation and degradation of RNA in bacteria. Curr Opin Microbiol. 2007;10(3):271–8.
165. Deutscher MP. Degradation of stable RNA in bacteria. J Biol Chem. 2003;278(46):45041–4.
166. Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH. Structure and complexity of a bacterial transcriptome. J Bacteriol. 2009;191(10):3203–11.
167. Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, et al. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. PLoS Genet. 2009;5(7):e1000569.
168. Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, et al. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. Proc Natl Acad Sci U S A. 2009;106(10):3976–81.
169. Pallen MJ, Loman NJ, Penn CW. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr Opin Microbiol. 2010;13(5):625–31.
170. Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet. 2010;11(1):9–16.
171. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, et al. The Listeria transcriptional landscape from saprophytism to virulence. Nature. 2009;459(7249):950–6.
172. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, et al. The primary transcriptome of the major human pathogen Helicobacter pylori. Nature. 2010;464(7286):250–5.
173. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. Nat Rev Microbiol. 2012;10(9):618–30.
174. Westermann AJ, Barquist L, Vogel J. Resolving host-pathogen interactions by dual RNA-seq. PLoS Pathog. 2017;13(2):e1006033.
175. He J, Abdel-Wahab O, Nahas MK, Wang K, Rampal RK, Intlekofer AM, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. Blood. 2016;127(24):3004–14.
176. Kolata G. In treatment for leukemia, glimpses of the future. New York Times. 2012 July 8, 2012;Sect. A1.
177. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

miRNA Expression Assays

5

Cornelia Braicu, Diana Gulei, Beatriz de Melo Maia, Ioana Berindan-Neagoe, and George A. Calin

Introduction

The most studied transcripts of the human genome were until recently those related to protein-coding genes. But, the coding exons of these genes represent less than 1.5%, this value being increased for approximate 2% if the untranslated regions are included [1]. In recent years, an intensified interest of the role of the noncoding RNA (ncRNA) was observed, its decisive function in multiple physiological and pathological processes being demonstrated [1].

ncRNAs are sequences that comprise between 19 and 200 nucleotides long and are generally classified into two main categories according to their length: short and long noncoding RNAs (lncRNAs). They include transcript variants such as microRNAs (miRNAs), small interfering RNAs (siRNA), small nucleolar RNAs (snoRNAs), transfer RNAs, piwi-interacting RNAs (piRNAs), transcribed ultraconserved regions (T-UCRs), and large intergenic noncoding RNAs (lincRNAs) [1–4]. In 2006, Andrew Fire and Craig Mello received a Nobel Prize for their discovery of RNA interference—gene silencing by double-stranded RNA [5], a proof of the tremendous scientific interest generated by their discovery.

miRNAs are single-stranded RNA structures that are 19–25 nucleotides long. They regulate the expression of genes by binding to different sites in the 3′ untranslated regions (3′ UTR) of several target mRNAs, which causes mRNA cleavage (degradation) or translational repression [6]. Moreover, a single miRNA can modulate the expression of multiple genes, where one gene can be targeted by more than one miRNA sequence. This is due to the imperfect complementarity rules, where the hybridization process is based on one short seed sequence.

Through their differential expression, miRNAs have the ability to sustain pathological phenotypes. Therefore, differences in miRNA expression between normal and pathological states are currently explored as diagnostic, prognostic, and therapeutic tools in cancer and cardiovascular and autoimmune diseases, to name a few [3, 4, 6–9]. Transcriptional modulation of target genes can be achieved by administration of exogenous miRNA mimics or inhibitors in order to repress aberrant translation of target proteins involved in the development of a specific disease [2–4, 8].

In the era of precision medicine, clinicians aim to use miRNAs as tools for individualized diagnostic and therapeutic approaches. This year marked the first clinical trial involving the therapeutic administration of miRNA mimic

C. Braicu
Research Center for Functional Genomics, Biomedicine and Translational Medicine,Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

D. Gulei
Medfuture Research Center for Advanced Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

B. de Melo Maia
Research and Development, PHD Laboratory – Surgical and Molecular Pathology, São Paulo, SP, Brazil

I. Berindan-Neagoe
Research Center for Functional Genomics, Biomedicine and Translational Medicine,Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
Medfuture Research Center for Advanced Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Department of Functional Genomics and Experimental Pathology, The Oncology Institute "Prof. Dr. Ion Chiricuta", Cluj-Napoca, Romania

G. A. Calin (✉)
Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: gcalin@mdanderson.org

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_5

(MRX34) in solid and hematological tumors (ClinicalTrials.gov Identifier: NCT01829971).

Currently, there are a wide range of platforms for miRNA identification, confirmation, and profiling, each having specific strengths and limitations [7, 9, 10]. In this sense, focusing on miRNA extraction, quality control, molecular profiling, expression profiling, and sequencing in different types of biological samples (cells, tissues [in vitro and in vivo], and biological fluids) is of paramount importance as these methods can be useful for pathologists as additional methods to classical immunohistochemistry in order to validate specific diagnoses.

In this chapter, we emphasize methods for the assessment of miRNA expression profiles in cells, tissues, and body fluids and highlight the main advantages and disadvantages of these methods in specific biological applications. This complete picture can help pathologists combine new methods and classical ones for diagnosis and therapeutic purposes.

miRNA Target Key Pathways with Important Implications in Human Disease

In spite of the fact that initially miRNAs were considered as a part of the dark matter, the latest investigations characterize miRNAs as key elements of the transcriptional and translational machinery with repercussions on cell fate [4, 8, 10–15]. miRNAs exhibit a tissue specificity profile and are present in complex regulatory networks of cellular function, tissue differentiation, and maintenance of cell identity and evolution from embryogenesis to adult life. Also, multiple fundamental processes are modulated by miRNAs, such as cell differentiation, apoptosis, tumor initiation, invasion, and metastasis [13, 16, 17].

Alterations of particular signaling pathways or biological processes are often correlated with a specific pathology [12, 16]. The capacity of miRNAs to modulate key signaling pathways has been extensively documented [12, 17]. One relevant example is the miR-34 family, which includes important components of the p53 pathways. TP53 tumor suppressor gene was demonstrated to directly transactivates miR-34a/b/c followed by modulation of the cell cycle and apoptosis [18, 19]. miRNAs such as miR-125b [20] and miR-101 [21] were demonstrated to regulate the major controller of cancer growth in the prostate, namely, androgen receptor signaling. Similarly, miR-106b [22, 23] and miR-23b [23] can target members of the phosphatidylinositol 3-kinase/Akt/PTEN signaling [24]. miRNA implication in cancer is also related to their ability to regulate apoptosis, cellular growth, and proliferation [12]. For example, miR-14 is required for growth control, and *let-7* family miRNAs are regulators of the proto-oncogene *RAS*. Interestingly, the mir-17 cluster is located on human chromosome 13 that is frequently amplified in B-cell lymphomas [19, 25, 26].

Understanding miRNA roles in different disease processes is an ongoing process that remains far from complete, although preclinical and clinical data are so far encouraging [12, 26]. Below we provide a broad overview of miRNAs as biomarkers of prognosis, therapy prediction, and targets of therapy, demonstrating their possible roles in clinical management in the era of precision medicine [11].

Cancer

Several miRNAs are located at genomic regions linked to cancer [25]. During the cellular transformation, some miRNAs are specifically deregulated, and their altered expression and functions lead to important disease phenotypes [11].

Calin and colleagues provided the first identification of miRNA involvement in cancer more than a decade ago (2002) in chronic lymphocytic leukemias (CLL) [27]. In this study, miR-15a and miR-16-1 were deleted or downregulated in most CLL samples indicating the potential of miRNA regulatory control over target genes [27], including BCL-2 [28]. The same group subsequently described a unique miRNA signature associated with prognostic factors and disease progression in CLL [29]. Additionally, TCL-1, an important oncogene in B-cell CLL responsible for the aggressive form of the disease, was shown to be regulated by miR-29 and miR-181, two miRNAs differentially expressed in CLL [30], and the role of miRNAs and ncRNAs was broadly described in this disease [31]. The expression levels of miR-21 were shown to be significantly higher in patients with poor prognosis and were able to predict the overall survival. A score termed 21FK based on miR-21 evaluation by quantitative reverse transcription-PCR (qRT-PCR), fluorescence in situ hybridization (FISH), and karyotype was proposed to predict patients' survival. Patients with low 21FK score demonstrated significantly better survival [32]. More recently, miR-155 was shown to be overexpressed in B cells from individuals with monoclonal B-cell lymphocytosis (MBL, a premalignant condition) and was successfully identified circulating in microvesicles of both MBL and CLL patients. miR-155 overexpression was also observed, in patients with CLL who did not reach complete responses after therapy, pointing to miR-155 as a powerful biomarker for progression in individuals with MBL and for prediction to therapy in individuals with CLL [32].

In solid tumors, various miRNAs are recognized to control the expression of tumor suppressor genes and oncogenes, whereas others have predictive value for treatment response and survival, including miR-21 and miRNA-221/miRNA-222, which are usually overexpressed in various cancers.

miR-21 is the most upregulated miRNA in solid tumors, as compared with matching noncancerous tissue [7, 10, 33]. miR-21 has been shown to promote tumor proliferation and invasion in gastric cancer through the suppression of PTEN expression [34]. Increased expression of miR-21 in non-small cell lung cancer (NSCLC) was also shown to be significantly associated with worse survival and increased lymph node metastasis [35]. In another study, an eight-miRNA signature was shown to discriminate, with high sensitivity and accuracy, histologic types of lung cancers. miR-21 and miR-29b are significantly overexpressed in NSCLC compared to small cell lung cancer (SCLC), while miR-129 and miR-205 are differentially expressed in squamous versus non-squamous lung cancers [36]. In colorectal cancer patients, serum levels of miR-21 were correlated with recurrence and mortality, making it a potential prognostic marker in this type of tumors [37]. In invasive ductal carcinomas of the breast, high expression of miR-21 was associated with clinicopathological features such as tumor size, stage, grade, negative estrogen receptor (ER) expression, human epidermal growth factor receptor 2 (HER2) expression, high Ki-67 expression, mastectomy, and lower overall survival, making it a potentially important prognostic factor in breast cancer [38]. Other well-studied miRNAs are the miR-17-92 clusters, consisting of six miRNAs with the same seed sequence, part of a cluster located on human chromosome 13q31. This region has been known to be frequently amplified in several types of lymphoma [39] as well as solid tumors [40] including childhood solid tumors [41]. In oral squamous cell carcinomas, miR-17/20a was shown to regulate cell migration inhibition and negatively correlate with TNM-stage and lymphatic metastasis [42]. High levels of miR-17, miR-20a, and miR-92-1, along with miR-15a and miR-16-1, are associated with poor prognosis in multiple myeloma (MM) with shorter progression-free survival [43]. In colon cancer, upregulation of this cluster was also correlated with poor prognosis, and miR-17-92 expression has been identified as an independent prognostic factor [44].

As noninvasive biomarkers, urine miRNA assessment in urologic cancers such as bladder, prostate, and renal cell carcinomas has been explored [24]. It is demonstrated that miRNAs are secreted by the tumor and have important role in the intercellular communication [45]. Increased levels of miR-126, miR-182, and miR-199a have been documented in urine of bladder cancer patients indicating the potential of miRNAs as a biomarker of the disease [46].

miRNA detection in exosomal plasma has revealed that a combination of miR-126 and miR-449a (or miR-34b-5p) has the potential to detect renal carcinoma with high specificity and sensitivity [45].

Recently, potential synergy between miR-155 inhibition and chemotherapy response for the treatment of lung cancer has been proposed. Administration of miR-155 inhibitor sensitized tumors to the action of chemotherapeutic agents as a result of impairment of miR-155/TP53 feedback loop involved in chemoresistance [47].

miRNAs in Drug Resistance

Drug resistance, where cells become insensitive to the cytotoxic action of therapeutic agents, is considered one of the main reasons for increased mortality rates among patients, especially in the oncological field.

miRNAs play an important role in drug resistance mechanisms, being able to regulate pathways involved in drug transport and cell metabolism. miR-21 upregulation was associated with resistance to NSC265450 (nogamycin) and NSC67055014 agents when administrated in three different lung cancer lines [48]. Experimental inhibition of such upregulation could impair the resistant phenotype of cancer cell and increase the action of chemotherapeutic drugs. In breast cancer cell lines, miR-451 may affect response to DOX through modulation of MDR1 (multidrug resistance 1) gene. Inhibition of mdr1 via miR-451 overexpression sensitized MCF-7/DOX-resistant cells, indicating a possible molecular therapeutic strategy for potentiating DOX efficacy [49]. In colorectal cancer, miR-222 also plays a role in the development of multidrug resistance by modulation of ADAM-17 [50]. Similarly, miR-122 and miR-29a were shown to contribute to resistance to adriamycin and docetaxel in breast cancer [51]. Downregulation of miR-29 is capable of increasing resistance to cisplatin in ovarian cancer cells [52]. An increased level of miR-181a is associated with resistance to radiotherapy, by regulating the proapoptotic PRKCD protein in cervical [53, 54] and ovarian cancers [55]. Finally, miRNA-21 has been linked to increased resistance to cisplatin by negatively regulating PTEN in ovarian cancers [56]. In this sense, future studies targeting miRNAs related to drug resistance can be of paramount importance in clinical practice.

miRNAs as Epigenetic Regulator

Epigenetic regulatory mechanisms have emerged as a central pathway in disease pathogenesis [57, 58]. miRNAs are important effectors of epigenetic machinery [58, 59]. Studies on the impact of methylation and acetylation on miRNA expression in cancer are actively pursued with important clinical implications [60]. miR-129-2 is frequently methylated in hepatocellular carcinoma cells (HCC) [61] and CLL, adversely impacting survival in the latter [62]. Members of miR-29 family can negatively modulate DNA methyltransferase DNMT3A and DNMT3B enzymes in lung cancer. Lower survival rates are observed in lung cancer patients with higher

levels of DNMT3A [63]. This indicates that interactions between the miRNome and the epigenome can provide new grounds for future cancer therapy studies. Additional examples of miRNA interactions with epigenetics in lung cancer include the associations of miR-29, miR-141, and miR-499 with histone modification and involvement of miR-34b, miR-126, or miR-212 in DNA methylation [57]. In colorectal cancer, miR-125b and miR-125a (miR-125 family) are downregulated through hypermethylation-associated mechanisms. Therefore, expression status of miR-125 family has been proposed as a potential biomarker for colorectal cancer [64], some relevant examples being presented in Table 5.1.

In nonneoplastic disease, DNMT1-related microRNAs (miR-21, miR-126, miR-148a, miR-181s, and miR-29s) as well as renal function-associated microRNAs (miR-26a, miR-30b, miR-29c, miR-130, and miR-150) have been proposed as biomarkers in lupus [65].

Histone methyltransferase EZH2 was proved to strongly reactivate the miR-212 expression in lung cancer cells; meanwhile in the case of prostate cancer, downregulation of miR-101 was correlated with upregulation of EZH2 [66]. Furthermore, histone deacetylase (HDAC) inhibition has been shown to alter miRNA levels [67]. A negative association among HDAC5 and miR-589-5p was demonstrated in lung cancer [66]. miR-206 and miR-9 appear to target HDACs and histone acetyl transferases (HATs) [68]. HATs have been shown to acetylate androgen receptor (AR), sustaining the transcriptional activity of these receptors via upregulation of p300/CBP-associated factor (PCAF) in prostate cancer [69].

Table 5.1 Some relevant examples related to the implication of miRNAs in epigenetic-related mechanism

Epigenetic event	Epigenetic gene	Target miRNA	References
DNA methylation	–	miR-34b, miR-126 or miR-212, miR-129	[4]
Epigenetic effector DNA methyltransferase	DNMT1	miR-21, miR-126, miR-148a, miR-181s, and miR29	[177]
	DNMT3A and DNMT3B	miR-29	[63]
Histone methyltransferase (HMTs)	EZH2	miR-101, miR-212, *miR-181a, miR-181b, miR-200b, miR-200c*, and *miR-203*	[178]
Histone deacetylase inhibition (HDACs)	HDACs	miR-27a and miR-27b	[67]
	HDAC5	miR-589	[66]
	HDAC4 and HDAC5	miR-9	[68]
Histone acetyltransferases (HATs)	Myst3	miRNA-9* and miRNA-206	[68]
	PCAF	miR-17	[69]

Dicer ribonuclease to generate the mature miRNA and antisense miRNA star (miRNA) products

Adipogenesis

miRNAs are dysregulated in adipose tissue of obese patients. miRNAs appear to play regulatory roles in many biological processes associated with obesity, including adipocyte differentiation, insulin action, lipid storage processes, and fat metabolism [70, 71]. Several miRNAs were described in adipocytes and appear to have a role in the modulation of adipogenesis, which offer an opportunity for targeting adipogenesis dysfunction by controlled delivery of miRNA structures [70, 72]. Most of the studies on miRNAs in adipogenesis are based on murine models, reporting the activity of miRNAs in various processes regulating adipogenesis. In mice models, the cluster miR-17-92, miR-200, and miR-103 has a pro-adipogenic role, and let-7 and miR-27a/b are anti-adipogenic [70, 72, 73]. During adipogenesis, miRNAs can accelerate or inhibit adipocyte differentiation and hence regulate fat cell development. In addition, miRNAs may regulate adipogenic lineage commitment in multipotent stem cells and hence govern fat cell numbers. Recent findings suggest miR-519d is associated with obesity, but larger case-control studies are needed. Few miRNA targets have been experimentally validated in adipocytes. Both miR-27 and miR-519d target PPAR family members, well-established regulators of fat cell development [48].

miR-181 is also shown to be involved in regulation of fat metabolism [74], mainly via IDH1 [75]. miR-181 family is linked to reduced expression of genes that regulate lipid synthesis mechanisms and to the overexpression of the genes related to b-oxidation, thus leading to reduced lipid accumulation [75]. miR-181a's overexpression accelerates adipocyte differentiation, the mechanism being related to regulation of TNF-α [76] or via TGFβ/Smad signaling pathway [77].

Cardiovascular Diseases

Recent investigations provide ample evidence that miRNAs modulate a wide range of cardiac functions with developmental, pathophysiological, and clinical implications [78]. miRNA expression analysis has pointed to the potential role for miR-1, miR-16, miR-27b, miR-30d, miR-126, miR-133, miR-143, and the let-7 family in mammalian heart development [79].

Recently, deregulated expression of miR-1 and miR-133 was reported in heart failure patients [80, 81] and cardiac hypertrophy [82]. Increased expression of miR-1, miR-133, miR-499, and miR-208 has been shown to be associated with cardiac injury following acute myocardial infarction, while reduced levels of miR-126 is linked to the development of coronary artery disease and diabetes [83]. Recently, circulating miR-181a was proposed as a noninvasive biomarker for acute myocardial infarction given its altered expression in necrotic processes. Furthermore, it is highly overexpressed in association with the differentiation of immune response effectors related to infarction and vascular inflammation [84].

Autoimmune Diseases

Immune modulatory genes are highly regulated by miRNAs. Toll-like receptors, known mediators in microbial infections, and their ligands induce miR-155 expression, immune cell survival, and cytokine signaling suppression [85]. miRNA is heavily implicated in the molecular mechanisms that regulate the immune system during the development of autoimmune diseases. Examples of specific types of autoimmune disorders in which miRNAs are involved in immune response include rheumatoid arthritis, where miR-155 and miR-146a are deregulated, multiple sclerosis where miR-18b and miR-599 are associated with disease relapse, and systemic lupus erythematous where miR-155 and miR-146a appear to be involved in upregulating and downregulating disease activity, respectively. miR-941 is potentially involved in the pathogenesis of inflammatory bowel disease [85]. In Sjogren's syndrome, an autoimmune disease affecting salivary and lacrimal glands, miR-146a and miR-155 are deregulated during disease development. miR-21 has been shown to be activated in T cells in psoriasis, enhancing dermal inflammation [86]. Several miRNAs such as miR-510, miR-191, and miR-342 were found to be deregulated in association with type I diabetes. In this condition, miR-21 expression was also highly correlated with the severity of the disease [87].

Which Types of Biological Samples Can Be Used for miRNA Studies?

An important issue in the evaluation of miRNA expression is the quality of biological samples and the RNA isolation method used to obtain an appropriate quantity of total RNA in order to achieve reproducible, reliable results [88]. Because of their short sequences, miRNAs are known to be relatively stable and homogeneous compared to RNA and DNA. They are well preserved in a wide variety of biological samples, including various body fluids such as blood serum/plasma, saliva, and urine [3, 4, 6, 88–90]. miRNAs can also be isolated from fresh tissues, from cell cultures, and more recently from formalin-fixed paraffin-embedded (FFPE) tissue blocks and archived materials, using various methods (Fig. 5.1). The extraction of miRNAs from FFPE samples allows pathologists to retrospectively use archival paraffin blocks from previously diagnosed patients with invaluable corresponding clinical follow-up data [33, 91, 92] to assess the relationship with response to treatment and overall survival. The in situ *hybridization* (ISH) technique for miRNA evaluation has also

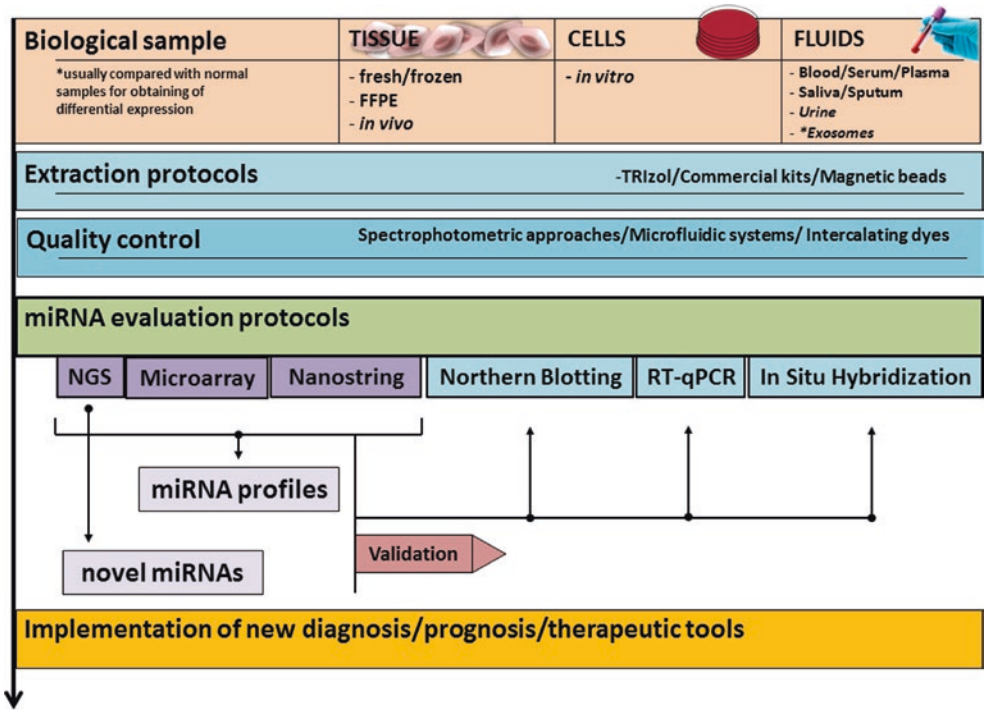


Fig. 5.1 Biological samples and methods for microRNA evaluation. The figure illustrates a variety of biological samples and evaluation methods for miRNA assessment, targeting the achievement of routine early diagnosis, prognostication, and therapy for these molecules. *LCM* laser capture microdissection

been developed for FFPE samples as a useful method for comparison with immunohistochemistry (IHC).

The major limitation of using FFPE tissue samples in molecular biology applications is the less than optimal quality of RNA extracted from them; specifically, the nucleic acids are degraded to fewer than 300 base pairs in length and are chemically modified during formalin fixation [33, 91]. However, due to the short sequence of miRNAs, they can be successfully extracted from FFPE tissue blocks using several commercially available kits [91, 92]. Additionally, miRNA expression profiles in FFPE tissue are closely similar to those in fresh tissue, proving that, if prepared appropriately, FFPE tissue samples are excellent resources for miRNA expression investigations [33, 91, 92]. FFPE tissues also present a valuable resource for biomarker discovery particularly for the identification of cancer subtypes. An example is the identification of specific transcript panel for various genomic subtypes of breast cancer including basal-like (miR-18a, miR-135b, miR-93, and miR-155), HER2 type (miR-142-3p and miR-150), and normal-like (miR-145, miR-99a, miR-100, miR-130) [10].

Several preparation methods are commonly used prior to miRNA isolation [93, 94]. During extraction and purification of RNA, preventing the loss of small RNA species is the main concern. Therefore, using a robust miRNA isolation technique and ensuring the stability of stored miRNA samples that were isolated using these methods are highly important [14, 93, 94].

miRNA expression and localization studies have demonstrated the potential for their utility as screening and early diagnosis markers for various pathologies through their identification in body fluids. Therefore, development of reliable instruments for the assessment and quantification of these miRNAs, which are found at minute concentrations, is important. Fresh tissue samples, leukocytes, and cell lines usually furnish large quantities of high-quality miRNAs, which are required for miRNA profiling studies (Fig. 5.1). Several kits are now available for miRNA extraction of good quality and quantity, including from FFPE tissues.

Regarding the need for matched pairs of normal and tumor samples, both normal and tumor tissue samples ideally should be obtained from the same patient, as described by Yanaihara et al. [90]. Access to both normal and tumor cells can be highly insightful for pathologists when evaluating miRNA expression. Nevertheless, tumor cell lines can also be suitable models for studying the effects of downregulated and upregulated miRNA expression, in spite of the fact that these may contain genetic abnormalities never identified in patient tumors. A recommended step at the start of a miRNA-related study is to select clinical samples with different characteristics (e.g., colorectal tumor samples with and without microsatellite instability, blood cells obtained from CLL patients with poor and good prognosis) in relatively similar numbers (more than 30 in each group for performance of various statistical analyses). If the number of normal controls is not sufficient for conferring statistical power, the data can still be analyzed by comparing the two (or more) clinically distinct sets of samples. Because disease-oriented profiling uses few types of samples (e.g., malignant and normal control samples in the simplest example), replicates of each sample are not needed for profiling. In more detailed studies using limited numbers of samples that are very different biologically (e.g., cells transfected with a specific reagent and non-transfected controls), analysis of three replicates of each sample is critical.

Investigators have observed large differences in miRNA expression in serum and plasma samples. The possibility has been raised that plasma and serum might display several differences in their miRNA content. These differences in miRNA concentration may be due to the fact that the coagulation process modifies the pattern of blood miRNA [95]. However, differences can also be caused by platform-dependent variations in measurement of miRNAs [14, 96].

miRNA Extraction Protocols

Sample preparation and RNA extraction approaches can have direct consequences for miRNA analysis and profiling, especially with samples that are prone to miRNA degradation (Fig. 5.2). All protocols for miRNA extraction have a basic step of cell/tissue lysis with the exception of those for body fluids. For tissue samples, performance of this step using mechanical disruption of the sample is recommended, as researchers have observed higher efficiency in quantity when using a homogenizer. The samples must be processed according to the manufacturer's extraction protocol, and every effort needs to be made in order to eliminate contaminants, such as xylene, chloroform, and TRIzol.

Extraction of miRNAs using TRIzol or TRI Reagent is a method initially proposed as an approach to facilitate the elimination of proteins from nucleic acids. This method is recommended primarily for cells and tissues with increased expression of endogenous RNase or when separation of cytoplasmic RNA from nuclear RNA is required [97, 98]. The TRIzol/TRI Reagent extraction procedure is a reliable method for isolation of miRNA species, as reported in studies that compared it with other commercially available kits for miRNA extraction. When correctly stored and managed using these procedures, miRNA samples have not exhibited degradation [97–99], information summarized in Table 5.2. However, at least one study has called attention to the loss of short structured RNAs with low GC content during extraction with this reagent when using a small number of cells [100].

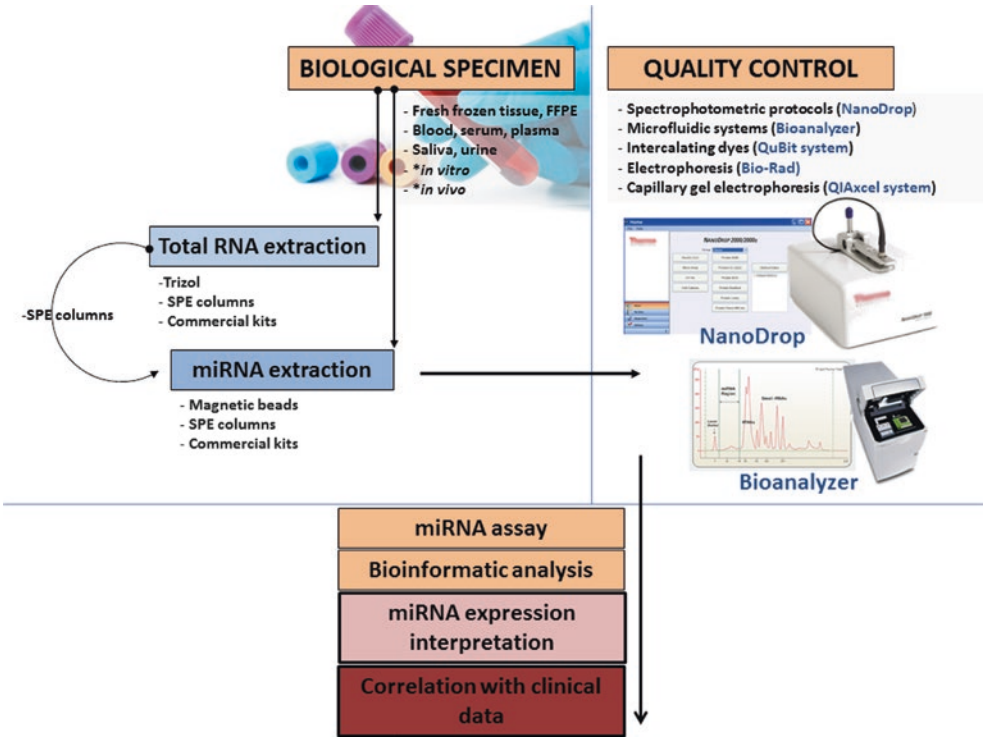


Fig. 5.2 miRNA extraction protocols, quality control, and quantification. The recovery of miRNA from biological samples through TRIzol/TRI Reagent, SPE columns, and magnetic beads is shown on the *top* of the figure. miRNA quality control through the use of a NanoDrop or a Bioanalyzer is highly recommended prior to miRNA assays, statistics and bioinformatics analysis, and data correlation with clinical outcome in order to generate accurate results. *SPE* solid-phase extraction. *overexpressed level for the analysed transcript

Table 5.2 Biological samples used for miRNA evaluation and their utility

Sample type		miRNA quantity (ng)	Utility	References
Cell lines		≥1000	Identification of new miRNA; validation of action mechanisms and target pathways; identification of novel target therapies	[10, 19, 20, 23, 97]
Biopsy	Macrodissection	1–100	Identification of novel early diagnostic biomarkers, novel prognostic biomarkers, biomarkers for prediction of the response to treatment or resistance to treatment, clinical investigation, clinical trials, translational research	[10, 14, 19, 20, 23, 97]
	LCM	≤1–10		
Fresh/ frozen tissue	Macrodissection	≥1000		
	LCM	10–200		
FFPE	Macrodissection	10–200		
	LCM	≤1–40		
Serum		1–100		
Plasma		1–100		
Saliva		≤1–20		
Urine		≤1–40		

FFPE formalin-fixed paraffin-embedded, *LCM* laser capture microdissection

The development of commercial miRNA extraction kits, such as the mirVana miRNA isolation kit (Ambion), mirPremier microRNA isolation kit (Sigma-Aldrich), miRNeasy mini kit (Qiagen), and miRCURY (Exiqon), was based on the principle of solid-phase extraction (SPE), which significantly improves the retention of small RNA species on SPE columns in a selective manner. Consequently, any redundant materials in the extraction procedures, including large RNAs, are eliminated prior to the final step of the protocol, in which small RNAs are eluted. The use of SPE appears to be the most widely applied and effective approach for miRNA extraction.

The use of magnetic beads also ensures an easy and rapid workflow for miRNA extraction, thus avoiding the use of hazardous chemicals. Specifically, miRNAs are immobilized on magnetic beads, decanting the contaminants in a solution. miRNA purification is done by a magnetic field to extract the magnetic beads from the solution, after decontamination [101]. This extraction method has many advantages for different samples with different concentrations of miRNAs, such as serum/plasma, saliva, and urine.

miRNA Quantity and Quality Control

Quantification of nucleic acids, including miRNAs, extracted from different types of samples is essential for quality and quantity control. It is recommended to use the same amounts of miRNA when comparing various biological samples [99]. Recent experimental data demonstrated that the miRNA expression profile is affected by RNA integrity [91, 93, 97]. Samples with low RNA integrity exhibited the highest miRNA concentrations, because when RNA is degraded, this results in the formation of small RNA species, leading to an overestimation of the miRNA amount. However, assessment of RNA integrity must be a routine step in assessing miRNA expression patterns [94, 96, 99]. Importantly, measuring RNA concentrations using different platforms is a challenging process, so comparing miRNA profiles in samples evaluated using different platforms can be very complex.

In practice, miRNAs are quantified using the same techniques as those used to quantify DNA and mRNA (Fig. 5.2). The lack of specificity that arises during miRNA quantification and quality control is the main drawback of the available techniques. The most commonly used methods for quantifying miRNAs are spectrophotometric approaches (μ-volume quantification, e.g., by the NanoDrop instrument series, based on absorbance determination), which measure the concentrations of total nucleic acid species but are not able to differentiate miRNA species; microfluidic systems (2100 Bioanalyzer [Agilent Technologies]); intercalating dyes (QuBit system, using Invitrogen, SYBR Green); Experion automated electrophoresis system (Bio-Rad); and capillary gel electrophoresis (QIAxcel Advanced System, Qiagen) [102].

Staining with fluorescent RNA-binding dyes (e.g., RiboGreen RNA Assay kit [Promega, Ambion]) is preferred when assessing nucleic acids at low concentrations. Most of the methods used for integrity evaluation are based on an assessment of 18S and 28S ribosomal RNA, followed by different algorithms for analysis. These methods have different sensitivities, producing a wide range of results, and therefore require objective comparisons of the experimental data [96, 99].

miRNA Stability

For all miRNA assay procedures sample processing should be performed in an RNase-free environment in order to prevent degradation of nucleic acids and to produce consistent results. At present, a wide range of RNase-free materials and reagents is commercially available. To prevent degradation of nucleic acids, including miRNA species, protocols for storage indicate that RNAs should be deposited at −80 °C for prolonged periods to enable the generation of reproducible and reliable data. In the case of cDNA synthesized based on miRNA sequence, the appropriate storage temperature for miRNA expression profile studies is −20 °C. In a recent study, researchers presented the TRIzol/TRI Reagent system as the gold standard for miRNA extraction [94]. Rapid degradation of miRNAs and cDNA appeared to be independent of the extraction method. In order to reduce the degradation processes, proper storage conditions for RNA and cDNA are essential and should be used to ensure the accuracy of the experimental data and to allow correlation of different studies presented in the literature [103].

Degradation of nucleic acids material can occur during surgical procedures. For example, certain transurethral resection of bladder tumor (TURBT) technique may lead to significant DNA/RNA degradation in resected tumor samples. Luckily, miRNAs are relatively spared [104] and can still be used for biomarker discovery.

Advanced Techniques for Examining miRNA Expression

Assessing miRNA expression profiles as a diagnostic, prognostic, or therapeutic tool is performed using molecular and biological methodologies including Northern blotting, polymerase chain reaction (PCR), qRT-PCR, ISH, miRNA microarrays, and NGS techniques [105–107].

As indicated above, miRNAs can be evaluated in a wide range of biological samples with variable content quantities and qualities including fresh and FFPE tissue samples. Laser capture microdissection (LCM) is required to obtain homogeneous tissue samples even at a lower yield (<10 ng of miRNA/sample) than in macrodissection. Plasma/serum, saliva, and urine are other types of samples in which low concentrations of miRNA are obtained and where quantification of miRNA is a major challenge.

Once investigators observe alterations of miRNA profiles in patients with different pathologies [108], quantification of primary miRNA (pri-miRNA), precursor miRNA (pre-miRNA) transcripts, and mature miRNAs is required to distinguish among different isomiR species, i.e., sequence variants of miRNAs [109, 110], as well as changes in mature form processing.

Northern Blotting

Northern blotting was the first technique used to identify miRNAs [111, 112], and up until recently, it was the only standardized and most widely used assay for small RNA research. This technique can reliably detect the expression profiles of miRNAs of interest, determine their sizes, and

accurately quantify and identify a predictive population of miRNAs with a specific role in a disease [113].

Despite its frequent use, Northern blotting has several technical limitations. These include the relatively large sample quantity requirement and poor sensitivity of routine analysis once mature miRNAs are very short and their prevalence in total RNA is very low. Furthermore, the close sequence similarity among miRNAs in the same families poses a challenge to the specificity of miRNA detection using Northern blotting [114].

New Northern blotting versions have been developed in an attempt to increase the sensitivity of the procedure and reduce total assay time. Nevertheless, Northern blotting and its variants are considered to be medium-throughput miRNA-screening techniques. Recently, a multiplexed Northern blotting system based on mechanism of hybridization chain reaction was developed [115].

Quantitative Reverse Transcription-Polymerase Chain Reaction

qRT-PCR is one of the most frequently used approaches to achieve gene expression quantification, including for mature miRNA and pre-miRNA expression profiling [93, 116]. Small quantities of miRNA or total RNA are reverse-transcribed into cDNA, followed by a quantitative PCR analysis in which accumulation of reaction products is observed in real time [10]. qRT-PCR has excellent sensitivity and sequence specificity, and it is the most often used method for expression profiling and confirmation of miRNA findings that were obtained by other methodologies, particularly by microarray assays. Most qRT-PCR-based miRNA expression quantification approaches are specific to 3′ end sites of targeted miRNAs [93, 116].

In qRT-PCR applications using hydrolysis probes, both mature miRNAs and their precursors must be assayed. There are several advantages in using hydrolysis probes, such as the capacity to detect specific miRNA precursors [10] and the achievement of higher reaction efficiency for mature rather than miRNA precursors. Other available qRT-PCR protocols for miRNAs use a single nonspecific dye, such as SYBR Green. These methods require treatment with DNase for removing genomic DNA [98] but are less costly and have a lower detection limit and a higher sequence specificity and accuracy than hydrolysis probe-based methods [78, 116].

Given that most miRNAs are 21–22 nucleotides long, a classic PCR primer of approximately the same size length imposes limitations on miRNA evaluation, because at least two non-overlapping primers are required for the exponential amplification phase in qRT-PCR [116, 117]. One way to circumvent this limitation is to extend the length of the miRNA, generally via polyadenylation [118, 119]. Because all RNA structures with 3′ ends will be polyadenylated using this procedure, it cannot differentiate among pre-miRNAs and mature miRNAs that are capable of activation of the RISC protein complex for the silencing of genes. Another restriction of this polyadenylation-based technique is its inability to quantify species containing 2′-oxymethyl modifications at their 3′ ends that will block polyadenylation. Companies like Qiagen, Stratagene, Agilent Technologies, and Invitrogen (NCode) developed commercial qRT-PCR assays for miRNA. Additionally, Exiqon and Eurogentec have developed assays that require proprietary reagents, such as locked nucleic acid (LNA), and complex modification steps that restrict their routine implementation. LNA and TaqMan miRNA qPCR assays offer the ability to distinguish different isomiRs that differ in internal base pairs, such as let-7 family members [110].

In Situ Hybridization

ISH is used for validation of experimental data and evaluation of relative expression levels [120, 121] and can be applied to discern biological differences between pre-miRNAs and mature miRNAs [122, 123]. Therefore, ISH provides powerful complementary data for confirmation of target miRNA, which enables progression from high-throughput investigations to more focused examination of the roles and localization of individual miRNA [93, 122]. Despite the technical difficulty associated with adapting ISH methods to miRNA quantification using fluorescent dyes [122, 124], ISH is a robust assay that has been applied to miRNA studies in a large number of organisms from bacteria to various eukaryote cell types [93, 116, 120, 123].

Paralogous miRNAs such as miR-29a and miR-29b differ by unique central nucleotides and a few nucleotides at their 3′ ends, existing in different cellular compartments (nuclear and cytoplasmic, respectively). Such differences in localization can only be illustrated using ISH techniques [93, 116]. ISH techniques also offer the advantage of delineating the cellular distribution of miRNA(s) and lncRNAs in tissue samples composed of different cell types (Fig. 5.3).

Microarrays

DNA microarray technology continues to evolve following its initial use for the measurement of differences in the amount of target DNA and RNA sequences among biological samples [106, 125–127]. Recently, DNA microarrays have been shown to be a powerful tool for the evaluation of alterations in an abundance of miRNA species [128, 129]. Although

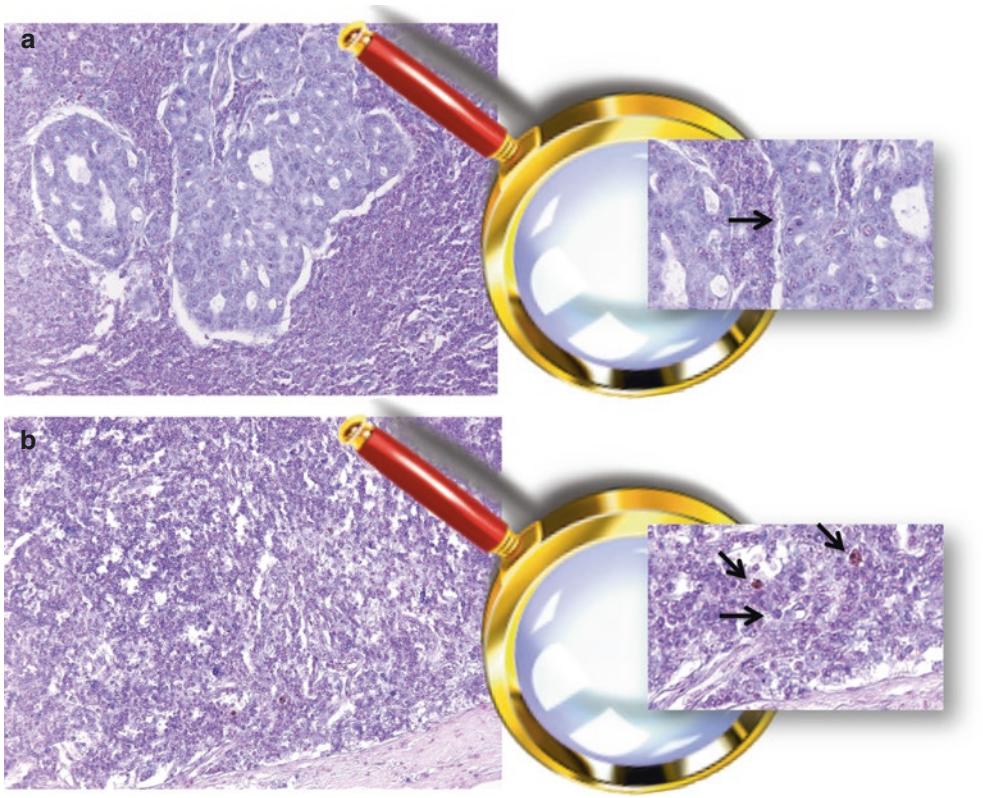


Fig. 5.3 miRNA staining using in situ hybridization (*ISH*) technique. Examples of ISH are provided. (**a**). let-7i expression in breast cancer (**b**). miR-194 expression in breast cancer. ISH staining for the target miRNAs is shown in *brown* in the *small box* (higher magnification) and in the *bigger box* (lower magnification) being marked with black arrows

microarrays enhance the throughput of miRNA analysis with a high sensitivity and specificity, microarray-derived findings must be confirmed using qRT-PCR assays [125–127, 130]. A comprehensive workflow for microarray data production, analysis, and interpretation is provided in Fig. 5.4.

Despite the large amount of experimental data obtained from DNA microarray studies evaluating miRNA, only a few studies have focused on intraplatform and interplatform correlations [93]. The reproducibility of data obtained using different miRNA-detection technologies with a focus on comparing the sensitivity and specificity of microarray platforms has also rarely been explored [106, 125]. This is particularly important given the fact that such studies have pointed to a poor correlation of mRNA gene expression among different microarray platforms in terms of reproducibility and equivalence. This is in part due to the lack of optimization of microarray protocols and the deficiency in complete and accurate data annotation of the commercially available platforms, lack of correct probe matching among technologies, discrepancies in data normalization, and interlaboratory or intralaboratory variability in technical expertise [130].

The short length of miRNA sequences represents the main drawback for primer design, selection of specifically labeled probes, and, generally, optimization of reactions for the concomitant detection of collective miRNAs by microarray technologies [14, 126]. Also, because miRNAs are short, even small variations in their length and/or their GC nucleotide content have major consequences on their biochemical properties, which are particularly visible at the melting temperature (T_m) of the miRNA in a hybridization reaction [14].

Commercially available microarray platforms for the analysis of miRNA expression are listed in Table 5.3. Selecting a microarray platform is difficult, as each platform has benefits and limitations in printing and surface technology, slide design, labeling and detection chemistry (one or two colors), hybridization conditions, probe design, and cost per sample [126]. Microarray has the advantage of a continuous updating of the new transcripts and remains a good alternative for miRNome profile. Most of microarray platforms have their design based on latest MirBase released variant [131].

Affymetrix has developed a microarray that includes miRNAs from several species [127]. This platform and that developed by Illumina contain probes for human and murine species [127, 129]. The Exiqon platform uses LNAs to increase the specificity [127]. Agilent Technologies offers a microarray design with considerable flexibility in analysis [129]. Microarray technologies are especially suitable for high-throughput miRNA expression profiling given the large number of samples that can be analyzed [7]. Recently a miRNA 3D array (Toray, Tokyo, Japan) was launched, adding a new dimension to biomarker profiling with good data reproducibility [131]. The 3D aray can be successfully applied to FFPE tissue.

Another innovative approach based on hybridization is represented by the NanoString nCounter that utilizes a multiplexed probe library, generated based on two

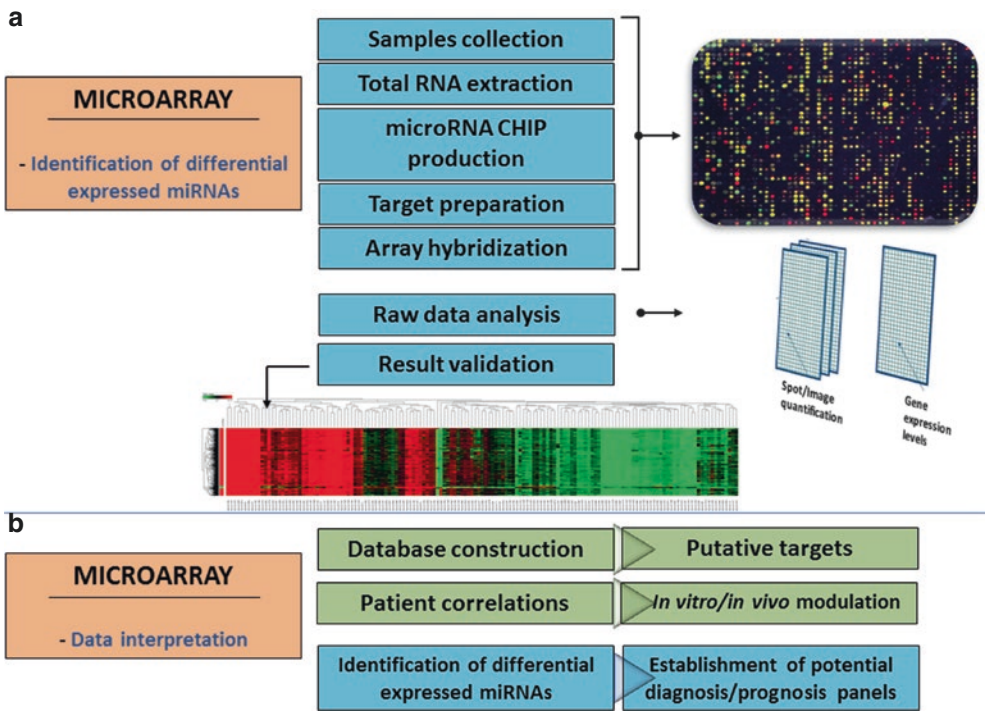


Fig. 5.4 Comprehensive workflow for microarray data production, analysis, and interpretation. Obtaining a differentially expressed miRNA list through the use of microarray technology is possible following the workflow shown in (**a**). The identification and validation of targets for these differentially expressed miRNAs follow the workflow shown in (**b**)

Table 5.3 Microarray technology platforms for miRNA expression

No.	Technology	Probe technology	Labeling technology	Detection chemistry	miRbase version/ number of human miRNAs	Quantity (ng)	References
1	Affymetrix GeneChip miRNA 4.0 Array	High-density arrays of 25 oligonucleotide probes	FlashTag™ biotin RNA labeling kit (Genisphere)	One color	20/1733 human,	130–500	[73, 80, 127]
2	Illumina, BeadChip miRNA arrays v2	Biotin-labeled oligo-dT primer with a universal sequence at the 5′ end	Adaptation of the DASL (cDNA-mediated annealing, selection, extension, and ligation)	One color	12/1146 human and 656 mouse	40–200	[80, 128]
3	Agilent, miRNA microarray	60-mer SurePrint technology	Agilent labeling (v1) uses CIP (GE) and T4 RNA ligase	One color	21/2549 human miRNA	100	[129, 145]
4	Exiqon, human miRCURY LNA miRNA	LNA-based capture probes	miRCURY LNA microRNA, hi-power labeling kit, Hy3/Hy5	Two color	20/3100 human, mouse, and rat	30	[73, 145]
5	Invitrogen, the NCode human miRNA microarray V3	34–44 complementary oligonucleotides	NCode™ rapid miRNA labeling system	Two color	10/380 human	30–1000	[73, 145]
6	3D-gene microarray	20–25 complementary oligonucleotides	3D-gene® scanner 1000	Two color	21/2565	1000 ng	[131]
7	NanoString		NanoString nCounter	Four color	21/800	1–100 ng	[10, 132]

sequence-specific capture probes designed to target a specific miRNA. This method allows the discrimination between related variants with high precision [10]. It can be also used with success for study of the tumor microenvironment or single cell analysis from FFPE, reducing the issues of RNA degradation or limited sample quantity and evaluating over 800 transcripts in a single reaction [132].

Next-Generation Sequencing (NGS)

NGS approaches have increased the number of developed application for discovery of novel biomarkers. NGS-based miRNA profiling provides important methodological benefits such as higher throughput and reduced RNA input. It allows for a superior consistency and quality of data,

avoiding batch effects related to hybridization. NGS miRNA profiling also allows for a wider range of detection limit [133, 134].

NGS is used for the assessment of known miRNA sequence patterns and for identifying novel miRNA species, which traditional approaches are incapable of detecting [10, 135, 136]. Sequencing miRNAs provides information not only about the levels of expression but also about single-nucleotide polymorphisms (SNPs), posttranscriptional RNA modifications particularly in 3′-terminal regions extension with a single nucleotide, and variations in miRNA length. Posttranscriptional modifications of miRNAs produce multiple mature miRNA variants, which are referred to as isomiRs [10, 137]. Presently several RNA-Seq analysis tools are available for miRNA or isomiR analysis and prediction of their interaction with mRNA [138, 139], These are integrated in databases like IsomiR Bank that furnish valuable information related to target prediction and enrichment analysis [140].

High-throughput sequencing for miRNA expression profiling is currently performed on several commercial NGS platforms including SOLiD (Applied Biosystems), HiSeq2000 (or Genome Analyzer IIx; Illumina), and GS FLX+ 454 (Roche) as well as smaller-scale NGS platforms, such as Ion Torrent and Ion Proton (Applied Biosystems) and MiSeq (Illumina). Sequencing technologies have also been applied to miRNA analysis but are expensive and not widely accessible. NGS platforms, which use less than 1 μg of total RNA, start with the preparation of a cDNA library [136, 141] from total RNA and miRNA samples [141] followed by massively parallel sequencing of the millions of distinct DNA molecules in the library [142]. They are able to generate millions of short sequence reads in order to provide miRNA profiles or to identify novel miRNAs [135, 137].

NGS platforms for miRNA high-throughput sequencing and novel miRNA characterization have high sensitivity and resolution and can differentiate among highly similar sequences, such as isomiRs. The main limitation of these approaches is that they are not suitable for absolute quantification of miRNA amounts. They also require highly complex bioinformatics analysis of sequence reads. Specialized bioinformatics software programs as miRDeep [104] and miRExpress [135, 143] and Web-based tools such as miRanalyzer [135, 143] and miRCat [144] are available for miRNA expression profiling as well as discovery [135] making the identification of known and novel miRNAs an easier process.

As a summary, Table 5.4 lists the main principles used in miRNA assays and techniques, and Table 5.5 lists the main platforms used for evaluation of miRNA expression patterns along with their strengths and limitations.

Examination of miRNAs is considerably more complex and technically demanding compared to that of mRNAs due to several factors: short miRNA lengths, difficulty in distinguishing different miRNA types (pre-miRNA, pri-miRNA, and mature miRNA), fluctuating $T_m s$ for primers and probes, different RNA ligase conditions, and high homology between probes among miRNA families.

Data Processing and Normalization

Prior to data normalization, data preprocessing for miRNA pattern analysis is required and comprises a set of corrections of values specific for the platform used: baseline corrections and threshold settings for qRT-PCR approaches, background adjustment for microarrays, and/or screening for small amounts of RNA-sequencing data [145]. Subsequent to these initial steps, data interpretation requires selection of the optimal normalization strategy for proper estimation of the biological variation among samples to avoid systematic and technical errors [146].

Statistical tools are important in any high-throughput technology, including the examination of miRNA species. The statistical analyses needed differ according to the technology used. The analysis should be carried out with consideration of the biochemistry, biological material characteristics [145, 146], and intrinsic limitations of each miRNA profiling platform [146, 147]. Regarding microarrays, to reduce the overall variance in data interpretation, researchers developed different normalization methods for specific miRNA microarray platforms, taking into account the type of sample, method of RNA extraction, dye labeling (one or two colors), hybridization and washing conditions, and efficacy of scanning [145–147].

Raw data analysis, ideally, should be performed by two distinct bioinformatician using two independent methods of analysis. Using this approach, we were able, for example, to identify a unique miRNA signature associated with prognostic factors and disease progression in patients with B-cell CLL [29]. Another strategy for analyzing raw data was generated by microarray images using the GenePix Pro software program (Molecular Devices). For a detailed review, we suggest reading Volinia et al. [148] as well as multiple additional publications describing various methods [149–152].

The precision of miRNA analysis is dependent on proper data normalization. A common method of normalization of qRT-PCR inputs is the use of constant endogenous controls or reference miRNAs, such as RNU6, RNU6B, RNU44, and RNU48 [147], or exogenous normalizers (cell-39 or cell-54 spike-in). When using an endogenous control, the standardized approach is applying the $2 - \Delta^{CT}$ method [147] with the formula $\Delta CT = CT_{miRNA} - CT_{endogenous\ control}$.

Presently the literature shows a high degree of inconsistency among different platforms due to different extraction protocols with different rates of recovery. Increased accuracy

Table 5.4 Principles of miRNA assays and technologies

Assay		Technologies	Method principle	Applications	References
miRNA extraction protocols		TRIzol/TRI reagent	Single-extraction assay for DNA, RNA, and proteins; chloroform extraction and ethanol precipitation of RNA—Recovers small RNA	Fresh frozen tissue, cell cultures	[10, 20, 23, 94]
		Column commercial kits	Solid-phase extraction (SPE)	FFPE samples	
		Magnetic beads	Magnetic microspheres for binding small RNA species, purification in magnetic field	Low concentration of miRNA (serum/plasma, saliva, or urine)	
miRNA purification protocols		Column commercial kits	SPE	Enrich miRNA fraction from biological specimens, especially for NGS applications	[10, 20, 24]
		Magnetic beads	Magnetic microspheres for binding small RNA species, purification in magnetic field		
qRT-PCR	RNA → cDNA reverse transcription, specific amplification for target miRNA	Stem loop qRT-PCR and TaqMan qRT-PCR	Polyadenylation of the miRNA at the 3′ end; reverse transcription reactions are based on stem-loop primers specific for 3′ end of microRNA. Then the amplicons are synthetized based on specific forward and reverse primer	Validation of microarray and NGS	[37, 38, 42, 118, 119, 130]
		Poly(T) adaptor primer for poly(A) miRNA and SYBR green qRT-PCR	Polyadenylated miRNA to cDNA and cDNA detection by qRT-PCR using technology	Discriminates between miRNAs that differ by as few as a single nucleotide	
NanoString	Hybridization	nCounter Prep Station and nCounter Digital Analyzer	Hybridization using a multiplexed probe library using two sequence-specific capture probes for miRNA of interest	Profiling and validation tool of the microarray and NGS data	[10, 132]
Microarrays	RNA → cDNA, cDNA labeling, hybridization with solid-phase probes, washing, detection of signal, data analysis	Affymetrix GeneChip miRNA 3.0 array	High-density arrays of 25 oligonucleotide probes	Genome-wide miRNA expression pattern for multiple samples can be processed in parallel in a standardized protocol, analysis less difficult than NGS	[10, 98, 145]
		Illumina BeadChip miRNA arrays v2	Biotin-labeled oligo-dT primer with a universal sequence at the 5′ end and adaptation of the DASL (cDNA-mediated annealing, selection, extension, and ligation)		
		Agilent, miRNA microarray	60-mer SurePrint technology and Agilent labeling (v1) uses CIP (GE) and T4 RNA ligase		
		Exiqon human miRCURY LNA miRNA	LNA-based capture probes and miRCURY LNA microRNA, hi-power labeling kit, Hy3/Hy5		
		Invitrogen—The NCode human miRNA microarray V3	34–44 complementary oligonucleotides and NCode™ rapid miRNA labeling system		

(continued)

Table 5.4 (continued)

Assay		Technologies	Method principle	Applications	References
NGS—High-throughput sequencing platforms	Library generation (miRNA transcriptomic profile, genome/DNA of interest, fragmentation, adaptor ligation, PCR amplification), simultaneous sample and amplification product sequencing	*SOLiD* (applied biosystems)	Ligation-based chemistry with dibase-labeled probes	miRNA expression pattern and novel miRNA to sequence millions of fragments; simultaneously in system multiplex; differentiate among similar sequences like isomiRs	[93, 94, 135, 136, 142]
		HiSeq 2000 (Illumina)	Sequencing by synthesis, using *HiSeq2000* (or genome analyzer IIx from Illumina)		
		GS FLX + 454 sequencing (Roche)	Pyrosequencing		
Smaller-scale next-generation sequencing platforms		*Ion Torrent* and *Ion Proton* (applied biosystems)	Sequencing chemistry is based on the proton release when a nucleotide is incorporated by the polymerase in the DNA molecule, resulting in a detectable local change of pH		
		MiSeq (Illumina)	When a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a by-product		

Table 5.5 Technology platforms for miRNA expression analysis: advantages and disadvantages

Technology	Advantages	Limitations	References
Northern blotting	"Gold standard"	Low sensitivity and reproducibility, requires a large amount of biological material	[32, 112]
Real-time PCR	Low sample amount, high sensitivity and sequence specificity; discriminates between pre- and mature miRNA	High price	[34, 35, 37, 38, 42, 44, 118, 119]
In situ hybridization	Cellular localization	Low sensitivity, background effect	[37, 38, 71, 120, 122, 124]
Microarray	High-throughput assay	Low sensitivity; it requires data validation; low price	[98, 145]
Next-generation sequencing	High-throughput assay; discovering novel or rare miRNAs	High price	[135, 142] [6, 56, 57, 59]
NanoString	High sensitivity, no amplification or cloning, multiplexing	High price	[10, 132]

and reliability of the experimental data can be achieved by selection of proper normalization method. The utilization of multiple stable normalizers will be useful for elimination of systematic bias in order to ensure validity of differences detected among evaluated groups or to monitor the response to therapy in preclinical trials [153].

The high variation may be related to variation in the expression level of a particular miRNA in tumor tissue among different studies or the differential expression level among pathological tissue and the expression level in biological fluids like serum/plasma or urine. A classic example is miR-16, demonstrated to be downregulated in CLL [28] but recently proposed as normalizer [153].

Analysis of miRNA Function and Gene Interaction

miRNAs primarily function by inhibiting translation or degrading the target mRNA or by altering the mRNA stability. This is facilitated by the partial sequence homology of the miRNA seed sequence with the 3′ untranslated regions of the target mRNAs [154, 155]. Because of this unique feature, an individual miRNA can have multiple targets and regulate a large number of protein-coding genes [156].

As master gene regulators, miRNAs can impact a variety of cellular pathways and functions [157]. Consequently, miRNAs regulate diverse biological processes that are critical to development, cell death, proliferation, and differentiation. Likewise, given their presence in all eukaryotic cells, miRNAs are implicated in the deregulation of multiple pathways, leading to a variety of diseases in humans, animals, and plants [155, 156]. Thus, understanding how miRNA expression is regulated in normal as well as disease-specific cellular processes is critical. For example, a variety of miRNAs play important roles in various aspects of cellular immunity. miR-29 family members play critical roles in

determining the molecular bases for innate and adaptive immune responses toward intracellular bacterial infections [155]. Other examples are miR-208a, miR-208b, and miR-499, which, along with MHC genes, have important functions in the formation of a regulatory circuit that controls cardiac hypertrophy and leads to heart failure [158, 159].

Interactions between miRNAs and biologically important targets can be direct (sequence complementarity) or indirect (e.g., via a transcription factor influenced by the miRNA). Therefore, the step of target identification is one of the most important ones for the biological characterization of a miRNA role in the pathogenic mechanisms of a specific disease. The best approach for target identification is to confirm the negative expression correlation using protein samples obtained from the same study subjects who had miRNA expression profiled. If a negative correlation is found between array miRNA expression and Western blot protein expression, it should be confirmed by transfecting the miRNA of interest in at least two cell lines that express the protein that is thought to be a putative target of that miRNA. To date, investigators have mainly found negative correlations between miRNA expression and their targets (miRNA inhibition of transcription or translation). This does not, however, preclude the existence of positive correlations. The finding that miR-122 has a positive influence on replication of hepatitis C virus by interacting with the 5′ noncoding region of the virus is an indication of such positive correlations [160].

miRNA Mimics and Inhibitors

Manipulation of mRNAs with specific oligonucleotides that mimic miRNAs provides new opportunities for a better understanding of the molecular mechanisms of diseases, opening up new avenues of research aimed at treatment of pathologies. The concept that the cause of a specific pathogenesis is solely based on alteration of the expression of protein-coding genes is no longer tenable with the discovery of miRNAs. Pathogenic miRNA alterations can be reversed by positively or negatively modulating the expression of miRNAs, highlighting the potential of these small molecules as pharmacological targets.

Because of their small size, miRNAs are easier to transfer than other molecules, such as DNA or mRNA [161]. Therapeutic manipulation of miRNAs is primarily divided into two categories: (i) downregulation of miRNAs that inhibit the expression of genes involved in pathological impairment and (ii) upregulation of miRNAs that normally suppress the action of disease-promoting genes. In this sense, several therapeutic strategies have been developed in order to increase the efficiency of the exogenous miRNAs, with a marked domination of experimental designs in malignant pathologies. AntagomiRs, anti-miRs, agonist microRNAs, circular RNA (ciRNA) under the form of miRNA sponges, and miRNA masks are novel chemically designed oligonucleotides that prevent the interaction of miRNAs with their target mRNA molecules. Anti-miR oligonucleotides represent the most widely used therapeutic molecules and have been subjected to several modifications in order to prolong their effect within the cell. One such strategy consists in addition of functional groups like methylated hydroxyl groups (2′-OMe), phosphorothioate, methoxyethyl, and fluorine (2′-F). AntagomiRs are similar in action to anti-miR oligonucleotides, the difference consisting in the prior conjugation with cholesterol upon administration in order to prolong the stability of the exogenous sequence. Other inhibitory approaches include impairment of the enzymatic processing steps from miRNA biogenesis or attacking of incompletely processed sequences: pre- and pri-miRNA. miRNA mimics have the opposite effect, mimicking the effect of miRNAs in the cells. Both antagomiRs and miRNA mimics can be readily introduced into cell lines via cold transfection using electroporation or viral and nonviral vectors [162].

The latest class of ncRNAs discovered, ciRNA can be used as therapeutic tools for targeting overexpressed miRNAs, due to their capacity to modulate gene expression, with important effects on pathological processes [163]. There is the case of circular RNA hsa_circ_0001564 that promotes tumorigenesis by sponging of miR-29c [164]. circMTO1 have the capacity to sponge miR-9, leading to the inhibition of hepatocellular carcinoma progression [163].

Viruses that are widely used for miRNA transfer include retroviruses, lentiviruses, adenoviruses, adeno-associated viruses, and herpes simplex viruses [165]. Their favorable rate of cell uptake and gene and/or tissue specificity has led to their wide use as transfer vectors. Their main drawbacks are production difficulties and limited administration capacity due to host-induced acute inflammatory responses [166]. The antitumoral effect achieved with the administration of adeno-associated viral vector scAAV8 expressing miR-26a, in transgenic mice with conditional expression of Myc in the liver, illustrates the potential role of miRNA in therapeutics [167]. In another study, overexpression of miR-133 delivered into a mouse model using an adenovirus vector led to a significant reduction in the number of cardiac myocytes in the left ventricle and increased expression of fetal genes pointing to miR-133 as a therapeutic target in cardiac hypertrophy [80].

Nonviral miRNA delivery systems are constructed from synthetic or natural compounds. These systems are generally less toxic and less immunogenic than viral systems, and they can be easily secured and repeatedly utilized. Their main disadvantages are reduced efficiency and short expression durations. For example, Rai et al. [168] used a liposomal system carrying miR-7 to block the expression of mutant epidermal growth factor receptor (EGFR) in lung cancer cells.

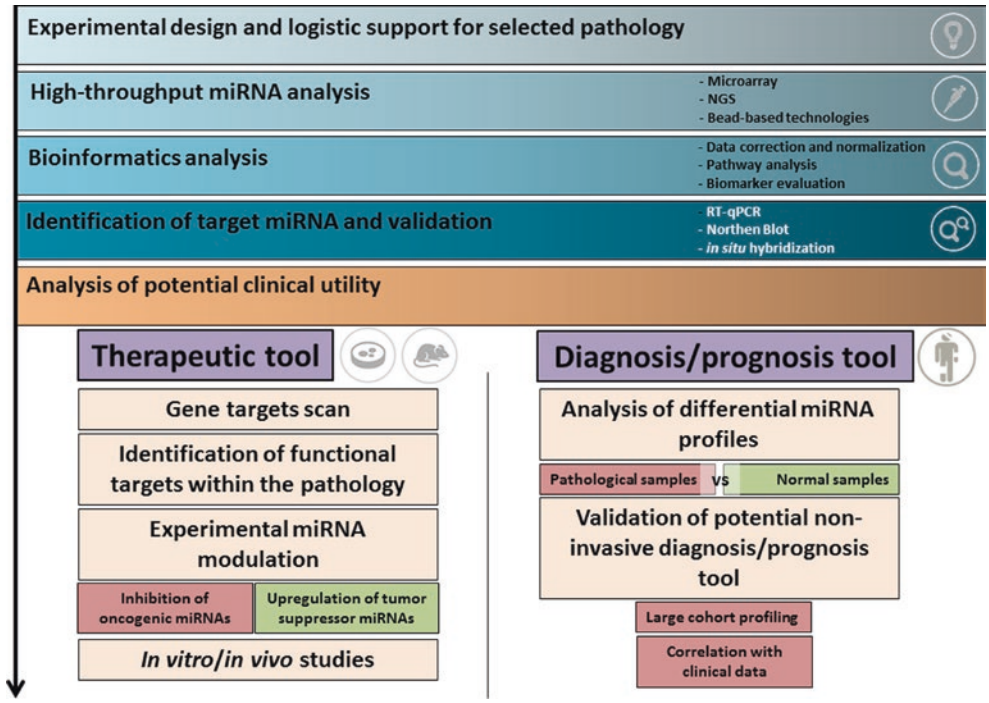


Fig. 5.5 Experimental design and logistic support for evaluation of miRNAs in a selected pathology. A comprehensive experimental design and logistic support start with sample processing, quality control, and high-throughput miRNA expression pattern and bioinformatics analysis. Once a candidate miRNA is identified, a validation process takes place, and functional studies are aimed at the elucidation of diagnostic, prognostic, or therapeutic roles for the found miRNA

The main obstacle for clinical applications of such treatment strategies remains the limited efficiency in miRNA delivery. Over the years, researchers have developed many delivery systems, but the focus is still on emerging concepts for the delivery of therapeutic miRNA mimics and antagomiRs [169].

As mentioned above, miRNAs inhibit the expression of genes relevant to different biological and pathological processes. Not surprisingly, alteration of the miRNA profile has been linked to multiple human illnesses, including cancer and cardiovascular and autoimmune diseases [170–172]. In this sense, miRNAs continue to have a strong potential as biomarkers for diagnosis [107, 173, 174], prognosis, and therapeutic response prediction [174, 175] (Fig. 5.5).

Conclusions

miRNAs, first discovered over two decades ago, have proven to play a major role in gene regulation and transcription in normal as well as abnormal cells and in many different diseases. Cancer, cardiovascular diseases, immune system deregulation, and adipogenesis, all represent platforms for investigation into miRNA communication and molecular profiling in order to better establish an early diagnostic and a therapeutic option for different patient subgroups. From a pathologists' point of view, miRNA's discovery was a breakthrough for diagnosis, prognostication, and therapy prediction. Bringing these molecules and all their complexity from the bench to the bedside is, at present, a big challenge. The investigative frontier in miRNA biology is currently centered on the detection of novel structures of miRNAs, identification of the molecular targets, and potential as yet unknown roles.

Investigative discoveries in miRNAs that could be applied to clinical practice are increasingly reported albeit mainly in the cancer field. Important examples are miR-15a and miR-16-1 downregulation in CLL; miR-155 expression levels and miR-21 panels as powerful biomarkers in CLL; association of miR-21 with important clinicopathological characteristics in non-small cell lung cancer, colorectal cancer, and invasive ductal carcinomas of the breast; and association of the miR-17-92 family with many solid tumors, lymphomas, and myelomas.

From the therapeutic perspective, based on clinical trial data, controlling the levels of a single miRNA has resulted in only a limited effect on the expression level for a target gene. The latter could be due to the activation of compensatory pathways or an inefficient delivery. In order to limit the activation of the immune response-related genes, (see miR-34 cytokine storm), the capacity of miRNAs to target multiple genes needs to be controlled in order to limit unwanted sides effects [176].

The changes in miRNA pattern have relevant biological significance and may provide useful information for identification of the early stages of disease, for clinical diagnoses, or for the identification of therapeutic targets with important implications in personalized treatment.

By using ISH, the pathologist is able to use FFPE or fresh tissue samples with high-performance quality and reproducible results. This has made it possible to refine and validate the data obtained by microarrays and to localize particular miRNAs in a given tissue. This new technique indicates the remarkable stability of miRNAs in archival human tissues. Additional functional genomic techniques are still necessary to profile the upregulated and downregulated miRNAs in a patient. Microarray technology can be used for the assessment of miRNAs in normal and tumor tissues. As the expression profiles and function of miRNAs during disease development and progression are further elucidated using NGS, their role as biomarkers and therapeutic targets will likely continue to increase.

Acknowledgments This work was supported by the Competitively Operational Program, 2014–2020, entitled "Clinical and economical impact of personalized targeted anti-microRNA therapies in reconverting lung cancer chemoresistance" (CANTEMIR), grant no. 35/01.09.2016, MySMIS-103375.

Dr. Calin is the Alan M. Gewirtz Leukemia & Lymphoma Society Scholar. He is supported as a fellow of the University of Texas MD Anderson Cancer Center Research Trust, as the University of Texas System Regents Research Scholar, and by the CLL Global Research Foundation. Work in Dr. Calin's laboratory is supported in part by the National Institutes of Health/National Cancer Institute (CA135444); Department of Defense Breast Cancer Idea Award; Developmental Research Awards in Breast Cancer, Ovarian Cancer, Brain Cancer, Prostate Cancer, Multiple Myeloma, Leukemia (P50 CA100632), and Head and Neck (P50 CA097007) Specialized Program of Research Excellence grants; Sister Institution Network Fund grants in CLL and colon cancer; the Laura and John Arnold Foundation; the RGK Foundation; and the Estate of C. G. Johnson, Jr. This research is supported in part by the MD Anderson Cancer Center Support Grant CA016672.

References

1. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74.
2. Enfield KSS, Pikor LA, Martinez VD, Lam WL. Mechanistic roles of noncoding RNAs in lung cancer biology and their clinical implications. Genet Res Int. 2012;2012:16.
3. Shah MY, Calin GA. The mix of two worlds: non-coding RNAs and hormones. Nucleic Acid Ther. 2013;23(1):2–8.
4. Redis RS, Calin S, Yang Y, You MJ, Calin GA. Cell-to-cell miRNA transfer: from body homeostasis to therapy. Pharmacol Ther. 2012;136(2):169–74.
5. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998;391(6669):806–11.
6. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–97.
7. Negrini M, Nicoloso MS, Calin GA. MicroRNAs and cancer–new paradigms in molecular oncology. Curr Opin Cell Biol. 2009;21(3):470–9.
8. Eastlack S, Alahari S. MicroRNA and breast cancer: understanding pathogenesis, improving management. Non-Coding RNA. 2015;1(1):17.
9. Irimie AI, Braicu C, Cojocneanu-Petric R, Berindan-Neagoe I, Campian RS. Novel technologies for oral squamous carcinoma biomarkers in diagnostics and prognostics. Acta Odontol Scand. 2015;73(3):161–8.
10. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. Nat Rev Genet. 2012;13(5):358–69.
11. Berindan-Neagoe I, Monroig Pdel C, Pasculli B, Calin GA. MicroRNAome genome: a treasure for cancer diagnosis and therapy. CA Cancer J Clin. 2014;64(5):311–36.
12. Braicu C, Catana C, Calin GA, Berindan-Neagoe I. NCRNA combined therapy as future treatment option for cancer. Curr Pharm Des. 2014;20(42):6565–74.
13. Catana CS, Calin GA, Berindan-Neagoe I. Inflamma-miRs in aging and breast cancer: are they reliable players? Front Med. 2015;2:85.
14. Etheridge A, Lee I, Hood L, Galas D, Wang K. Extracellular microRNA: a new source of biomarkers. Mutat Res. 2011;717(1–2):85–90.
15. Krichevsky AM. MicroRNA profiling: from dark matter to white matter, or identifying new players in neurobiology. Sci World J. 2007;7:155–66.
16. Berindan-Neagoe I, Calin GA. Molecular pathways: microRNAs, cancer cells, and microenvironment. Clin Cancer Res. 2014;20(24):6247–53.
17. Braicu C, Calin GA, Berindan-Neagoe I. MicroRNAs and cancer therapy – from bystanders to major players. Curr Med Chem. 2013;20(29):3561–73.
18. Bommer GT, Gerin I, Feng Y, Kaczorowski AJ, Kuick R, Love RE, Zhai Y, Giordano TJ, Qin ZS, Moore BB, et al. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. Curr Biol. 2007;17(15):1298–307.
19. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. Cell. 2012;148(6):1172–87.
20. Shi XB, Xue L, Yang J, Ma AH, Zhao J, Xu M, Tepper CG, Evans CP, Kung HJ, deVere White RW. An androgen-regulated miRNA suppresses Bak1 expression and induces androgen-independent growth of prostate cancer cells. Proc Natl Acad Sci U S A. 2007;104(50):19983–8.
21. Cao P, Deng Z, Wan M, Huang W, Cramer SD, Xu J, Lei M, Sui G. MicroRNA-101 negatively regulates Ezh2 and its expression is modulated by androgen receptor and HIF-1alpha/HIF-1beta. Mol Cancer. 2010;9:108.
22. Poliseno L, Salmena L, Riccardi L, Fornari A, Song MS, Hobbs RM, Sportoletti P, Varmeh S, Egia A, Fedele G, et al. Identification of the miR-106b~25 microRNA cluster as a proto-oncogenic PTEN-targeting intron that cooperates with its host gene MCM7 in transformation. Sci Signal. 2010;3(117):ra29.
23. Zaman MS, Thamminana S, Shahryari V, Chiyomaru T, Deng G, Saini S, Majid S, Fukuhara S, Chang I, Arora S, et al. Inhibition of PTEN gene expression by oncogenic miR-23b-3p in renal cancer. PLoS One. 2012;7(11):e50203.
24. Mlcochova H, Hezova R, Stanik M, Slaby O. Urine microRNAs as potential noninvasive biomarkers in urologic cancers. Urol Oncol. 2014;32(1):41.e41–9.
25. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999–3004.
26. Yang BF, Lu YJ, Wang ZG. MicroRNA and disease associations. Clin Exp Pharmacol Physiol. 2009;36(10):951–60.
27. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524–9.

28. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102(39):13944–9.
29. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793–801.
30. Pekarsky Y, Santanam U, Cimmino A, Palamarchuk A, Efanov A, Maximov V, Volinia S, Alder H, Liu CG, Rassenti L, et al. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. Cancer Res. 2006;66(24):11590–3.
31. Calin GA, Pekarsky Y, Croce CM. The role of microRNA and other non-coding RNA in the pathogenesis of chronic lymphocytic leukemia. Best Pract Res Clin Haematol. 2007;20(3):425–37.
32. Rossi S, Shimizu M, Barbarotto E, Nicoloso MS, Dimitri F, Sampath D, Fabbri M, Lerner S, Barron LL, Rassenti LZ, et al. microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. Blood. 2010;116(6):945–52.
33. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6(11):857–66.
34. Wang JL, Hu Y, Kong X, Wang ZH, Chen HY, Xu J, Fang JY. Candidate microRNA biomarkers in human gastric cancer: a systematic review and validation study. PLoS One. 2013;8(9):e73683.
35. Wang Y, Li J, Tong L, Zhang J, Zhai A, Xu K, Wei L, Chu M. The prognostic value of miR-21 and miR-155 in non-small-cell lung cancer: a meta-analysis. Jpn J Clin Oncol. 2013;43(8):813–20.
36. Faruq O, Vecchione A. microRNA: diagnostic perspective. Front Med. 2015;2:51.
37. Menendez P, Padilla D, Villarejo P, Palomino T, Nieto P, Menendez JM, Rodriguez-Montes JA. Prognostic implications of serum microRNA-21 in colorectal cancer. J Surg Oncol. 2013;108(6):369–73.
38. Lee JA, Lee HY, Lee ES, Kim I, Bae JW. Prognostic implications of MicroRNA-21 overexpression in invasive ductal carcinomas of the breast. J Breast Cancer. 2011;14(4):269–75.
39. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, Yoshida Y, Seto M. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. 2004;64(9):3087–95.
40. Goblirsch M, Richtig G, Slaby O, Berindan-Neagoe I, Gerger A, Pichler M. MicroRNAs as a tool to aid stratification of colorectal cancer patients and to guide therapy. Pharmacogenomics. 2017;18(10):1027–38.
41. Leichter AL, Sullivan MJ, Eccles MR, Chatterjee A. MicroRNA expression patterns and signalling pathways in the development and progression of childhood solid tumours. Mol Cancer. 2017;16:15.
42. Chang CC, Yang YJ, Li YJ, Chen ST, Lin BR, Wu TS, Lin SK, Kuo MY, Tan CT. MicroRNA-17/20a functions to inhibit cell migration and can be used a prognostic marker in oral squamous cell carcinoma. Oral Oncol. 2013;49(9):923–31.
43. Gao X, Zhang R, Qu X, Zhao M, Zhang S, Wu H, Jianyong L, Chen L. MiR-15a, miR-16-1 and miR-17-92 cluster expression are linked to poor prognosis in multiple myeloma. Leuk Res. 2012;36(12):1505–9.
44. Yu G, Tang JQ, Tian ML, Li H, Wang X, Wu T, Zhu J, Huang SJ, Wan YL. Prognostic values of the miR-17-92 cluster and its paralogs in colon cancer. J Surg Oncol. 2012;106(3):232–7.
45. Butz H, Nofech-Mozes R, Ding Q, Khella HWZ, Szabo PM, Jewett M, Finelli A, Lee J, Ordon M, Stewart R, et al. Exosomal microRNAs are diagnostic biomarkers and can mediate cell-cell communication in renal cell carcinoma. Eur Urol Focus. 2016;2(2):210–8.
46. Hanke M, Hoefig K, Merz H, Feller AC, Kausch I, Jocham D, Warnecke JM, Sczakiel G. A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. Urol Oncol. 2010;28(6):655–61.
47. Van Roosbroeck K, Fanini F, Setoyama T, Ivan C, Rodriguez-Aguayo C, Fuentes-Mattei E, Xiao L, Vannini I, Redis RS, D'Abundo L, et al. Combining anti-miR-155 with chemotherapy for the treatment of lung cancers. Clin Cancer Res. 2017;23(11):2891–904.
48. Blower PE, Chung JH, Verducci JS, Lin S, Park JK, Dai Z, Liu CG, Schmittgen TD, Reinhold WC, Croce CM, et al. MicroRNAs modulate the chemosensitivity of tumor cells. Mol Cancer Ther. 2008;7(1):1–9.
49. Kovalchuk O, Filkowski J, Meservy J, Ilnytskyy Y, Tryndyak VP, Chekhun VF, Pogribny IP. Involvement of microRNA-451 in resistance of the MCF-7 breast cancer cells to chemotherapeutic drug doxorubicin. Mol Cancer Ther. 2008;7(7):2152–9.
50. Xu K, Liang X, Shen K, Sun L, Cui D, Zhao Y, Tian J, Ni L, Liu J. MiR-222 modulates multidrug resistance in human colorectal carcinoma by down-regulating ADAM-17. Exp Cell Res. 2012;318(17):2168–77.
51. Zhong S, Li W, Chen Z, Xu J, Zhao J. MiR-222 and miR-29a contribute to the drug-resistance of breast cancer cells. Gene. 2013;531(1):8–14.
52. Yu PN, Yan MD, Lai HC, Huang RL, Chou YC, Lin WC, Yeh LT, Lin YW. Downregulation of miR-29 contributes to cisplatin resistance of ovarian cancer cells. Int J Cancer. 2014;134(3):542–51.
53. Chen Y, Ke G, Han D, Liang S, Yang G, Wu X. MicroRNA-181a enhances the chemoresistance of human cervical squamous cell carcinoma to cisplatin by targeting PRKCD. Exp Cell Res. 2014;320(1):12–20.
54. Ke G, Liang L, Yang JM, Huang X, Han D, Huang S, Zhao Y, Zha R, He X, Wu X. MiR-181a confers resistance of cervical cancer to radiation therapy through targeting the pro-apoptotic PRKCD gene. Oncogene. 2013;32(25):3019–27.
55. Woyengo TA, Ramprasath VR, Jones PJ. Anticancer effects of phytosterols. Eur J Clin Nutr. 2009;63(7):813–20.
56. Yu X, Chen Y, Tian R, Li J, Li H, Lv T, Yao Q. miRNA-21 enhances chemoresistance to cisplatin in epithelial ovarian cancer by negatively regulating PTEN. Oncol Lett. 2017;14(2):1807–10.
57. Son JW. Year-in-review of lung cancer. Tuberc Respir Dis. 2012;73(3):137–42.
58. Kita Y, Vincent K, Natsugoe S, Berindan-Neagoe I, Calin GA. Epigenetically regulated microRNAs and their prospect in cancer diagnosis. Expert Rev Mol Diagn. 2014;14(6):673–83.
59. Strmsek Z, Kunej T. MicroRNA silencing by DNA methylation in human cancer: a literature analysis. Non-Coding RNA. 2015;1(1):44.
60. Irimie AI, Ciocan C, Gulei D, Mehterov N, Atanasov AG, Dudea D, Berindan-Neagoe I. Current insights into oral cancer epigenetics. Int J Mol Sci. 2018;19(3):670.
61. Wong KY, Yim RL, Kwong YL, Leung CY, Hui PK, Cheung F, Liang R, Jin DY, Chim CS. Epigenetic inactivation of the MIR129-2 in hematological malignancies. J Hematol Oncol. 2013;6:16.
62. Lu CY, Lin KY, Tien MT, Wu CT, Uen YH, Tseng TL. Frequent DNA methylation of MiR-129-2 and its potential clinical implication in hepatocellular carcinoma. Genes Chromosomes Cancer. 2013;52(7):636–43.
63. Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A. 2007;104(40):15805–10.

64. Chen H, Xu Z. Hypermethylation-associated silencing of miR-125a and miR-125b: a potential marker in colorectal cancer. Dis Markers. 2015;2015:345080.
65. Wang Z, Chang C, Peng M, Lu Q. Translating epigenetics into clinic: focus on lupus. Clin Epigenetics. 2017;9:78.
66. Suzuki H, Maruyama R, Yamamoto E, Kai M. Epigenetic alteration and microRNA dysregulation in cancer. Front Genet. 2013;4:258.
67. Scott GK, Mattie MD, Berger CE, Benz SC, Benz CC. Rapid alteration of microRNA levels by histone deacetylase inhibition. Cancer Res. 2006;66(3):1277–81.
68. Roccaro AM, Sacco A, Jia X, Azab AK, Maiso P, Ngo HT, Azab F, Runnels J, Quang P, Ghobrial IM. microRNA-dependent modulation of histone acetylation in Waldenström macroglobulinemia. Blood. 2010;116(9):1506–14.
69. Gong A-Y, Eischeid AN, Xiao J, Zhao J, Chen D, Wang Z-Y, Young CYF, Chen X-M. miR-17-5p targets the p300/CBP-associated factor and modulates androgen receptor transcriptional activity in cultured prostate cancer cells. BMC Cancer. 2012;12:492.
70. Sacco J, Adeli K. MicroRNAs: emerging roles in lipid and lipoprotein metabolism. Curr Opin Lipidol. 2012;23(3):220–5.
71. Romao JM, Jin W, Dodson MV, Hausman GJ, Moore SS, Guan LL. MicroRNA regulation in mammalian adipogenesis. Exp Biol Med (Maywood, NJ). 2011;236(9):997–1004.
72. Ortega FJ, Mercader JM, Catalan V, Moreno-Navarrete JM, Pueyo N, Sabater M, Gomez-Ambrosi J, Anglada R, Fernandez-Formoso JA, Ricart W, et al. Targeting the circulating microRNA signature of obesity. Clin Chem. 2013;59(5):781–92.
73. McGregor RA, Choi MS. microRNAs in the regulation of adipogenesis and obesity. Curr Mol Med. 2011;11(4):304–16.
74. Chen Z, Shi H, Sun S, Xu H, Cao D, Luo J. MicroRNA-181b suppresses TAG via target IRS2 and regulating multiple genes in the Hippo pathway. Exp Cell Res. 2016;348(1):66–74.
75. Chu B, Wu T, Miao L, Mei Y, Wu M. MiR-181a regulates lipid metabolism via IDH1. Sci Rep. 2015;5:8801.
76. Li H, Chen X, Guan L, Qi Q, Shu G, Jiang Q, Yuan L, Xi Q, Zhang Y. MiRNA-181a regulates Adipogenesis by targeting tumor necrosis factor-α (TNF-α) in the porcine model. PLoS One. 2013;8(10):e71568.
77. Ouyang D, Xu L, Zhang L, Guo D, Tan X, Yu X, Qi J, Ye Y, Liu Q, Ma Y, et al. MiR-181a-5p regulates 3T3-L1 cell adipogenesis by targeting Smad7 and Tcf7l2. Acta Biochim Biophys Sin. 2016;48(11):1034–41.
78. Ono K, Kuwabara Y, Han J. MicroRNAs and cardiovascular diseases. FEBS J. 2011;278(10):1619–33.
79. Thum T, Catalucci D, Bauersachs J. MicroRNAs: novel regulators in cardiac development and disease. Cardiovasc Res. 2008;79(4):562–70.
80. Care A, Catalucci D, Felicetti F, Bonci D, Addario A, Gallo P, Bang ML, Segnalini P, Gu Y, Dalton ND, et al. MicroRNA-133 controls cardiac hypertrophy. Nat Med. 2007;13(5):613–8.
81. Yang B, Lin H, Xiao J, Lu Y, Luo X, Li B, Zhang Y, Xu C, Bai Y, Wang H, et al. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. Nat Med. 2007;13(4):486–91.
82. Oliveira-Carvalho V, da Silva MM, Guimaraes GV, Bacal F, Bocchi EA. MicroRNAs: new players in heart failure. Mol Biol Rep. 2013;40(3):2663–70.
83. Fichtlscherer S, Zeiher AM, Dimmeler S. Circulating microRNAs: biomarkers or mediators of cardiovascular diseases? Arterioscler Thromb Vasc Biol. 2011;31(11):2383–90.
84. Zhu J, Yao K, Wang Q, Guo J, Shi H, Ma L, Liu H, Gao W, Zou Y, Ge J. Circulating miR-181a as a potential novel biomarker for diagnosis of acute myocardial infarction. Cell Physiol Biochem. 2016;40(6):1591–602.
85. Hu R, O'Connell RM. MicroRNA control in the development of systemic autoimmunity. Arthritis Res Ther. 2013;15(1):202.
86. Meisgen F, Xu N, Wei T, Janson PC, Obad S, Broom O, Nagy N, Kauppinen S, Kemeny L, Stahle M, et al. MiR-21 is up-regulated in psoriasis and suppresses T cell apoptosis. Exp Dermatol. 2012;21(4):312–4.
87. Mi QS, He HZ, Dong Z, Isales C, Zhou L. microRNA deficiency in pancreatic islet cells exacerbates streptozotocin-induced murine autoimmune diabetes. Cell Cycle. 2010;9(15):3127–9.
88. Yoshizawa JM, Wong DT. Salivary microRNAs and oral cancer detection. Methods Mol Biol (Clifton, NJ). 2013;936:313–24.
89. Allegra A, Alonci A, Campo S, Penna G, Petrungaro A, Gerace D, Musolino C. Circulating microRNAs: new biomarkers in diagnosis, prognosis and treatment of cancer (review). Int J Oncol. 2012;41(6):1897–912.
90. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9(3):189–98.
91. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, Guenther SM, O'Leary JJ, Sheils O. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Biotechnol. 2007;7:36.
92. Liu A, Xu X. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. Methods Mol Biol (Clifton, NJ). 2011;724:259–67.
93. Nelson PT, Wang WX, Wilfred BR, Tang G. Technical variables in high-throughput miRNA expression profiling: much work remains to be done. Biochim Biophys Acta. 2008;1779(11):758–65.
94. Mraz M, Malinova K, Mayer J, Pospisilova S. MicroRNA isolation and stability in stored RNA samples. Biochem Biophys Res Commun. 2009;390(1):1–4.
95. Wang K, Yuan Y, Cho JH, McClarty S, Baxter D, Galas DJ. Comparing the MicroRNA spectrum between serum and plasma. PLoS One. 2012;7(7):e41561.
96. Becker C, Hammerle-Fickinger A, Riedmaier I, Pfaffl MW. mRNA and microRNA quality control for RT-qPCR analysis. Methods (San Diego, Calif). 2010;50(4):237–43.
97. Rio DC, Ares M Jr, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc. 2010;2010(6):pdb.prot5439.
98. Ach RA, Wang H, Curry B. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol. 2008;8:69.
99. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55(4):611–22.
100. Kim YK, Yeo J, Kim B, Ha M, Kim VN. Short structured RNAs with low GC content are selectively lost during extraction from a small number of cells. Mol Cell. 2012;46(6):893–5.
101. Yoo CE, Kim G, Kim M, Park D, Kang HJ, Lee M, Huh N. A direct extraction method for microRNAs from exosomes captured by immunoaffinity beads. Anal Biochem. 2012;431(2):96–8.
102. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW. Comparison of DNA quantification methods for next generation sequencing. Sci Rep. 2016;6:24067.
103. Bravo V, Rosero S, Ricordi C, Pastori RL. Instability of miRNA and cDNAs derivatives in RNA preparations. Biochem Biophys Res Commun. 2007;353(4):1052–5.
104. Pop LA, Pileczki V, Cojocneanu-Petric RM, Petrut B, Braicu C, Jurj AM, Buiga R, Achimas-Cadariu P, Berindan-Neagoe I. Normalization of gene expression measurement of tissue samples obtained by transurethral resection of bladder tumors. OncoTargets Ther. 2016;9:3369–80.

105. Kim SW, Li Z, Moore PS, Monaghan AP, Chang Y, Nichols M, John B. A sensitive non-radioactive northern blot method to detect small RNAs. Nucleic Acids Res. 2010;38(7):e98.
106. Wang B, Howel P, Bruheim S, Ju J, Owen LB, Fodstad O, Xi Y. Systematic evaluation of three microRNA profiling platforms: microarray, beads array, and quantitative real-time PCR array. PLoS One. 2011;6(2):e17167.
107. Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R, Bhattacharya A. Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. BMC Genomics. 2010;11:288.
108. Jensen SG, Lamy P, Rasmussen MH, Ostenfeld MS, Dyrskjot L, Orntoft TF, Andersen CL. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. BMC Genomics. 2011;12:435.
109. Sablok G, Milev I, Minkov G, Minkov I, Varotto C, Yahubyan G, Baev V. isomiRex: web-based identification of microRNAs, isomiR variations and differential expression using next-generation sequencing datasets. FEBS Lett. 2013;587(16):2629–34.
110. Magee R, Telonis AG, Cherlin T, Rigoutsos I, Londin E. Assessment of isomiR discrimination using commercial qPCR methods. Non-Coding RNA. 2017;3(2):18.
111. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol. 2004;5(3):R13.
112. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.
113. Setoyama T, Ling H, Natsugoe S, Calin GA. Non-coding RNAs for medical practice in oncology. Keio J Med. 2011;60(4):106–13.
114. de Planell-Saguer M, Rodicio MC. Analytical aspects of microRNA in diagnostics: a review. Anal Chim Acta. 2011;699(2):134–52.
115. Schwarzkopf M, Pierce NA. Multiplexed miRNA northern blots via hybridization chain reaction. Nucleic Acids Res. 2016;44(15):e12.
116. Kumar P, Johnston BH, Kazakov SA. miR-ID: a novel, circularization-based platform for detection of microRNAs. RNA (New York, NY). 2011;17(2):365–80.
117. Schmittgen TD, Jiang J, Liu Q, Yang L. A high-throughput method to monitor the expression of microRNA precursors. Nucleic Acids Res. 2004;32(4):e43.
118. Reichenstein I, Aizenberg N, Goshen M, Bentwich Z, Avni YS. A novel qPCR assay for viral encoded microRNAs. J Virol Methods. 2010;163(2):323–8.
119. Ro S, Park C, Jin J, Sanders KM, Yan W. A PCR-based method for detection and quantification of small RNAs. Biochem Biophys Res Commun. 2006;351(3):756–63.
120. Leung AK, Sharp PA. Function and localization of microRNAs in mammalian cells. Cold Spring Harb Symp Quant Biol. 2006;71:29–38.
121. Nass D, Rosenwald S, Meiri E, Gilad S, Tabibian-Keissar H, Schlosberg A, Kuker H, Sion-Vardy N, Tobar A, Kharenko O, et al. MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. Brain Pathol (Zurich, Switzerland). 2009;19(3):375–83.
122. Bustos-Sanmamed P, Laffont C, Frugier F, Lelandais-Briere C, Crespi M. Analyzing small and long RNAs in plant development using non-radioactive in situ hybridization. Methods Mol Biol (Clifton, NJ). 2013;959:303–16.
123. Obernosterer G, Leuschner PJ, Alenius M, Martinez J. Post-transcriptional regulation of microRNA expression. RNA (New York, NY). 2006;12(7):1161–7.
124. Shi Z, Johnson JJ, Stack MS. Fluorescence in situ hybridization for MicroRNA detection in archived oral cancer tissues. J Oncol. 2012;2012:903581.
125. Yauk CL, Rowan-Carroll A, Stead JD, Williams A. Cross-platform analysis of global microRNA expression technologies. BMC Genomics. 2010;11:330.
126. Aldridge S, Hadfield J. Introduction to miRNA profiling technologies and cross-platform comparison. Methods Mol Biol (Clifton, NJ). 2012;822:19–31.
127. Duttagupta R, DiRienzo S, Jiang R, Bowers J, Gollub J, Kao J, Kearney K, Rudolph D, Dawany NB, Showe MK, et al. Genome-wide maps of circulating miRNA biomarkers for ulcerative colitis. PLoS One. 2012;7(2):e31241.
128. Chen J, April CS, Fan JB. miRNA expression profiling using Illumina Universal BeadChips. Methods Mol Biol (Clifton, NJ). 2012;822:103–16.
129. D'Andrade PN, Fulmer-Smentek S. Agilent microRNA microarray profiling system. Methods Mol Biol (Clifton, NJ). 2012;822:85–102.
130. Benes V, Castoldi M. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. Methods (San Diego, Calif). 2010;50(4):244–9.
131. Ono S, Lam S, Nagahara M, Hoon DSB. Circulating microRNA biomarkers as liquid biopsy for cancer patients: pros and cons of current assays. J Clin Med. 2015;4(10):1890–907.
132. M'Boutchou MN, van Kempen LC. Analysis of the tumor microenvironment transcriptome via NanoString mRNA and miRNA expression profiling. Methods Mol Biol (Clifton, NJ). 2016;1458:291–310.
133. Tam S, de Borja R, Tsao M-S, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. Lab Investig. 2014;94(3):350–8.
134. Wang H, Ach RA, Curry B. Direct and sensitive miRNA profiling from low-input total RNA. RNA (New York, NY). 2007;13(1):151–9.
135. Liu J, Jennings SF, Tong W, Hong H. Next generation sequencing for profiling expression of miRNAs: technical progress and applications in drug development. J Biomed Sci Eng. 2011;4(10):666–76.
136. Yang Q, Lu J, Wang S, Li H, Ge Q, Lu Z. Application of next-generation sequencing technology to profile the circulating microRNAs in the serum of preeclampsia versus normal pregnant women. Clin Chim Acta. 2011;412(23–24):2167–73.
137. Schulte JH, Marschall T, Martin M, Rosenstiel P, Mestdagh P, Schlierf S, Thor T, Vandesompele J, Eggert A, Schreiber S, et al. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. Nucleic Acids Res. 2010;38(17):5919–28.
138. Urgese G, Paciello G, Acquaviva A, Ficarra E. isomiR-SEA: an RNA-Seq analysis tool for miRNAs/isomiRs expression level profiling and miRNA-mRNA interaction sites evaluation. BMC Bioinform. 2016;17:148.
139. Amsel D, Vilcinskas A, Billion A. Evaluation of high-throughput isomiR identification tools: illuminating the early isomiRome of Tribolium castaneum. BMC Bioinform. 2017;18(1):359.
140. Zhang Y, Zang Q, Xu B, Zheng W, Ban R, Zhang H, Yang Y, Hao Q, Iqbal F, Li A, et al. IsomiR Bank: a research resource for tracking IsomiRs. Bioinformatics (Oxford, England). 2016;32(13):2069–71.
141. Lu C, Meyers BC, Green PJ. Construction of small RNA cDNA libraries for deep sequencing. Methods (San Diego, Calif). 2007;43(2):110–7.
142. Buermans HP, Ariyurek Y, van Ommen G, den Dunnen JT, t Hoen PA. New methods for next generation sequencing based microRNA expression profiling. BMC Genomics. 2010;11:716.

143. Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. BMC Bioinform. 2009;10:328.
144. Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, Moulton V. A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics (Oxford, England). 2008;24(19):2252–3.
145. Bargaje R, Hariharan M, Scaria V, Pillai B. Consensus miRNA expression profiles derived from interplatform normalization of microarray data. RNA (New York, NY). 2010;16(1):16–25.
146. Meyer SU, Pfaffl MW, Ulbrich SE. Normalization strategies for microRNA profiling experiments: a 'normal' way to a hidden layer of complexity? Biotechnol Lett. 2010;32(12):1777–88.
147. Deo A, Carlsson J, Lindlof A. How to choose a normalization strategy for miRNA quantitative real-time (qPCR) arrays. J Bioinforma Comput Biol. 2011;9(6):795–812.
148. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103(7):2257–61.
149. microarrays Ssao. http://www-stat.stanford.edu/~tibs/SAM/index.html.
150. microarrays Ppaf. http://www-stat.stanford.edu/~tibs/PAM/index.html.
151. Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, et al. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc Natl Acad Sci U S A. 2004;101(26):9740–4.
152. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;65(16):7065–70.
153. Huang X, Liang M, Dittmar R, Wang L. Extracellular microRNAs in urologic malignancies: chances and challenges. Int J Mol Sci. 2013;14(7):14785–99.
154. Zhang L, Miller D, Yang Q, Wu B. MicroRNA regulatory networks as biomarkers in obesity: the emerging role. Methods Mol Biol (Clifton, NJ). 2017;1617:241–60.
155. Mo YY. MicroRNA regulatory networks and human disease. Cell Mol Life Sci. 2012;69(21):3529–31.
156. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353–8.
157. Irimie AI, Zimta AA, Ciocan C, Mehterov N, Dudea D, Braicu C, Berindan-Neagoe I. The unforeseen non-coding RNAs in head and neck cancer. Genes. 2018;9(3):134.
158. Ojamaa K. Signaling mechanisms in thyroid hormone-induced cardiac hypertrophy. Vasc Pharmacol. 2010;52(3–4):113–9.
159. Callis TE, Pandya K, Seok HY, Tang RH, Tatsuguchi M, Huang ZP, Chen JF, Deng Z, Gunn B, Shumate J, et al. MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. J Clin Invest. 2009;119(9):2772–86.
160. Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science (New York, NY). 2005;309(5740):1577–81.
161. Yue J. miRNA and vascular cell movement. Adv Drug Deliv Rev. 2011;63(8):616–22.
162. Li W, Szoka FC Jr. Lipid-based nanoparticles for nucleic acid delivery. Pharm Res. 2007;24(3):438–49.
163. Han D, Li J, Wang H, Su X, Hou J, Gu Y, Qian C, Lin Y, Liu X, Huang M, et al. Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. Hepatology (Baltimore, Md). 2017;66:1151.
164. Li JF, Song YZ. Circular RNA hsa_circ_0001564 facilitates tumorigenesis of osteosarcoma via sponging miR-29c-3p. Tumour Biol. 2017;39(8):1010428317709989.
165. Walther W, Stein U. Viral vectors for gene transfer: a review of their use in the treatment of human diseases. Drugs. 2000;60(2):249–71.
166. Al-Dosari MS, Gao X. Nonviral gene delivery: principle, limitations, and recent progress. AAPS J. 2009;11(4):671–81.
167. Geisler A, Fechner H. MicroRNA-regulated viral vectors for gene therapy. World J Exp Med. 2016;6(2):37–54.
168. Rai K, Takigawa N, Ito S, Kashihara H, Ichihara E, Yasuda T, Shimizu K, Tanimoto M, Kiura K. Liposomal delivery of MicroRNA-7-expressing plasmid overcomes epidermal growth factor receptor tyrosine kinase inhibitor-resistance in lung cancer cells. Mol Cancer Ther. 2011;10(9):1720–7.
169. Gondi CS, Rao JS. Concepts in in vivo siRNA delivery for cancer therapy. J Cell Physiol. 2009;220(2):285–91.
170. Lujambio A, Lowe SW. The microcosmos of cancer. Nature. 2012;482(7385):347–55.
171. Ferracin M, Zagatti B, Rizzotto L, Cavazzini F, Veronese A, Ciccone M, Saccenti E, Lupini L, Grilli A, De Angeli C, et al. MicroRNAs involvement in fludarabine refractory chronic lymphocytic leukemia. Mol Cancer. 2010;9:123.
172. Ru P, Steele R, Newhall P, Phillips NJ, Toth K, Ray RB. miRNA-29b suppresses prostate cancer metastasis by regulating epithelial-mesenchymal transition signaling. Mol Cancer Ther. 2012;11(5):1166–73.
173. Kandalam MM, Beta M, Maheswari UK, Swaminathan S, Krishnakumar S. Oncogenic microRNA 17-92 cluster is regulated by epithelial cell adhesion molecule and could be a potential therapeutic target in retinoblastoma. Mol Vis. 2012;18:2279–87.
174. Fassina A, Marino F, Siri M, Zambello R, Ventura L, Fassan M, Simonato F, Cappellesso R. The miR-17-92 microRNA cluster: a novel diagnostic tool in large B-cell malignancies. Lab Invest. 2012;92(11):1574–82.
175. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834–8.
176. Slabakova E, Culig Z, Remsik J, Soucek K. Alternative mechanisms of miR-34a regulation in cancer. Cell Death Dis. 2017;8(10):e3100.
177. Zeng Y, Xu Y, Shu R, Sun L, Tian Y, Shi C, Zheng Z, Wang K, Luo H. Altered expression profiles of circular RNA in colorectal cancer tissues from patients with lung metastasis. Int J Mol Med. 2017;40(6):1818–28.
178. Cao Q, Mani RS, Ateeq B, Dhanasekaran SM, Asangani IA, Prensner JR, Kim JH, Brenner JC, Jing X, Cao X, et al. Coordinated regulation of polycomb group complexes through microRNAs in cancer. Cancer Cell. 2011;20(2):187–99.

Circulating Tumor Cells: Enrichment and Genomic Applications

6

Dorraya El-Ashry, Marija Balic, and Richard J. Cote

Introduction

Worldwide, cancer remains among the most common causes of morbidity and mortality constantly presenting challenges in diagnosis and clinical management. The effective implementation of cancer screening methods and improvements in treatment strategies have led to a decrease in cancer mortality in the last decades. However, despite the advances in multidisciplinary treatment strategies and the development of more efficacious systemic therapy, patients with metastatic disease remain currently incurable.

Cancer spreads locally and through blood and lymphatic vessels, leading to distant metastases. Although the processes underlying metastasis remain largely unknown, single cells or cell clusters detaching from the primary tumor and invading surrounding vessels are regarded to be the carriers of metastatic spread [1–6]. Surrogate markers of metastatic spread have been studied for decades [7]. Regional lymphatic spread has long been evaluated in lymph nodes, with a more recent focus on the assessment of sentinel nodes. Whereas the detection of metastatic hematogenous dissemination was initially focused on bone marrow analysis, it has increasingly shifted to peripheral blood, given the invasive nature of the bone marrow aspiration procedure. Occult metastatic spread is not detectable by routine diagnostic and staging methods and is known in literature as the presence of occult tumor cells, disseminated tumor cells, micrometastases, or, in peripheral blood, circulating tumor cells (CTCs).

D. El-Ashry
Department of Laboratory Medicine and Pathology, Masonic Comprehensive Cancer Center, University of Minnesota, Minneapolis, MN, USA

M. Balic
Division of Oncology, Department of Internal Medicine, Medical University of Graz, Graz, Austria

R. J. Cote (✉)
Department of Pathology, University of Miami Miller School of Medicine, Miami, FL, USA
e-mail: rcote@med.miami.edu

Enrichment Techniques for CTC

The greatest challenge in the detection of CTCs is their rarity in peripheral blood. Very few CTCs will be present, even in patients with advanced stages of metastatic disease. These cells have to be detected among white blood cells ($5–10 \times 10^6$/ml), red blood cells ($5–9 \times 10^9$/ml), and platelets ($2.5–4 \times 10^8$/ml). The frequency of CTCs is often less than 1 CTC per ml of peripheral blood.

A variety of techniques are currently in use for enrichment and detection of CTCs. All these techniques are based on biological and/or physical properties that help distinguish CTC from all the normal blood cells, with the exception of the recently developed RareCyte technology [8]. A brief overview is provided here and the reader is referred to excellent, detailed reviews on the topic [9–13].

Density-Based Enrichment

Among the oldest approaches for cell enrichment are those that take advantage of differences in cell density (a physical property of the cell). An example of such a method is Ficoll-Hypaque separation [14], which separates red blood cells from nucleated cells in the peripheral blood or bone marrow aspirate, including tumor cells. Despite low recovery yield and poor enrichment of the tumor cells by density gradients, it remained the standard approach for many years [15, 16].

Affinity-Based Enrichment

By far, affinity-based enrichment is the most commonly employed strategy. Affinity-based methods take advantage of antigens that are differentially expressed by CTC but not by blood cells (e.g., EpCAM, positive selection) [17, 18] or vice versa (e.g., CD45, negative selection) [19, 20]. A significant drawback of positive enrichment strategies is that they are only effective for CTC that show high expression of the target capture antigen; CTC with low or no expression are not

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_6

enriched. For example, EpCAM is heterogeneously expressed, even by tumor types that are considered to have high expression levels (such as breast cancer). Furthermore, some epithelial tumors may completely lack EpCAM expression (e.g., renal cell carcinoma). In addition, non-epithelial cancers such as melanoma and sarcomas do not express EpCAM.

The most common method of affinity-based selection is immunomagnetic separation using magnetic beads equipped with antibodies that bind to either CTC or to blood cell antigens [21]. Other forms of immunomagnetic enrichment use columns or cartridges that allow for automation of the process [18, 22]. In recent years, microchip-based affinity methods have been described [17, 23, 24]. One of the first reported microchips for CTC enrichment was an affinity-based microfluidic chip with microposts coated with EpCAM antibodies [17, 23]. CTC were successfully isolated in each clinical blood sample tested in all study subjects that included metastatic lung, prostate, pancreatic, breast, and colon cancer patients. Using this method, monitoring of CTC was performed in metastatic non-small cell lung cancer patients demonstrating correlation of CTC count with tumor response [24]. The affinity-based microfluidic chip was also capable of capturing tumor cells from which DNA could be extracted for *EGFR* mutation analysis [23]. A limitation of such chips is their requirement for a very slow flow of blood for efficient capture of CTC, often taking more than 10 h to process 7.5 ml of blood [17].

Size-Based Enrichment (a more even coverage of methods)

An alternative property of tumor cells that has long been considered to be potentially useful for enrichment is cell size [25]. Tumor cells, particularly those derived from epithelial tissues, are larger than most blood cells [26]. An advantage of this approach is that a broader range of tumors is potentially amenable to size-based separation, without dependence on their inherent heterogenous antigen expression. Whereas the use of size to enrich CTC in blood has been considered for almost 50 years [25], it is only recently that size-based separation techniques have become commercially available [27, 28].

Several platforms that use size as the capture method have been described [26]. We have developed a size-based microfilter for enrichment and detection of CTC [29], which is highly efficient and faster than affinity-based separation techniques. The ability to fabricate high-density pore filters allowed for enhancement of both the enrichment factor and the recovery rate of CTC. In initial studies, the achieved recovery rate of tumor cells, spiked into peripheral blood, was >90%. The filtration of a 7.5 ml blood sample could be performed within a few minutes. Using the microfilter, we were able to detect CTC in peripheral blood samples from 51 of 57 metastatic cancer patients that included prostate, colon, breast, and bladder cancer patients, compared to 26 of 57 patients for whom the CellSearch™ method was employed. The mean number of recovered CTC was 5.5 times higher by the microfilter device compared to CellSearch™ [29]. After capture, our platform allows for integrated downstream RNA, DNA, and multi-marker protein characterization [30], which is of particular interest for genomic analysis of CTC. Because cells about 8 microns in size or larger are retained on the filter, most recently, we have also used our platform to capture and enumerate circulating non-cancer, stromal cells from both human patient and mouse blood, an area of intense research interest (see below) [31]. We have further detected both CTCs and circulating non-cancer stromal cells in blood from breast cancer patients [31]. Another advantage of the parylene filter is that parylene is optically transparent making it possible for the capture platform to simultaneously be used as the analysis platform.

We have developed an automated version of our platform (Fig. 6.1) which is completely fluidically integrated to operate in a user-friendly "sample-to-slide" format. This robust platform obviates many manual steps prior to staining and imaging which has resulted in eliminating user-to-user variability and provided an extremely fast processing time of 6 min for a 10 mL blood sample.

Alternative size-based separation platforms have also been described [32]. The main difference between the parylene filter and other sized-based separation platforms is the density and regularity of the pores. Most filters are track-etched, produced by ionizing radiation. This results in an irregular pore distribution, with a low density and significant overlap of pores. Some holes are large enough to allow CTC to pass through. The parylene filter is designed with advanced lithography techniques, resulting in a highly regular and dense pore pattern. The size and shape of the pores can be precisely controlled. The two different designs are depicted in Fig. 6.2.

Other Methods

Among existing alternative methods are a microfluidic device that utilizes deterministic hydrodynamic flow and size-based separation [33–35]. Inside a microfluidic chamber, the device contains a micropost array, and the diameter of the circular micropost, the distance between the microposts in individual rows, and the row-to-row shift determine its performance. With such a device, separation of plasma from blood cells, different types of blood cells from each other, and DNA fragments of different size can be performed.

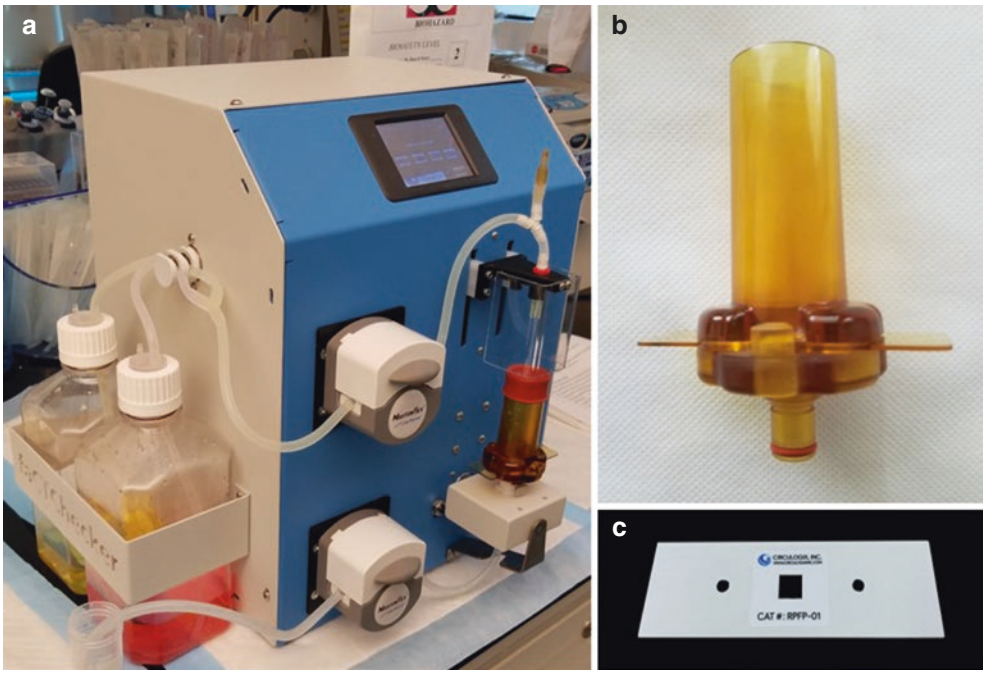


Fig. 6.1 (**a**) Circulogix FaCTChecker automated fluid handler allows end user to capture fixed and live CTC and tumor microenvironment cells in a sample-to-slide format. (**b**) Custom cartridge houses the microfilter slide. After processing the blood sample, the cartridge is disassembled to recover (**c**) the microfilter slide for downstream analysis of captured CTC and tumor microenvironment cells

The time required for relatively large volumes of blood samples is considerably longer.

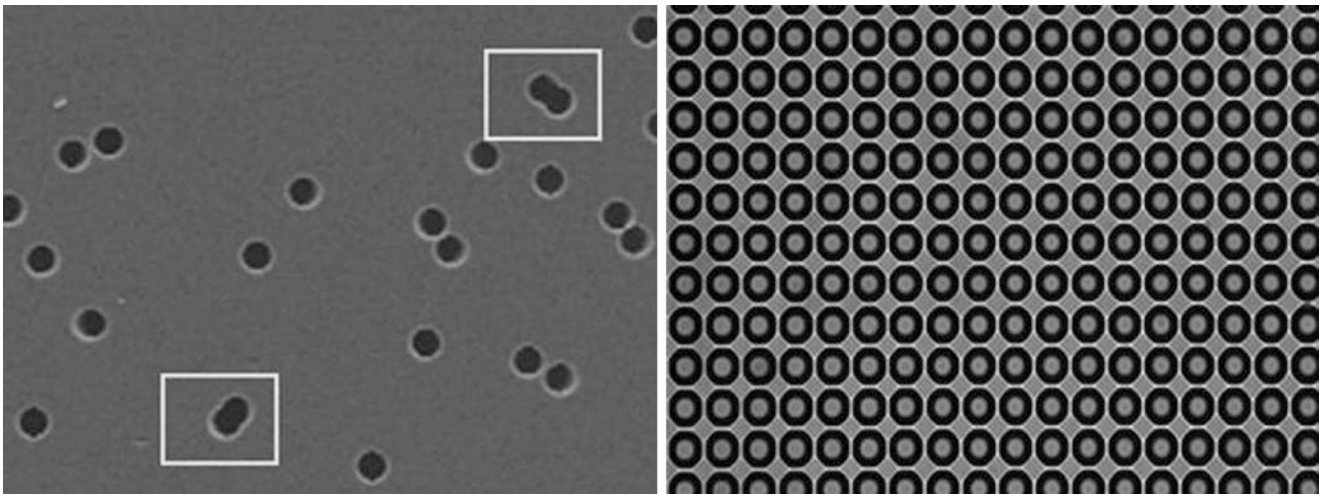


Fig. 6.2 Comparison of the design of track-etch filter and a parylene filter. On the left, holes large enough to allow CTC to pass through are framed. The parylene filter is designed with advanced lithography techniques, resulting in a highly regular and dense pore pattern (right image) [28, 29]

Some devices have been developed using dielectrophoretic forces applied through the microelectrode arrays onto the field. Based on their electric properties, cells are positioned at a specific distance from electrodes [34, 36]. This technique has demonstrated 100% separation efficiency when defined numbers of cultured human breast cancer MDA-231 cells were spiked into peripheral blood [37, 38]. A severe limitation of this technique is that only a very small sample volume (30 μl) can be processed. Given the low number of CTC (often less than 1 CTC per ml), this technique would not be expected to be useful for the detection or isolation of CTC in clinical samples.

Another recent methodology of CTC detection that does not rely upon targeted cell enrichment, but simply takes a sample of blood, distributes it among microscopic slides, and identifies cancer cells by immunohistochemical methods using epithelial cell markers was described by Campton et al. [8]. These methods then use microscopy and image analysis software to identify the tumor cells. Identified cells can be extracted and further analyzed for mutations or gene expression.

Detection Methods

CTC enrichment methods allow for the ratio of target CTC to background cells to be significantly enhanced. However, none of the currently available enrichment methods results in a pure population of tumor cells. Following enrichment, all separation techniques require a method to distinguish CTC from the nonspecifically captured cells. Several approaches can be performed to identify tumor cells: cytomorphological characterization of CTC, immunohistochemical/immunofluorescent (IHC/IF) detection of tumor-specific antigens, or various real-time polymerase chain reaction (RT-PCR) approaches. Cytomorphological characterization relies on classification of tumor cells based on their distinct morphological features [28]. Immunocytochemical detection of CTC relies on antibody-based detection of cells using antibodies specific for epithelial cells. Most commonly used antibodies are cytokeratins, including both low- and high-molecular-weight cytokeratins. This method is now often combined with markers such as CD45 that identify the background blood (non-CTC) cells. A representative cytokeratin-positive cell in a background of blood cells is depicted in Fig. 6.3.

Multiplex IHC/IF approaches enable simultaneous visualization of multiple markers on a single cell [39], such as in the representative image for detection of cytokeratins and the putative breast cancer stem cell (CSC) marker aldehyde dehydrogenase (ALDH) (Fig. 6.4). Detection of CTC by IHC/IF has one major potential drawback, namely, the potential to miss cells that lack the expression of the targeted antigen. The potential for such omission is further suggested by recent literature demonstrating epithelial-mesenchymal transition (EMT) and expression of mesenchymal markers by epithelial CTC [40]. The simultaneous use of multiple cytokeratins of high and low molecular weight, including stem cell-associated cytokeratins, has minimized this concern [41, 42].

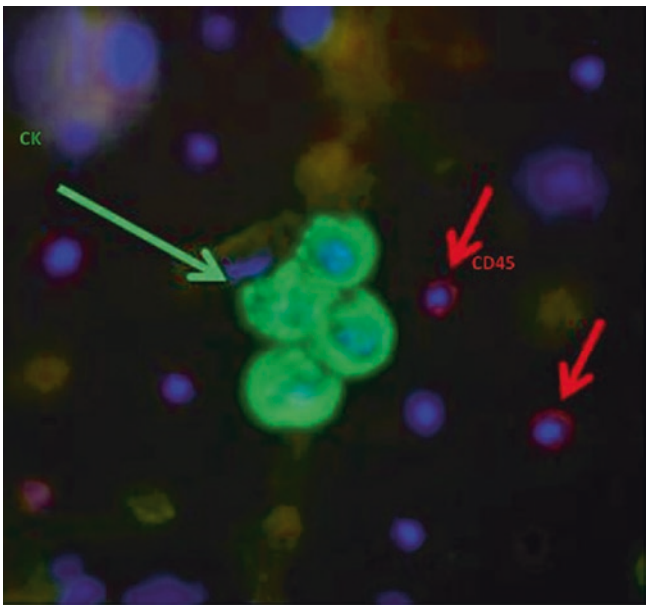


Fig. 6.3 Detection of a circulating tumor cell in a background of hematopoietic cells. CTCs are identified as cells that are large in size (i.e., 12–25 μm), nucleated (as demonstrated on 4′6-diamidino-2-phenylindole (DAPI) positivity), cytokeratin (CK) positive (green), and CD45 negative and display other morphologic features consistent with malignant cells (e.g., a high nuclear to cytoplasmic ratio). In contrast, the lymphocytes display a CK-negative/CD45-positive (red) profile

Cells Associated with CTCs

Recent studies show CTCs are not the only tumor elements that reach the peripheral blood. An accumulating body of evidence has demonstrated the pivotal role of stromal cells in promoting cancer progression [43–46], metastasis, and poor clinical outcome [47]. Cancer-associated fibroblasts (CAFs), an essential component of the tumor microenvironment (TME) [48] in many cancers, comprise the majority of BC stromal cells. Stromal cells such as CAFs are not only present in the TME of the primary tumor but exist in pre-metastatic and metastatic niches, and mouse metastasis models demonstrate that cancer cells take their CAFs with them to metastatic sites, and the survival and establishment of these CAFs promote metastatic seeding and growth of the cancer cells [3]. Since successful metastatic seeding by CTCs is dependent on a productive interactive relationship with their environment as well as avoidance of immune surveillance, elucidation of interactions between CTCs and their microenvironment is critical. Such studies have been hindered primarily due to lack of a suitable platform to interrogate multiple cell types simultaneously; the size-based microfilter platform enables such multicellular analyses as well as the capture and release of live circulating cells from patient blood. Using this platform, we discovered that non-cancer, nonimmune cells were being isolated in association with CTCs and identified these as circulating CAFs (cCAFs) [31]. In a pilot study in which cCAFs and CTCs were enumerated in the blood of breast cancer patients, we found cCAFs present in patients with Stage IV (metastatic) breast cancer but not in blood from patients with likely cured Stage I disease, while CTCs were detected in both (Fig. 6.5 and [31]). Jones et al. also found circulating fibroblast-like cells, identified as CK-/CD45-/vimentin+, in the blood of metastatic prostate cancer patients [49]. More recently, we have used this platform to capture and enumerate CTCs and stromal cells from mouse blood in mouse xenograft models.

In addition to CAFs, tumor-associated macrophages (TAMs) have also been identified in circulation [50]. Circulating immune cells such as cytotoxic T cells (T_{cyt}) play important roles in tumor suppression, and other immune cells such as myeloid-derived suppressor cells (MDSCs), regulatory T cells (T_{regs}), and TAMs play important roles in tumor enhancement (reviewed in [51]). MDSCs are recruited to tumors and metastases and actively suppress T_{cyt}, and this

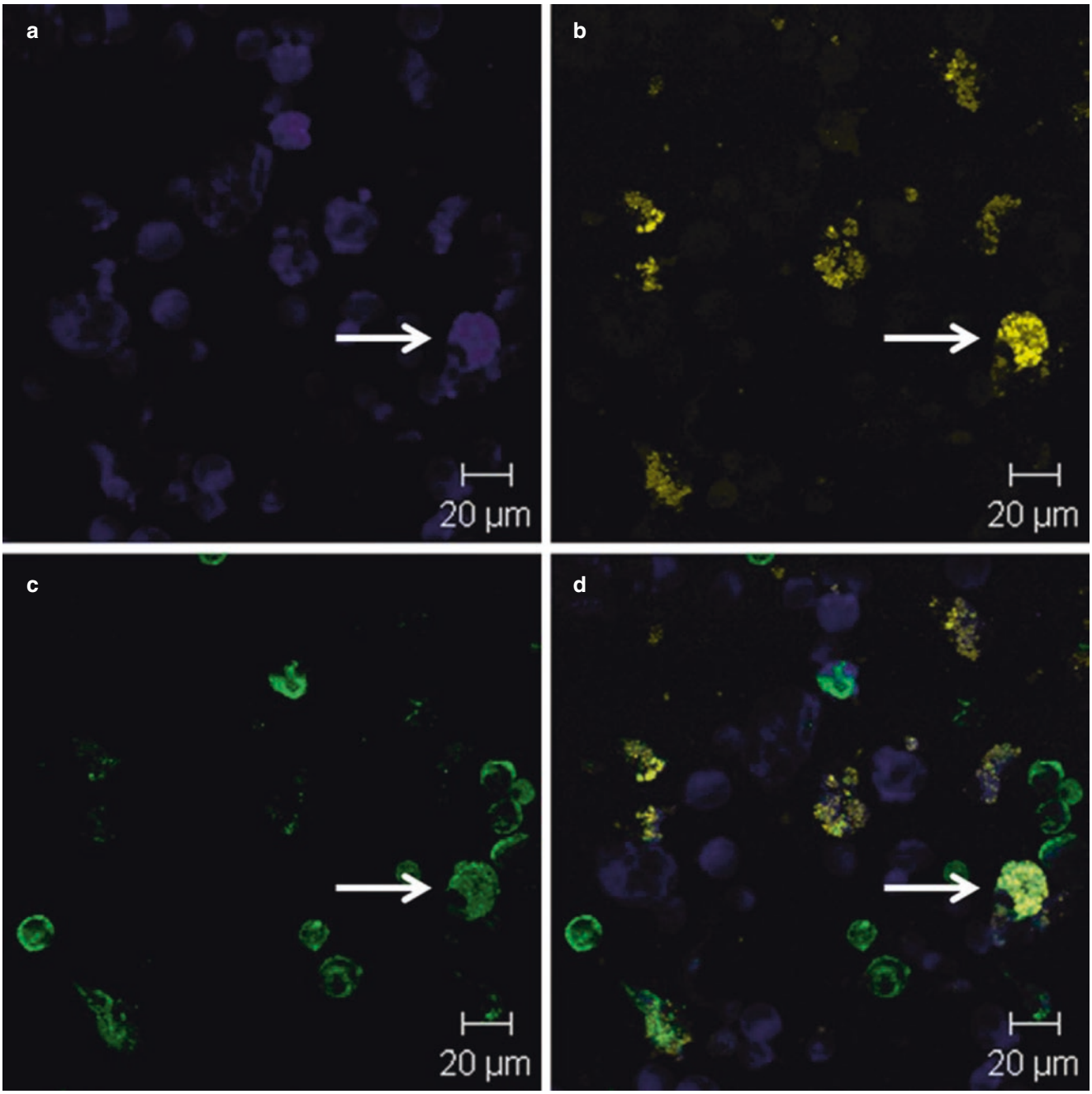


Fig. 6.4 Double immunofluorescence for a putative breast cancer stem cell marker ALDH and CK in a cytospin preparation from a mixture of cancer cell lines (MCF7, SUM159, SUM1315, and HEPG2). The sample was assessed for (**a**) DAPI (blue), (**b**) ALDH using a secondary antibody labeled with DyLight 550 (yellow), (**c**) CK with a secondary antibody labeled with Alexa Fluor 488 (green), and (**d**) all markers merged. The cell indicated with an arrow is a representative cell positive for ALDH and CK (DAPI, 4′6-diamidino-2-phenylindole; ALDH, aldehyde dehydrogenase; CK, cytokeratin)

suppression of immune response results in enhanced tumor growth and metastasis (reviewed in [52]). Depletion of MDSCs in mouse breast cancer models results in reduced tumor growth and inhibition of metastasis [53]. Using the MDSC marker S100A8 [53], we have been able to demonstrate the presence of S100A8+ cells in blood from mice with metastatic syngeneic breast tumors with the size-based microfilter platform (Fig. 6.6).

Cancer Stem Cells

Experimental evidence in support of a "cancer stem cell model" in various malignancies is mounting [54]. The CSC model assumes the presence of a small proportion of cancer cells on top of a hierarchy of tumor cells; these CSCs exhibit the capability to sustain tumor formation and growth, self-renewal, and differentiation. Several markers of CSC have

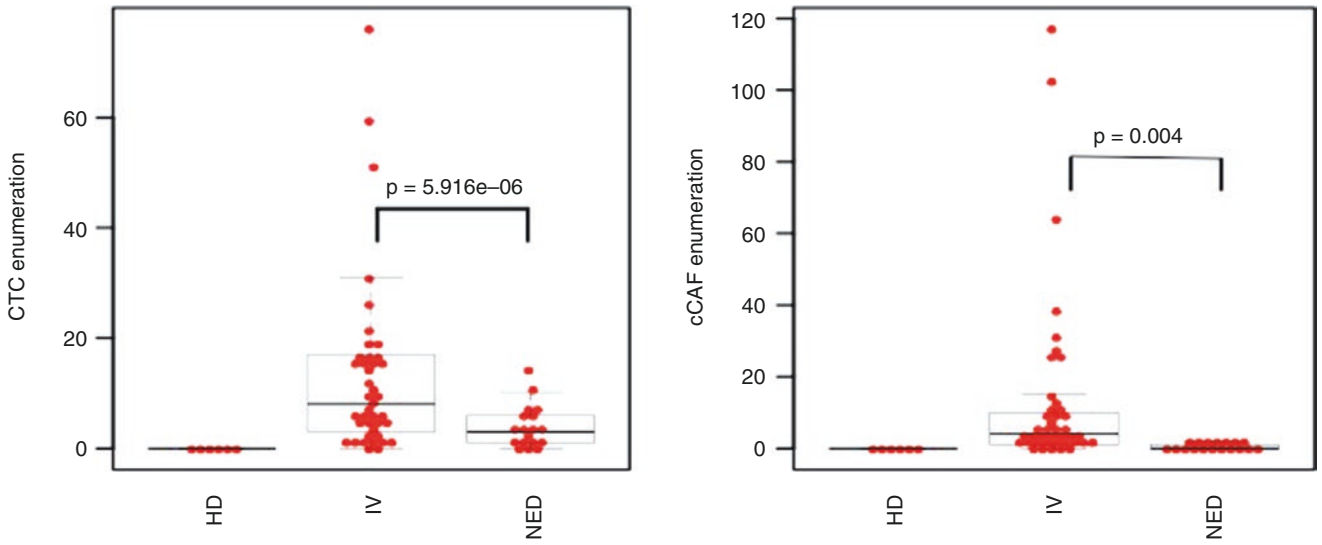


Fig. 6.5 Swarm plots of CTC (C) and cCAF (D) counts for breast cancer patients with Stage IV disease, patients cured of localized Stage I breast cancer and no evidence of disease (NED), and healthy donors (HD). (From Ao et al. [31] with permission of American Association for Cancer Research)

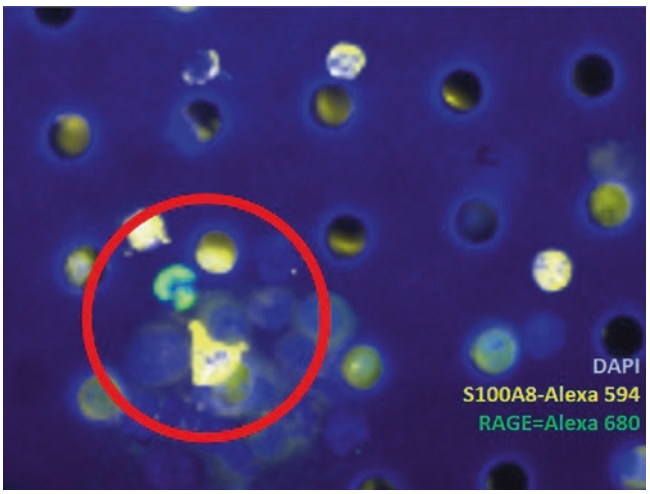


Fig. 6.6 S100A8+ cells (yellow) were detected in circulating clusters with RAGE+ (green) 4 T1 mammary cancer cells in the blood of mice challenged with syngeneic 4 T1 tumors, Figure generated in collaboration with Marc Lippman

been identified, including a $CD44^{+}CD24^{-/low}$ phenotype and ALDH expression in breast and other cancers [55, 56] and CD133 expression in non-small cell lung cancer and brain tumors [57, 58].

In breast cancer, and presumably other cancers, the identification of distinct properties and molecular biomarkers of CSC may help in the development of more effective treatment and novel therapeutic targets [59]. In general, the presence of occult micrometastases is the rationale behind the use of systemic adjuvant chemotherapy following a definitive local treatment of a primary tumor [60]. The occult dissemination of CSC in the bone marrow may be responsible for the failure of adjuvant chemotherapies in a proportion of early-stage breast cancer patients [61]. Several in vitro studies have demonstrated that putative breast CSCs are resistant to conventional treatment strategies, including radiation and chemotherapy [62–64]. Consequently, the identification of breast CSC among CTC may be a promising strategy to assess their malignant potential and identify novel therapeutic targets. A major hurdle for such an approach is the hereto limited available knowledge regarding CTC phenotypes and the fact that CSCs represent only a proportion of enriched CTC.

CSC in Dissemination

Pooled analysis of data from nine prior breast cancer studies which included a large number of early-stage patients revealed that the presence of micrometastatic tumor cells in the bone marrow (commonly referred to as disseminated tumor cells, or DTC) is associated with a poor prognosis. Surprisingly, a significant proportion of patients with DTC had a favorable survival outcome of 10 years or more following the diagnosis [7]. One potential reason for such an outcome could be the ability of DTC to remain dormant in distant organs. Based on the prior observations, we hypothesized that CSCs not only exist within the primary tumor but may represent the most potent and virulent cells metastasizing from primary breast cancer to distant locations. In order to test our hypothesis, we performed a study analyzing DTC from breast cancer patients enrolled in the ACOS-OG Z-00010 trial for the putative breast CSC phenotype $CD44^{+}CD24^{-/low}$. The large majority of DTC in examined patients had the putative CSC phenotype [61].

This study provided the first demonstration that DTCs/CTCs are primarily composed of CSC, in contrast to primary and metastatic tumors in which fewer than 10% of cells have a CSC phenotype [61, 65]. This finding has significant biologic implications, as it suggests that there is an enrichment for breast CSCs in the process of metastasis [66]. The fact that the ACOS-OG Z-00010 trial patients were early-stage I and II breast cancer patients in whom only 3% of BM samples were positive for DTC made the finding even more significant [67].

Several studies have since confirmed these findings. In a prospective analysis of bone marrow aspirates from high-risk breast cancer patients, using cell sorting by flow cytometry, Reuben et al. [68] were able to show a high percentage of CSC in DTC. Using a similar approach, Theodoropoulos et al. demonstrated the presence of CTC in 67% of patients with metastatic disease, with 35% of CTC displaying the $CD44^{+}/CD24^{low}$ CSC phenotype [69]. In another flow cytometry study, evaluating peripheral blood from breast cancer patients at variable stages, Wang et al. showed an increasing percentage of putative CSC in correlation with higher tumor stage [70]. The above findings further emphasize the need for reliable CTC enrichment methods allowing for detailed molecular characterization.

Epithelial-Mesenchymal Transition

Progression to an EMT phenotype is widely accepted as a contributing factor to tumor metastasis, and the ability of tumor cells to undergo EMT is crucial for local invasion and gaining access to the blood stream through intravasation [71]. EMT is associated with genetic changes that lead to increased tumor cell motility and an invasive phenotype. These changes are typically characterized by loss of E-cadherin expression and subsequent translocalization of β-catenin from the cell membrane into the nucleus, increased expression of vimentin, production of matrix metalloproteinase enzymes, and upregulation of various EMT-inducing transcription factors such as Twist, Snail, and Slug [72]. Thus, EMT provides a potential mechanistic basis for how CTCs intravasate in primary tumors to reach the circulation and subsequently extravasate from the circulation to seed tumor implants at distant secondary sites. Several studies have evaluated the expression of EMT-associated markers in CTC. In a study involving metastatic breast cancer patients, Aktas et al. revealed at least one of three EMT markers (Akt2, PI3K, and Twist1), assessed by RT-PCR, to be expressed by the CTC population in 62% of patients harboring CTC. Patients with CTC who were positive for EMT were more likely to fail to respond to palliative chemotherapy, antibody, or hormonal therapy [73]. Evaluating CTC expression of EMT markers Twist and vimentin by immunofluorescence, Kallergi et al. found vimentin/Twist expressing CTC in 77% of early-stage breast cancer patients compared to 100% of patients with metastatic disease [74]. In a recent study involving 11 breast cancer patients who were serially monitored for CTC phenotype, mesenchymal phenotype CTCs were more likely to be associated with disease progression. In one index patient, the authors were able to demonstrate a reversible shift between the epithelial and mesenchymal phenotype corresponding to response to treatment and disease progression, respectively [75].

Recently, an overlap between the EMT phenotype and the CSC phenotype has been described. Overexpression of EMT transcription factors increases CSCs [76–78], such that EMT may thus drive both tumor dissemination and increase CSC self-renewal to facilitate tumor metastasis. Alternatively or additionally, the plasticity of tumor cells and their capability to transform and acquire mesenchymal characteristics may be derived from CSC [79]. To further complicate the link between EMT and the CSC phenotypes, we recently found that the CSC compartment can be epithelial and distinct from the mesenchymal compartment in a primary lung cancer cell culture that we developed from a patient with primary resistant disease [80].

CTC Clusters

Additional potential mechanisms that could facilitate tumor cell dissemination include ameboid motility and collective migration of cell clusters [81]. In mouse breast cancer models, CTC clusters exhibit higher metastatic capacity compared with individual or single CTCs [2]. Additionally, it was demonstrated that polyclonal breast cancer metastasis resulted from CTC clusters composed of keratin 14+ CTCs [82]. Recently, CTC cluster enumeration in breast cancer patients has demonstrated that like CTC number, the presence of CTC clusters correlates with reduced progression-free survival and poor outcome [4–6, 82]. Recent data suggests that CTC clusters that are composed of CSCs, thus CSC clusters, are more metastatic than CTC clusters composed of non-CSC CTCs; these studies have also demonstrated the presence of CSC clusters in patient blood [83]. Interestingly, in these studies the CSCs exhibited a hybrid epithelial/mesenchymal phenotype. The occurrence of circulating tumor microemboli (CTM) in metastatic lung cancer patients was demonstrated by Hou et al. [84]. In this study, the authors showed that single CTC expressed apoptosis-related markers at a higher rate than CTM. These findings suggest that collective migration of tumor cells in circulation may offer a survival advantage to the tumor.

In our pilot study [31], we can see that cCAFs can cluster with CTCs as well as with each other (Fig. 6.7).

Thus, it may be that it is not solely CTC clusters that are responsible for metastatic seeding, but rather it is clusters

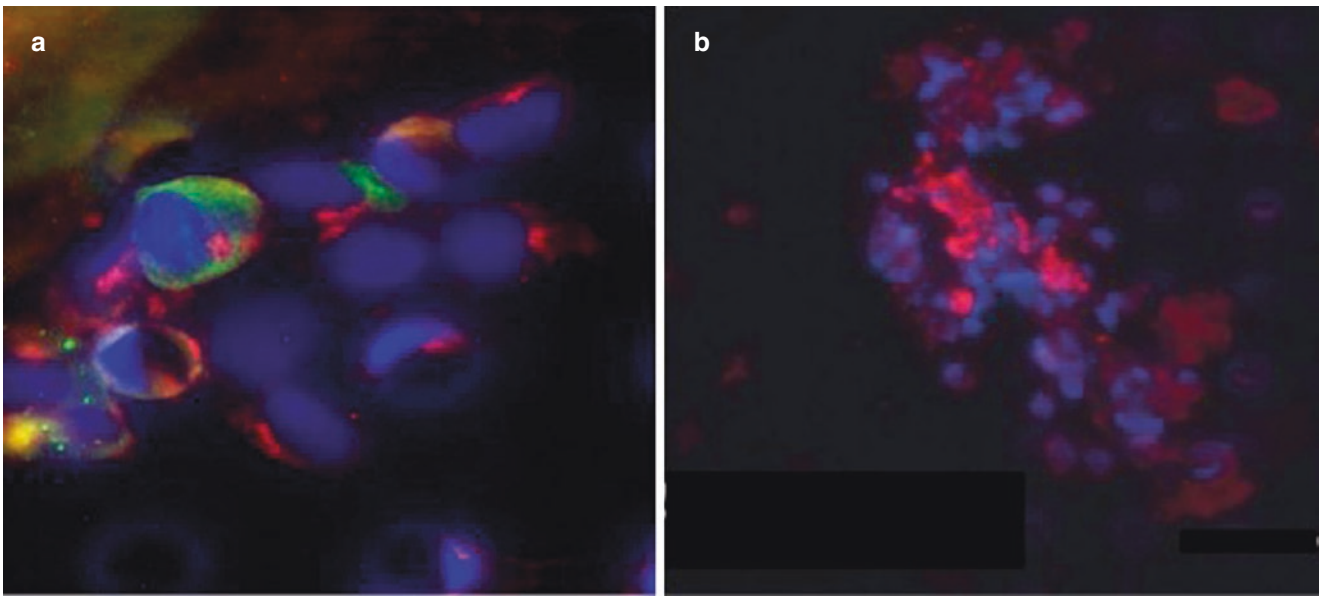


Fig. 6.7 Identification of CTC and cCAF clusters in patient blood on microfilter. DAPI (blue), pan-CK (green), FAP (red). (**a**) Representative picture of CTC/cCAF clusters and (**b**) cCAF/cCAF clusters. (From Ao et al. [31] with permission of American Association for Cancer Research)

comprised of CTCs along with their tumor microenvironment, including cCAFs that are the metastasis-driving factors. Given the recent evidence that at least a portion of CTCs are cells transitioning between the epithelial and mesenchymal state [75] that possess stem cell-like properties and the ability of reversible modulation [85], the functional characterization of these processes in CTCs, and in these CTC/stromal cell clusters, is crucial. Development of new technologies that will enhance sensitivity and efficiency of CTC and CTC cluster detection as well as detection of CTC-associated stromal cells will facilitate functional characterization of CTC invasiveness, aggressiveness, plasticity, and tumorigenic potential. Functional characterization will, in turn, help further clarify the mechanisms of tumor cell dissemination.

Recent advances in CTC enrichment have allowed for better molecular characterization of CTC. The latter is carried out by various strategies that include fluorescent in situ hybridization (FISH), comparative genomic hybridization (CGH), PCR-based techniques, RNA-seq, and immunofluorescence. These studies have shed light on the oncogenic profile and metastatic potential of CTC and have allowed the comparison of the genetic profile of tumor metastases and CTC to that of their primary tumor counterpart.

Molecular Characterization of CTC

It is becoming apparent that intratumoral heterogeneity is one of the many factors that could be responsible for therapeutic failure and drug resistance in cancer. Tumor cell heterogeneity is attributed, in part, to one of several key biologic mechanisms of cancer progression. Chief among these are the presence of a perpetual supply of cancer stem cells (CSCs) [86], the process of EMT, and the heterogeneity of expression of molecules that determine drug response and resistance, including those that are therapy targets among cancer cells [87–89]. For example, variability in estrogen and progesterone receptor expression and Her2-neu status among mammary cancer cells will result in heterogeneity of response to therapies directed against these targets [90, 91]. Given the likely role of CSC in tumor heterogeneity and resistance [92], molecular characterization of CSC in the setting of metastatic dissemination has increasingly gained interest.

Global Genomic Profiling of CTC

Approaches to the genetic characterization of CTC include evaluation of whole-genome copy number alterations in CTC in comparison to their primary tumor origin [59, 93–97]; Magbanua et al. [98] developed a novel approach for the molecular profiling of CTCs utilizing sequential immunomagnetic enrichment steps and flow cytometry sorting to isolate CTC, followed by whole-genome DNA amplification and array CGH analysis. These studies unveiled a wide range of copy number alterations in CTC obtained from peripheral blood samples of patients with advanced breast cancer by comparative analysis between the CTC CGH dataset and a previously published dataset of primary tumor CGH [99] that were more frequent in CTC compared to primary tumors. A sub-analysis performed in patients with known *HER-2* sta-

tus of primary tumors revealed focal amplification of *HER-2* in CTC obtained from two patients with *HER-2*-positive primary tumors. In contrast, in eight patients with *HER-2*-negative primary tumors, the CTC samples showed no *HER-2* copy number gain in six and low-level gains in the remaining two. Paris et al. demonstrated that copy number profiles of CTC detected in castration-resistant prostate cancer patients were similar to those of their paired solid tumor DNA and distinct from corresponding DNA from the remaining depleted mononuclear blood cells after EpCAM enrichment of CTC [100]. Similarly, Magbanua et al. were able to show the utility of immunomagnetic enrichment followed by fluorescence-activated cell sorting for isolating CTC in castration-resistant prostate cancer. The isolated CTCs were successfully used to perform copy number profiling, evaluate progression, and monitor response to therapy [101].

Recently, Heitzer et al. have tested whether tumor-specific copy number alterations can be detected in the peripheral blood of patients with cancer [102]. The authors evaluated the plasma DNA concentration and the fraction of DNA fragments in patients with colorectal and breast cancer and in healthy controls along with CTC detection by CellSearch. The presence of biphasic DNA size distribution was associated with increased CTC counts in cancer patients. Further, plasma DNA was screened for mutations with deep sequencing and an ultrasensitive mutation-detection method. In patients with biphasic DNA size distribution, an elevated concentration of mutated plasma DNA fragments was also detectable. The authors suggested that detection and characterization of plasma DNA in cancer patients may be useful for monitoring the response of cancer patients to ongoing treatment.

Our group has evaluated the feasibility of IHC/IF labeling of CTC captured by the parylene filter described above, for subsequent microdissection using a precise and contact-free laser microdissection system (PALM) and DNA extraction. Figure 6.8 shows MCF7 breast cancer cells that were captured and microdissected using this method. The captured CTC underwent DNA extraction and whole-genome amplification. The DNA quality was verified by multiplex PCR according to a previously published protocol for the evaluation of the DNA quality prior to CGH analysis [103]. Whole-genome array CGH analysis using a 44 K Agilent array successfully revealed the expected array CGH profile of MCF7 cells.

The greatest advances in the treatment of cancer have been made with combinations of targeted therapies. The identification of important pathways and their components and the characterization of the predictive value of specific molecular changes have led to an improvement in individualization of the treatment of cancer patients. With increasing knowledge of molecular targets and biomarkers, it may become necessary to perform genomic profiling of a large number of genomic changes in cancer tissue, metastatic sites, and CTC prior to the implementation of anticancer therapies. Once multi-targeted drugs or combinations of targeted therapies become clinically applicable, genomic profiling may help to optimize treatment. Such profiling may require next-generation sequencing of a larger number and combination of genes or profiling of many translational products. Currently, the technologies are limited. For example, the number of markers that can be analyzed is limited, regardless of their structures (i.e., proteins, mRNA, DNA, etc.). Whereas global profiling may add substantial information to the understanding of metastasis, heterogeneity of tumors, and the biology of disease, the current analyses are largely experimental. Analyses of specific changes, for now, may have a more specific clinical role in determining treatment strategies. The development of technologies capable of determining a large number of clinically relevant biomarkers is exigent.

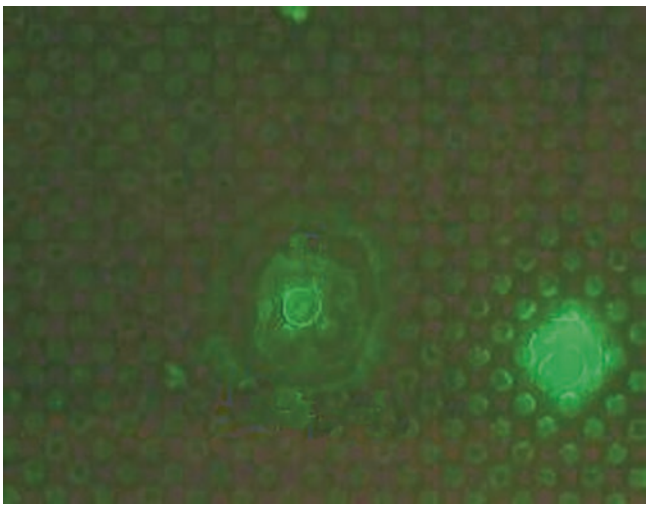


Fig. 6.8 Following CK (cytokeratin) immunostaining (*green*), laser microdissection of the parylene filter was performed under a precise and contact-free laser microdissection system (PALM). The microdissection of the filter represents a necessary step for single cell analyses

Mutation Analyses of CTC

Currently, mutation analysis of genetic alterations that predict response to targeted therapies in metastatic-stage cancer patients is performed on primary rather than metastatic tumor samples. Examples include assessment of *KRAS* and *BRAF* mutations and *EGFR* mutations in colorectal cancer and lung adenocarcinoma, respectively [104, 105]. Whereas determinants of response to a given targeted therapy would ideally be assessed in the metastatic tumor being treated, obtaining such samples through an invasive surgical or imaging procedure is associated with significant morbidity and/or

expense. Finding a source of tumor that can be noninvasively accessed, which represents the most relevant population of tumor cells being treated and could be serially accessed through the course of therapy, would be ideal. CTC may well represent such a source of tumor cells.

Early studies demonstrated the ability to detect specific mutations in captured CTC to predict response to targeted therapies [106] where the potential of CTC mutation analysis of *KRAS* along with further molecular analyses to provide real-time information on tumor biomarker status, including EGFR, Her-2neu, or ER on spiked cancer cells from various cancer cell lines, was shown. Maheswaran et al. demonstrated the feasibility of detecting *EGFR* mutations in CTC in metastatic lung cancer patients undergoing tyrosine kinase inhibitor treatment [23]. Dharmasiri et al. demonstrated the technical feasibility of detecting *KRAS* mutations in colon cancer cell lines spiked into peripheral blood [107]. Subsequently, Yang et al. [108] and Mostert et al. [109] were able to detect *KRAS* mutations in CTC in the peripheral blood of colorectal cancer patients. The discrepancies occasionally found between tissue and CTC *KRAS* mutation status were suggested to be due to the limited number of CTC available [109]. Gasch et al. have recently been able to analyze single CTC, obtained by CellSearch-based enrichment, for several genetic alterations in 49 metastatic colorectal cancer patients [102]. Considerable heterogeneity among patients and within individually analyzed cells from the same patient was found in regard to *EGFR* expression and genetic alterations in *EGFR*, *KRAS*, and *PIK3* [102]. Jiang et al. [110] have established an approach to detect androgen receptor mutations in CTC from castration-resistant prostate cancer patients after enrichment of CTC with the CellSearch method. Such an approach may facilitate the development of more effective treatment strategies in advanced prostate cancer.

Technological advances in CTC enrichment and sequencing have made it possible to perform genomic profiling on CTC. Many of the described and available technologies have the potential to be further developed for genomic profiling of CTC pools and single CTC in clinical samples. This will facilitate studies of CTC as a tool for liquid biopsy but also the evaluation of intra-CTC heterogeneity. For downstream analyses, there are many possibilities, depending on the questions to be addressed. Good single cell DNA quality may allow for global array CGH profiling and next-generation sequencing of larger regions or selected genes. Finally, specific PCR protocols on various platforms may be employed to detect particular genetic alterations and mutations.

Transcriptional CTC Profiling

Transcriptional profiling of CTC presents a significant technical challenge. A study by Smirnov et al. was one of the first to attempt global gene expression profiling of CTC in colorectal, prostate, and breast cancer patients [111]. Global gene expression profiles of CTC-enriched and corresponding CTC-depleted portions were generated, and a list of CTC-specific genes was obtained. Subsequently, using quantitative RT-PCR, the authors were able to differentiate the expression level of a set of CTC-specific genes in patients compared to normal controls. The study illustrated, for the first time, the feasibility of performing global gene expression profiling in CTC.

Barbazan et al. [112] performed whole-transcriptome amplification and gene expression analyses on affinity-enriched CTC from metastatic colorectal cancer patients. A 410-gene CTC signature was identified by hierarchical clustering, which included genes related to cell movement, cell adhesion, cell death, proliferation, cell signaling, and interaction. Confirmation of several genes was performed by quantitative RT-PCR in an independent set of patients. Sieuwerts et al. [113] brought attention to the fact that profiling a low number of CTC may result in discrepant estrogen receptor and HER-2 status profiles compared to primary tumor—a finding that could impact the use of current therapeutic strategies in breast cancer [87, 114].

Gene expression profiling studies, such as those evaluating the expression profiles of EMT-related and CSC signatures in CTC [41, 61, 64–67], have enabled a more detailed evaluation of the biologic events associated with CTC and cancer metastasis. These studies provide preliminary support for the utility of CTC genomic assessment as a tool for exploring the biology of metastasis.

Epigenomic and miRNA Characterization of CTC

Epigenetic events are fundamental to normal processes of development and differentiation and are increasingly found to play a substantial role in carcinogenesis. Aberrant DNA methylation profiles, histone modification, and the alterations in microRNA (miRNA) are examples of epigenetic alterations associated with cancer formation. Therefore, assessment of epigenomic alterations in CTC is crucial to further our understanding of the biology of cancer metastasis. As with all other types of genomic analyses, the rarity of CTC in patient samples presents technical challenges to epigenetic applications.

So far, only few studies have evaluated DNA methylation in CTC, attempting to correlate CTC occurrence with the methylation status of circulating DNA [68–70]. Likewise, few studies have either addressed the association of cancer miRNA alterations and CTC occurrence or the expression of miRNA in CTC. Sieuwerts et al. [113] were able to demonstrate the expression of ten miRNAs in CTC in metastatic breast cancer patients. Such studies are bound to become

more frequent as new prognostic and therapeutic applications related to epigenetic alterations in cancer emerge.

CTC and cStromal Cell (cStC) Live Cell Capture, Culture, and Propagation

Currently the vast majority of CTC studies have been performed on fixed cells and rely on the interrogation of an exceptionally rare population. With the recent ability to capture live, unfixed cells, additional and long-term analyses of CTCs and their associated circulating stromal cells can be performed. Ideally, the culture and propagation of CTCs would allow for study of mechanisms of metastasis, drug resistance, and dormancy. In the last few years, a small number of reports have been published on propagating CTC with the goal of creating CTC cultures from breast, colon, lung, and prostate cancers [2, 115–120]. All studies that attempt to propagate CTC share the following features: (a) requirement of high starting numbers of CTC in the blood (3–3000 cells/10–20 ml blood), (b) low culture efficiency, (c) use of non- or low adherent culturing conditions, and (d) development of initial cultures in small volumes. The reported successful cultures include development of one colon cancer CTC cell line (out of 30 patients) [115], one prostate cancer cell line (out of 17 patients) [117], and six breast cancer cell lines (out of 36 patients) [116, 119]. Recent studies have reported greater success in developing short-term cultures using smaller starting numbers of CTC co-cultured with fibroblasts [118], CTC clusters [2, 120], and successful xenograft propagation of lung cancer CTC [121].

While the capture of viable CTC directly from blood may have profound clinical implications, the strong retention of CTCs on the filter hinders downstream transfer of captured cells onto other platforms such as microfluidic Fluidigm™ single cell analysis system [122] or culture on specialized surfaces such as ultralow attachment plates. We have developed the capability of capture and thermoresponsive release of viable cancer cells from our slot pore filter [123] using PIPAAm to coat our filters. When spiked into blood, cancer cells (i.e., CTCs) could be captured using our slot filter, thermically released, and successfully cultured again. We have also investigated the ability of fibroblasts and CAFs to support the growth of viably released cancer cells from the PIPAAm filters. The mechanical properties (rigidity) and the composition of the extracellular matrix have also been identified as important factors regulating proliferation, function, and fate of a multitude of cell types, including stem cells [124], fibroblasts [125], lymphocytes [126], beta cells [127], muscle cells [128], and cancer cells [129, 130]. In particular, the mechanical signals of a tumor microenvironment can have profound effects of the initiation and propagation of metastasis [131]. In a study on a panel of 17 cancer cell lines, a majority of the cells exhibited substrate rigidity-dependent cell growth [130].

Invading cancer cells (i.e., CTC) from various malignancies are attracted to the bone marrow given the fertile environment it provides to establish secondary metastatic tumors. The process of bone invasion requires a close interaction of the extravasating CTC with local stromal residents of mesenchymal phenotype (hMSC), situated at the perivascular niche [132–134] and at the parenchyma [135, 136]. This cellular complex has been shown to regulate various biological phenomena implicated in determining the fate of the distant metastatic tumor, including cell proliferation, angiogenesis, chemoresistance, etc. [135, 137, 138], and MSCs are the progenitors of CAFs. Therefore, efforts to in vitro recapitulate the target tissue microenvironment to provide "natural" habitats for CTC to thrive are of paramount importance. In that regard, hMSC key role is twofold: first, as progenitors of bone extracellular matrix (ECM)-secreting osteoblasts and second, as perivascular cells (i.e., pericytes) interacting with invading CTC. Both phenotypic variations can be reproduced using in vitro/in vivo assays. hMSC can be harvested from various sources including the bone marrow, selected on the basis of plastic adherence, culture-expanded, characterized (immunotypification and tri-lineage differentiation potential assessment), and used as osteoblastic progenitors to produce bone tissue [139–141]. The differentiation to bone-producing osteoblasts can be performed either in vitro or in vivo [133, 141, 142]. Interestingly, these results were tightly reproduced with breast cancer cells.

Conclusions

The ability to detect and characterize CTC remains a technical challenge. Advancements in CTC enrichment, detection, and characterization methods are rapidly being made. In the past decade, molecular assessment of CTC at the single cell level has provided the foundation for improved understanding of the biology of metastatic cancer spread. We have witnessed an era of great technical advancement that has led to the improved sensitivity of CTC detection and a better definition of recently discovered molecular processes related to CTC occurrence. Advances in next-generation sequencing and bioinformatics will no doubt potentiate the field of CTC analyses and, through a better understanding of the biologic events associated with cancer metastasis, help establish novel strategies for cancer treatment. Definition of single molecular targets, such as mutations detected in CTC, may soon influence the treatment of cancer patients. Identification and analyses of circulating stromal cells may aid the prognostic capacity of CTCs as well as move forward studies in CTC/stromal cell clusters and their role in survival and immune evasion, metastatic seeding, and drug resistance.

And finally, the ability to capture live CTCs as well as cStCs will facilitate deeper interrogation of the molecular and phenotypic features of CTCs harboring metastatic ability, and the role of associated stromal cells in clusters plays in the metastatic seeding of these CTCs. In the not so distant future, analyses of "liquid biopsies," to define a large number of molecular targets and potential mechanisms of resistance in a given patient, will become a reality. Such analyses will dynamically guide the treatment of cancer patients and parallel their cancer progression status.

References

1. Malanchi I, et al. Interactions between cancer stem cells and their niche govern metastatic colonization. Nature. 2012;481(7379):85–9.
2. Aceto N, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110–22.
3. Duda DG, et al. Malignant cells facilitate lung metastasis by bringing their own soil. Proc Natl Acad Sci U S A. 2010;107(50):21677–82.
4. Jansson S, et al. Prognostic impact of circulating tumor cell apoptosis and clusters in serial blood samples from patients with metastatic breast cancer in a prospective observational cohort. BMC Cancer. 2016;16:433.
5. Mu Z, et al. Detection and characterization of circulating tumor associated cells in metastatic breast cancer. Int J Mol Sci. 2016;17(10).
6. Wang C, et al. Longitudinally collected CTCs and CTC-clusters and clinical outcomes of metastatic breast cancer. Breast Cancer Res Treat. 2017;161(1):83–94.
7. Braun S, et al. A pooled analysis of bone marrow micrometastasis in breast cancer. N Engl J Med. 2005;353(8):793–802.
8. Campton DE, et al. High-recovery visual identification and single-cell retrieval of circulating tumor cells for genomic analysis using a dual-technology platform integrated with automated immunofluorescence staining. BMC Cancer. 2015;15:360.
9. Alix-Panabieres C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. Clin Chem. 2013;59(1):110–8.
10. Balic M, et al. Micrometastasis: detection methods and clinical importance. Cancer Biomark. 2010;9(1–6):397–419.
11. Lin H, et al. Disseminated and circulating tumor cells: role in effective cancer management. Crit Rev Oncol Hematol. 2011;77(1):1–11.
12. Rawal S, et al. Identification and quantitation of circulating tumor cells. Annu Rev Anal Chem (Palo Alto, Calif). 2017;10(1):321–43.
13. Gabriel MT, et al. Circulating tumor cells: a review of non-EpCAM-based approaches for cell enrichment and isolation. Clin Chem. 2016;62(4):571–81.
14. Rosenberg R, et al. Comparison of two density gradient centrifugation systems for the enrichment of disseminated tumor cells in blood. Cytometry. 2002;49(4):150–8.
15. Cote RJ, et al. Immunopathology of adrenal and renal cortical tumors. Coordinated change in antigen expression is associated with neoplastic conversion in the adrenal cortex. Am J Pathol. 1990;136(5):1077–84.
16. Cote RJ, et al. Role of immunohistochemical detection of lymph-node metastases in management of breast cancer. International Breast Cancer Study Group. Lancet. 1999;354(9182):896–900.
17. Nagrath S, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. 2007;450(7173):1235–9.
18. Cristofanilli M, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. 2004;351(8):781–91.
19. Liu Z, et al. Negative enrichment by immunomagnetic nanobeads for unbiased characterization of circulating tumor cells from peripheral blood of cancer patients. J Transl Med. 2011;9:70.
20. Lustberg M, et al. Emerging technologies for CTC detection based on depletion of normal cells. Recent Results Cancer Res. 2012;195:97–110.
21. Naume B, et al. Immunomagnetic techniques for the enrichment and detection of isolated breast carcinoma cells in bone marrow and peripheral blood. J Hematother. 1997;6(2):103–14.
22. Iinuma H, et al. Detection of tumor cells in blood using CD45 magnetic cell separation followed by nested mutant allele-specific amplification of p53 and K-ras genes in patients with colorectal cancer. Int J Cancer. 2000;89(4):337–44.
23. Maheswaran S, et al. Detection of mutations in EGFR in circulating lung-cancer cells. N Engl J Med. 2008;359(4):366–77.
24. Sequist LV, et al. The CTC-chip: an exciting new tool to detect circulating tumor cells in lung cancer patients. J Thorac Oncol. 2009;4(3):281–3.
25. Fleischer RL. Cancer filter deja vu. Science. 2007;318(5858):1864.
26. Vona G, et al. Isolation by size of epithelial tumor cells : a new method for the immunomorphological and molecular characterization of circulating tumor cells. Am J Pathol. 2000;156(1):57–63.
27. Paterlini-Brechot P, Benali NL. Circulating tumor cells (CTC) detection: clinical impact and future directions. Cancer Lett. 2007;253(2):180–204.
28. Vona G, et al. Impact of cytomorphological detection of circulating tumor cells in patients with liver cancer. Hepatology. 2004;39(3):792–7.
29. Lin HK, et al. Portable filter-based microdevice for detection and characterization of circulating tumor cells. Clin Cancer Res. 2010;16(20):5011–8.
30. Zheng S, et al. Membrane microfilter device for selective capture, electrolysis and genomic analysis of human circulating tumor cells. J Chromatogr A. 2007;1162(2):154–61.
31. Ao Z, et al. Identification of cancer-associated fibroblasts in circulating blood from patients with metastatic breast Cancer. Cancer Res. 2015;75(22):4681–7.
32. Farace F, et al. A direct comparison of CellSearch and ISET for circulating tumour-cell detection in patients with metastatic carcinomas. Br J Cancer. 2011;105(6):847–53.
33. Davis JA, et al. Deterministic hydrodynamics: taking blood apart. Proc Natl Acad Sci U S A. 2006;103(40):14779–84.
34. Huang LR, et al. Continuous particle separation through deterministic lateral displacement. Science. 2004;304(5673):987–90.
35. Inglis DW, et al. Critical particle size for fractionation by deterministic lateral displacement. Lab Chip. 2006;6(5):655–8.
36. Wang XB, et al. Cell separation by dielectrophoretic field-flow-fractionation. Anal Chem. 2000;72(4):832–9.
37. Becker FF, et al. Separation of human breast cancer cells from blood by differential dielectric affinity. Proc Natl Acad Sci U S A. 1995;92(3):860–4.
38. Gascoyne PR, et al. Dielectrophoretic separation of cancer cells from blood. IEEE Trans Ind Appl. 1997;33(3):670–8.
39. Balic M, et al. Novel immunofluorescence protocol for multimarker assessment of putative disseminating breast cancer stem cells. Appl Immunohistochem Mol Morphol. 2011;19(1):33–40.
40. Raimondi C, et al. Epithelial-mesenchymal transition and stemness features in circulating tumor cells from breast cancer patients. Breast Cancer Res Treat. 2011;130(2):449–55.
41. Cote RJ, et al. Detection of occult bone marrow micrometastases in patients with operable lung carcinoma. Ann Surg. 1995;222(4):415–23. discussion 423-5

42. Bartkowiak K, et al. Discovery of a novel unfolded protein response phenotype of cancer stem/progenitor cells from the bone marrow of breast cancer patients. J Proteome Res. 2010;9(6):3158–68.
43. Tlsty TD. Stromal cells can contribute oncogenic signals. Semin Cancer Biol. 2001;11(2):97–104.
44. Allinen M, et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell. 2004;6(1):17–32.
45. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9(4):274–U65.
46. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
47. Kalluri R, Zeisberg M. Fibroblasts in cancer. Nat Rev Cancer. 2006;6(5):392–401.
48. Orimo A, Weinberg RA. Stromal fibroblasts in cancer: a novel tumor-promoting cell type. Cell Cycle. 2006;5(15):1597–601.
49. Jones ML, et al. Circulating fibroblast-like cells in men with metastatic prostate cancer. Prostate. 2013;73(2):176–81.
50. Adams DL, et al. Circulating giant macrophages as a potential biomarker of solid tumors. Proc Natl Acad Sci U S A. 2014;111(9):3514–9.
51. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423–37.
52. Nagaraj S, Gabrilovich DI. Myeloid-derived suppressor cells in human cancer. Cancer J. 2010;16(4):348–53.
53. Drews-Elger K, et al. Infiltrating S100A8+ myeloid cells promote metastatic spread of human breast cancer and predict poor clinical outcome. Breast Cancer Res Treat. 2014;148(1):41–59.
54. Azizi E, Wicha MS. Cancer stem cells–the evidence accumulates. Clin Chem. 2013;59(1):205–7.
55. Ginestier C, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell. 2007;1(5):555–67.
56. Al-Hajj M, et al. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983–8.
57. Wang S, et al. CD133+ cancer stem cells in lung cancer. Front Biosci. 2013;18:447–53.
58. Singh SK, et al. Cancer stem cells in nervous system tumors. Oncogene. 2004;23(43):7267–73.
59. Schardt JA, et al. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. Cancer Cell. 2005;8(3):227–39.
60. Cote RJ. Occult metastases: real harm or false alarm? J Thorac Cardiovasc Surg. 2003;126(2):332–3.
61. Balic M, et al. Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. Clin Cancer Res. 2006;12(19):5615–21.
62. Duru N, et al. HER2-associated radiation resistance of breast cancer stem cells isolated from HER2-negative breast cancer cells. Clin Cancer Res. 2012;18:6634.
63. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res. 2008;10(2):R25.
64. Kakarala M, Wicha MS. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. J Clin Oncol. 2008;26(17):2813–20.
65. Abraham BK, et al. Prevalence of CD44+/CD24−/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. Clin Cancer Res. 2005;11(3):1154–9.
66. Wicha MS. Cancer stem cells and metastasis: lethal seeds. Clin Cancer Res. 2006;12(19):5606–7.
67. Giuliano AE, et al. Association of occult metastases in sentinel lymph nodes and bone marrow with survival among women with early-stage invasive breast cancer. JAMA. 2011;306(4):385–93.
68. Reuben JM, et al. Primary breast cancer patients with high risk clinicopathologic features have high percentages of bone marrow epithelial cells with ALDH activity and CD44CD24lo cancer stem cell phenotype. Eur J Cancer. 2011;47(10):1527–36.
69. Theodoropoulos PA, et al. Circulating tumor cells with a putative stem cell phenotype in peripheral blood of patients with breast cancer. Cancer Lett. 2010;288(1):99–106.
70. Wang N, et al. Detection of circulating tumor cells and tumor stem cells in patients with breast cancer by using flow cytometry : a valuable tool for diagnosis and prognosis evaluation. Tumour Biol. 2012;33(2):561–9.
71. Lim J, Thiery JP. Epithelial-mesenchymal transitions: insights from development. Development. 2012;139(19):3471–86.
72. Lee K, Nelson CM. New insights into the regulation of epithelial-mesenchymal transition and tissue fibrosis. Int Rev Cell Mol Biol. 2012;294:171–221.
73. Aktas B, et al. Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. Breast Cancer Res. 2009;11(4):R46.
74. Ignatiadis M, et al. Prognostic value of the molecular detection of circulating tumor cells using a multimarker reverse transcription-PCR assay for cytokeratin 19, mammaglobin a, and HER2 in early breast cancer. Clin Cancer Res. 2008;14(9):2593–600.
75. Yu M, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580–4.
76. Mani SA, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704–15.
77. Morel AP, et al. Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS One. 2008;3(8):e2888.
78. Wellner U, et al. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat Cell Biol. 2009;11(12):1487–95.
79. Scheel C, Weinberg RA. Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? Int J Cancer. 2011;129(10):2310–4.
80. Tiran V, et al. Primary patient-derived lung adenocarcinoma cell culture challenges the association of cancer stem cells with epithelial-to-mesenchymal transition. Sci Rep. 2017;7(1):10040.
81. Friedlein R, et al. Solution-processed, highly-oriented supramolecular architectures of functionalized porphyrins with extended electronic states. Chem Commun (Camb). 2005;15:1974–6.
82. Cheung KJ, et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc Natl Acad Sci U S A. 2016;113(7):E854–63.
83. Jolly MK, et al. Inflammatory breast cancer: a model for investigating cluster-based dissemination. NPJ Breast Cancer. 2017;3:21.
84. Hou JM, et al. Circulating tumor cells as a window on metastasis biology in lung cancer. Am J Pathol. 2011;178(3):989–96.
85. May CD, et al. Epithelial-mesenchymal transition and cancer stem cells: a dangerously dynamic duo in breast cancer progression. Breast Cancer Res. 2011;13(1):202.
86. Lorico A, Rappa G. Phenotypic heterogeneity of breast cancer stem cells. J Oncol. 2011;2011:135039.
87. Fehm T, et al. Determination of HER2 status using both serum HER2 levels and circulating tumor cells in patients with recurrent breast cancer whose primary tumor was HER2 negative or of unknown HER2 status. Breast Cancer Res. 2007;9(5):R74.
88. Fehm T, et al. HER2 status of circulating tumor cells in patients with metastatic breast cancer: a prospective, multicenter trial. Breast Cancer Res Treat. 2010;124(2):403–12.
89. Tobin LA, et al. Targeting abnormal DNA repair in therapy-resistant breast cancers. Mol Cancer Res. 2012;10(1):96–107.

90. Liedtke C, et al. Prognostic impact of discordance between triple-receptor measurements in primary and recurrent breast cancer. Ann Oncol. 2009;20(12):1953–8.
91. Nogami T, et al. The discordance between primary breast cancer lesions and pulmonary metastatic lesions in expression of aldehyde dehydrogenase 1-positive cancer cells. Breast Cancer. 2014;21(6):698–702.
92. Chakrabarty A, et al. Trastuzumab-resistant cells rely on a HER2-PI3K-FoxO-survivin axis and are sensitive to PI3K inhibitors. Cancer Res. 2013;73(3):1190–200.
93. Sun YF, et al. Circulating tumor cells: advances in detection methods, biological issues, and clinical relevance. J Cancer Res Clin Oncol. 2011;137(8):1151–73.
94. Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer. 2009;9(4):302–12.
95. Klein CA, et al. Combined transcriptome and genome analysis of single micrometastatic cells. Nat Biotechnol. 2002;20(4):387–92.
96. Husemann Y, et al. Systemic spread is an early step in breast cancer. Cancer Cell. 2008;13(1):58–68.
97. Stoecklein NH, Klein CA. Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. Int J Cancer. 2010;126(3):589–98.
98. Magbanua MJ, et al. Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients. Cancer Res. 2013;73(1):30–40.
99. Fridlyand J, et al. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer. 2006;6:96.
100. Paris PL, et al. Functional phenotyping and genotyping of circulating tumor cells from patients with castration resistant prostate cancer. Cancer Lett. 2009;277(2):164–73.
101. Magbanua MJ, et al. Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer. BMC Cancer. 2012;12:78.
102. Heitzer E, et al. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. Int J Cancer. 2013;133(2):346–56.
103. van Beers EH, et al. A multiplex PCR predictor for aCGH success of FFPE samples. Br J Cancer. 2006;94(2):333–7.
104. Amado RG, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26(10):1626–34.
105. Paez JG, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500.
106. Punnoose EA, et al. Molecular biomarker analyses using circulating tumor cells. PLoS One. 2010;5(9):e12517.
107. Dharmasiri U, et al. High-throughput selection, enumeration, electrokinetic manipulation, and molecular profiling of low-abundance circulating tumor cells using a microfluidic system. Anal Chem. 2011;83(6):2301–9.
108. Yang MJ, et al. Enhancing detection of circulating tumor cells with activating KRAS oncogene in patients with colorectal cancer by weighted chemiluminescent membrane array method. Ann Surg Oncol. 2010;17(2):624–33.
109. Mostert B, et al. KRAS and BRAF mutation status in circulating colorectal tumor cells and their correlation with primary and metastatic tumor tissue. Int J Cancer. 2013;133(1):130–41.
110. Jiang Y, et al. Detection of androgen receptor mutations in circulating tumor cells in castration-resistant prostate cancer. Clin Chem. 2010;56(9):1492–5.
111. Smirnov DA, et al. Global gene expression profiling of circulating tumor cells. Cancer Res. 2005;65(12):4993–7.
112. Barbazan J, et al. Molecular characterization of circulating tumor cells in human metastatic colorectal cancer. PLoS One. 2012;7(7):e40476.
113. Sieuwerts AM, et al. mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. Clin Cancer Res. 2011;17(11):3600–18.
114. Munzone E, et al. Changes of HER2 status in circulating tumor cells compared with the primary tumor during treatment for advanced breast cancer. Clin Breast Cancer. 2010;10(5):392–7.
115. Cayrefourcq L, et al. Establishment and characterization of a cell line from human circulating colon cancer cells. Cancer Res. 2015;75(5):892–901.
116. Yu M, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. Science. 2014;345(6193):216–20.
117. Gao D, et al. Organoid cultures derived from patients with advanced prostate cancer. Cell. 2014;159(1):176–87.
118. Zhang Z, et al. Expansion of CTCs from early stage lung cancer patients using a microfluidic co-culture model. Oncotarget. 2014;5(23):12383–97.
119. Jordan NV, et al. HER2 expression identifies dynamic functional states within circulating breast cancer cells. Nature. 2016;537(7618):102–6.
120. Cheung KJ, Ewald AJ. A collective route to metastasis: seeding by tumor cell clusters. Science. 2016;352(6282):167–9.
121. Hodgkinson CL, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014;20(8):897–903.
122. Powell AA, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One. 2012;7(5):e33788.
123. Ao Z, et al. Thermoresponsive release of viable microfiltrated circulating tumor cells (CTCs) for precision medicine applications. Lab Chip. 2015;15(22):4277–82.
124. Engler AJ, et al. Matrix elasticity directs stem cell lineage specification. Cell. 2006;126(4):677–89.
125. Pelham RJ, Wang YL. Cell locomotion and focal adhesions are regulated by substrate flexibility. Proc Natl Acad Sci U S A. 1997;94(25):13661–5.
126. Zeng YY, et al. Substrate stiffness regulates B-cell activation, proliferation, class switch, and T-cell-independent antibody responses in vivo. Eur J Immunol. 2015;45(6):1621–34.
127. Narayanan K, et al. Extracellular matrix-mediated differentiation of human embryonic stem cells: differentiation to insulin-secreting beta cells. Tissue Eng Part A. 2014;20(1–2):424–33.
128. Engler AJ, et al. Myotubes differentiate optimally on substrates with tissue-like stiffness: pathological implications for soft or stiff microenvironments. J Cell Biol. 2004;166(6):877–87.
129. Schrader J, et al. Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. Hepatology. 2011;53(4):1192–205.
130. Tilghman RW, et al. Matrix rigidity regulates cancer cell growth and cellular phenotype. PLoS One. 2010;5(9):e12905.
131. Butcher DT, Alliston T, Weaver VM. A tense situation: forcing tumour progression. Nat Rev Cancer. 2009;9(2):108–22.
132. Caplan AI, Correa D. PDGF in bone formation and regeneration: new insights into a novel mechanism involving MSCs. J Orthop Res. 2011;29(12):1795–803.
133. Correa D, et al. Mesenchymal stem cells regulate melanoma cancer cells extravasation to bone and liver at their perivascular niche. Int J Cancer. 2016;138(2):417–27.
134. Correa D. Mesenchymal stem cells during tumor formation and dissemination. Curr Stem Cell Rep. 2016;2(2):174–82.
135. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9(4):274–84.
136. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer. 2009;9(4):239–52.

137. Dominici M, et al. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy. 2006;8(4):315–7.
138. Kang Y, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 2003;3(6):537–49.
139. Caplan AI. Review: mesenchymal stem cells: cell-based reconstructive therapy in orthopedics. Tissue Eng. 2005;11(7–8):1198–211.
140. Caplan AI. Mesenchymal stem cells. J Orthop Res. 1991;9(5):641–50.
141. Sacchetti B, et al. Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. Cell. 2007;131(2):324–36.
142. da Silva Meirelles L, Caplan AI, Nardi NB. In search of the in vivo identity of mesenchymal stem cells. Stem Cells. 2008;26(9):2287–99.

7 Circulating Cell-Free DNA for Molecular Diagnostics and Therapeutic Monitoring

Natasha B. Hunter, Julia A. Beaver, and Ben Ho Park

Introduction

The presence of circulating cell-free DNA (ccfDNA) in the human blood was originally discovered in 1948 by Mandel and Métais, a full 32 years before Frederick Sanger would win the Nobel prize for DNA sequencing [1]. At the time, the discovery was of purely biochemical interest, with no clinical utility. But with technological advances making molecular studies feasible and affordable in recent decades, interest in the potential medical applications of ccfDNA has resurfaced. Initially, scientists explored uses in maternal–fetal medicine and oncology, with more recent efforts broadening to include other disease processes such as sepsis, myocardial infarction, stroke, and diabetes [2–5].

This free-floating or "naked" DNA in the blood appears to derive from cells in both healthy and diseased states. Its origin is not fully understood and could involve multiple mechanisms. Some portion of ccfDNA is likely shed into the circulation by macrophage release of necrotic or apoptotic cellular debris [6], or some cells may actively secrete ccfDNA into the circulation [7].

Circulating cell-free DNA has been found in most human fluids: whole blood, serum, plasma, urine, and cerebral spinal fluid, with fragments ranging from 70 to 1200 base pairs in length [8–12]. Its half-life is short, on the order of 15 min to a few hours; it is quickly cleared by the kidney and liver [10, 13–15]. Patients with metastatic cancer, trauma, myocardial infarction, and sepsis display higher concentrations of overall ccfDNA than normal controls [9, 16–20], perhaps because all these processes involve high cell turnover.

Researchers are investigating numerous potential applications for ccfDNA-based assays in various disease states, but few have been approved by national regulatory bodies, and no standard testing platform exists. It should be mentioned that this review discusses free DNA in plasma, with an emphasis on oncologic applications; this is distinct from circulating tumor cells (CTCs), which are also being investigated as cancer biomarkers.

Applications in Maternal–Fetal Medicine

In the late 1970s, fetal cells were first discovered in the maternal circulation, and subsequent work demonstrated that small amounts of fetal ccfDNA was also present in maternal blood [21–23]. The prospect of noninvasive prenatal testing (NIPD) using maternal–fetal ccfDNA fueled interest in the field. Fetus-derived ccfDNA was determined to be likely placental in origin and shorter than maternal DNA (with fetal at <300 base pairs) [24]. The fetal fraction of ccfDNA accounts for approximately 10% of the total ccfDNA, although studies differ slightly, and the percentage may rise as gestation progresses [25–27].

The fetal ccfDNA in maternal circulation not only represents a small portion of the total, it is also haplotypically identical to the mother, so paternally inherited characteristics have been comparatively easier to decipher. For example, investigators in the 1990s were already demonstrating fetal rhesus D genotyping and fetal sex assessment by PCR for Y chromosome sequences [22, 28]. Later Fan et al. used direct shotgun sequencing followed by chromosome mapping to establish over- or underrepresentation of chromosomes in maternal plasma ccfDNA, thereby identifying potential aneuploidy (i.e., an abnormal number of chromosomes, most commonly trisomies 21, 18, or 13) [29]. Numerous other investigators have employed next-generation massively parallel sequencing (MPS) to detect aneuploidy, many in large-scale studies [30–32]. Indeed, NIPD using ccfDNA has already entered clinical practice: in 2015, the American College of Maternal Fetal Medicine released guidelines on cell-free DNA screening for fetal aneuploidy, cautiously endorsing ccfDNA testing for the most common trisomies and for sex chromosome analysis [33].

N. B. Hunter · J. A. Beaver · B. H. Park (✉)
The Sidney Kimmel Comprehensive, Cancer Center at Johns Hopkins, Baltimore, MD, USA
e-mail: bpark2@jhmi.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_7

Determining the maternally inherited portion of the fetal genome initially presented a difficult problem, as it required detection of genomic information that is identical to its background. Lo and colleagues were able to develop a method called "relative haplotype dosage analysis" (RHDO), which used highly accurate quantification to devise a solution; that is, if 10 percent of ccfDNA is fetal, then there should be a 5% overrepresentation of the maternal haplotype in the total ccfDNA [34]. Further studies are ongoing to detect monogenic diseases, which present similarly tough detection problems, with autosomal recessive conditions posing a particular challenge [32].

Applications in Cancer Diagnostics

Using radioimmunoassays, Leon et al. in 1977 first demonstrated higher levels of circulating ccfDNA in cancer patients compared to healthy individuals [35]. The range of ccfDNA concentrations in patients with malignancies varied substantially, however—between 0 and >1000 ng/ml, while normal subjects exhibited ccfDNA concentrations between 0 and 100 ng/ml [14, 36–38]. Given this overlap in ccfDNA concentrations in normal and cancer patients, it seemed unlikely that the total quantity of ccfDNA would prove a reliable diagnostic tool. This was confirmed in 2016 with the first large prospective study showing ccfDNA levels in NSCLC could not predict disease recurrence [39].

The promise of ccfDNA in cancer applications, therefore, depends on distinguishing a small population of DNA shed by the tumor from a larger population of normal patient DNA. In this regard, the task bears similarity to fetal DNA detection, with sometimes only a few small changes distinguishing the proverbial needle from the haystack of normal host genetic material. Indeed, studies have demonstrated that a patient with a solid tumor comprising 3×10^{10} cells, tumor DNA will make up about 3.3% of the already small amount of ccfDNA found in that patient's bloodstream daily [9].

The changes that signal the presence of plasma tumor DNA, or ptDNA, include mutations, epigenetic alterations, amplifications, and rearrangements resulting from translocations and deletions or insertions. Multiple groups have demonstrated that the size of ptDNA molecules is smaller than that of plasma DNA derived from normal cells and typically ranges from 70 to 200 bp [8, 9, 40]. Some mutations and rearrangements in tumor suppressor genes or oncogenes drive the development and progression of a given cancer; others are so-called "passenger" mutations, i.e., genetic alterations that probably result from tumor genetic instability but carry no functional consequence. All these mutations represent somatic changes (i.e., not present in the patient's germline DNA and therefore not heritable—but also not present in the patient's normal tissue) and thus potential cancer markers [41].

The implications of all of this are tantalizing: the short half-life of ptDNA lends itself to providing a snapshot of tumor burden and response to therapies. The ability to use a blood sample to perform a "liquid biopsy" offers a noninvasive, real-time assessment of both qualitative (molecular tumor genotype) and quantitative (burden of disease) aspects of a patient's tumor, providing a potentially more sensitive analysis than radiography or even surgery. Moreover, ptDNA could be a more specific biomarker than others in use clinically today, as it represents direct evidence of tumor, not just an associated nonspecific proxy of disease.

As an example, one can envision testing postsurgical patients for residual micrometastatic disease to assess the need for adjuvant treatment, possibly preventing the administration of unnecessary toxic systemic therapies. Such a validated technique could guide the substitution of various therapies (e.g., chemotherapy vs. hormonal vs. biologic) in the adjuvant or metastatic setting by triggering a change when personalized DNA markers do not respond to therapy or when new markers arise. Real-time assessment of the molecular profile of a tumor could drive rapid and mutation-enriched clinical trial enrollment and create new surrogate endpoints, expediting drug approval. The sections that follow will discuss current and developing clinical applications, along with anticipated directions and possible roadblocks.

The Rapid Advance of Genomic Technologies

As researchers hail ptDNA as a new frontier in cancer biomarkers, it is easy to forget that we have been performing blood-based molecular genetic testing in hematologic malignancies for decades. The Philadelphia chromosome that defines chronic myelocytic leukemia (CML) was identified in 1959 using chromosomal electron microscopy; its transcript was later characterized by the quantitative real-time PCR (qPCR) methods developed by Heid et al. [42]. This standardized assay has enabled real-time monitoring of disease burden and response to treatment using peripheral blood or bone marrow samples [43].

Because leukemias are by definition cancers of the blood, abundance of circulating leukemic (i.e., tumor) cells in the peripheral blood and bone marrow facilitates easier detection of fusion transcripts. Modern assays also identify other commonly occurring mutations, e.g., tandem repeats of the FLT3 gene. The majority of hematologic malignancies display a limited array of driver mutations, and many of these define the disease and establish the diagnosis. In contrast, the ability to identify the corresponding circulating tumor cells (CTCs) for most solid malignancies has been hindered by

low sensitivity (though newer capture methods have shown promise) [44]. And while cell-free tumor DNA in solid malignancies is more easily detected than CTCs, the rarity of predictable recurrent somatic rearrangements in these diseases has complicated assay development.

Over recent decades, new technologies have emerged that allow faster, cheaper gene exploration as well as highly sensitive detection of known genes, placing clinical applications for ptDNA in solid tumors within ever-easier reach. Simple PCR entered common practice in the late 1980s, and dramatically advanced the ability to detect genetic changes. In the 1990s, several labs independently developed variations on a new, highly sensitive PCR method by which individual strands of DNA were amplified separately, generating a binary result for each molecule and allowing both detection and quantification [45]; Kinzler and Vogelstein coined the term "digital" PCR for this technique in their 1999 paper describing sample partitioning in 384-well microplates [46]. The same group in 2005 developed an emulsion-based digital PCR (ddPCR) method called BEAMing (for Beads, Emulsion, Amplification, and Magnetics, outlining the steps involved); using this semiautomated technique, the group was able to identify patients with point mutations in mutant APC molecules in both early-stage and metastatic colorectal cancer patients [9].

In the mid-2000s, high-throughput DNA sequencing, termed next-generation sequencing (NGS), transformed genomics by enabling DNA processing on an order of magnitude larger than prior Sanger methods could accomplish in the same timeframe [47]. NGS technologies advanced rapidly, and new platforms along with fierce competition within the industry have driven costs down exponentially. Sequencing delivers not only detection and quantification of known mutations as digital PCR provides but also allows for identification of new mutations and alterations. This has dramatically accelerated the discovery of patterns of mutation and tumor evolution in cancer research. The newest methods also boast high sensitivity; e.g., a platform called TAM-seq described by Forshew et al. can detect mutations with allele frequencies as low as 2 percent in the ptDNA of patients with advanced cancer [48]. A technique called SafeSeqs employs a "barcoding" method to tag DNA strands before amplifying them, producing tagged "families" of clones that must be 95% identical to be called as a true variant, thereby reducing error. Duplex sequencing also employs tags, marking double-stranded DNA before PCR amplification and then establishing single-stranded consensus sequences that are compared with their complementary strands to yield a "duplex consensus sequence"; this has greatly improved accuracy, allowing detection of one mutation among 10^7 bases [49, 50].

Despite these refinements, NGS still lags behind PCR in sensitivity and lacks the speed and cost-efficiency of PCR for applications where discovery of new mutations is not an issue. Therefore, researchers developing ptDNA assays must weigh depth against breadth of DNA sequencing, as well as time constraints and cost. To elaborate, all sequencing involves some error; increasing the number of times one "reads" each string of bases reduces this error. The number of reads is the "depth" of sequencing. One can sequence larger portions of the genome at a lower depth and gain a broader range of information, or "breadth" of sequencing, but with less certainty as to its accuracy. In other words, knowing exactly which mutation to expect, e.g., a KRAS G12 V mutation in pancreatic cancer, allows one to deploy a sensitive, fast, and cheap test (PCR) for this specific alteration, with the understanding that if other mutations exist, they will not be detected. Conversely, one can design a broader, more expensive and usually slower assay using NGS to look for a variety of mutations, with the understanding that sensitivity for each of these may be lower.

These questions are particularly relevant in ptDNA, because of the small amounts of DNA in circulation. Leary and colleagues demonstrated a 0.61- to 1.97-fold copy number increase in the plasma of cancer patients compared to normalized controls, but the assay succeeded only when the percentage of ptDNA compared to ccfDNA was at least 0.75%, at which point it carried a sensitivity of >90% and specificity of >99% [51]. This study illustrates the critical consideration of depth of sequencing, given the low concentration of ptDNA at baseline, especially in the setting of early-stage cancers or minimal residual disease.

Without large amounts of DNA available for testing, one must balance breadth and depth when choosing an assay. For this reason, many studies use targeted NGS (focused on frequently mutated genes, e.g., TP53) to find alterations in more plentiful DNA from tumor tissue and then employ PCR to detect those mutations in plasma, which allows more sensitive, rapid, and inexpensive serial monitoring. This method, however, obviously requires the availability of tumor tissue, which is frequently scant or difficult to obtain. Moreover, initial biopsies may not be reflective of the tumor at the point of ptDNA assay, as the cancer cells may have developed new mutations.

The appropriate depth of coverage to achieve an adequate sensitivity for ptDNA studies therefore depends on the use indication. For assessing minimal residual disease in early-stage cancer, we would suggest depths of coverage in the 10,000- to 100,000-fold range to achieve a reliable sensitivity of 0.01% to 0.02% allelic frequency. This is best achieved with digital PCR and/or barcoding amplicon sequencing. For metastatic disease, finding such rare clonal populations may not be needed, so using NGS in the 1000- to-10,000-fold coverage range is likely adequate to obtain a 1% to 0.1% allelic frequency sensitivity.

Sensitivity of ptDNA and Concordance with Tissue

Two related questions arise when considering ptDNA as a reliable proxy for disease presence or burden: sensitivity of detection in the plasma and concordance between the mutations found in plasma and tumor tissue. Prior to NGS technologies becoming commonplace, investigators typically employed PCR to identify common mutations in tumor biopsies and then looked for these same mutations in plasma; in this case, the sensitivity of detection is defined by its concordance with tissue biopsy. But as sequencing technology has advanced, along with the field's understanding of tumor evolution and heterogeneity, tissue mutations may be proving less useful as a gold standard, with de novo sequencing of plasma providing a complementary and potentially broader look at the tumor mutational landscape.

Some of the earliest ptDNA assays employed PCR to analyze microsatellite instability and loss of heterozygosity (LOH) in plasma of breast cancer patients [41] and detected KRAS mutations in pancreatic cancer [52]. Other groups examined mutant KRAS in a primary tumor and identified corresponding KRAS mutations in the plasma, with higher sensitivity than in prior assays [53, 54]. These and studies like them provided an exciting proof of concept, spurring interest in ptDNA as a biomarker.

From the beginning, studies reported varying sensitivity. Mutations in TP53 were found in 42.9% of the plasma DNA from patients harboring TP53 mutations in their tumor [55]. Other studies reported 100% concordance [10]. Some of these discrepancies could be explained by differences in study design—e.g., retrospective vs. prospective trials. Lab techniques also differed—e.g., improperly collected or inadequately spun blood can result in a high fraction of white blood cell DNA diluting the plasma sample. Issues of ptDNA dilution by total ccfDNA may have complicated the detection of ptDNA in several studies [14, 56–58]. Also, the raw number of genome equivalents sampled by the investigators is a measure of DNA yield, which correlates with sensitivity; studies vary in the amount of plasma collected, which affects results [59].

Generally, metastatic disease has been consistently easier to detect in ptDNA, likely owing to bulkier disease shedding larger amounts DNA into the circulation. For example, Diehl and colleagues found 100% concordance between APC mutations in the plasma and solid tumor specimens of six metastatic colorectal cancer patients [9], but the group's subsequent analysis of early-stage colorectal cancer patients with proven APC mutations detected mutant APC DNA in only 63% of corresponding plasma samples. Similarly, Bettegowda and colleagues looked at multiple tumor types and were able to detect ptDNA in 75% of patients with advanced disease, but less than 50% in some tumor types with localized tumors (this also suggested varying ptDNA levels by tumor type, still an issue under investigation) [60]. However, this study used varying analytes (serum tumor DNA and ptDNA), as well as multiple methods for mutation detection (NGS and digital PCR), so it is unclear how much information from this study can be extrapolated into newer studies using higher-quality analytes and uniform detection platforms.

Sensitivity continues to improve for early-stage disease and in settings (e.g., postsurgical) where detection of micrometastatic disease is key. Technological advances are improving sensitivity, e.g., a 2016 study assay of KRAS in pancreatic cancer patients undergoing resection could detect down to a mutation prevalence of 0.01–0.1%, corresponding to 1 mutant copy per 1000–10,000 wild-type copies [61]. Investigators have also recognized that given lower concentrations of ptDNA in early-stage patients, increasing the number of genome equivalents sampled in early-stage cancer patients should also increase assay sensitivity [35].

As investigators worked to establish concordance between plasma and tissue in various tumor types, many noted issues arising when using archival specimens to assess mutational status in patients with metastatic disease [62, 63]. For example, our lab evaluated "hotspot" PIK3CA mutations (a gene commonly mutated in breast and other cancers) in metastatic breast cancer patients from 49 archival matched tumor and plasma samples for exon 9 (E542K and E545K) and 20 (H1047R) mutations. We found 100% concordance between the specific PIK3CA mutation in each tumor and its matched plasma sample. However, a subsequent prospective study by our group identified only 70% concordant PIK3CA mutations between tumor tissue and peripheral blood. This disparate result could be explained by tumor heterogeneity and clonal evolution: the prospective study used archived primary cancer tissues from any source and blood drawn at the time of study enrollment, while in the retrospective trial, blood and tissue samples were acquired concurrently [64]. Notably, in the prospective study, discordant results were only seen when tissue and blood were collected greater than 3 years apart.

Numerous studies have now characterized tumor heterogeneity in ptDNA; for example, its concordance with tissue mutations was confirmed recently by de Mattos Arruda and colleagues, who tracked mutations in both plasma and serial biopsies (primary tumor and metastatic disease) in a single breast cancer patient. They showed that high-depth massively parallel sequencing (MPS) could capture all mutations in the primary tumor and in liver metastases, indicating that from a mutation standpoint, ptDNA could represent an alternative to biopsy in the metastatic setting, if performed in a timely fashion [65]. Rothe and colleagues reported more discordant results in 17 patients with metastatic breast cancer, in which concurrent plasma and tissue samples showed 4 of

17 with different mutations identified (with the rest being concordant), indicating that ptDNA may in some cases provide complementary information to metastatic biopsy [66].

Regardless of these differences, one distinct advantage ptDNA offers is availability; we recently conducted a feasibility study, Individualized Mutational Analysis Guides Efforts (IMAGE), attempting to acquire metastatic biopsies and ptDNA for evaluation by our molecular tumor board within a 28-day period; the study was halted owing to inability to acquire tumor tissue in the requisite time frame, but plasma studies proved quite amenable to acquisition and evaluation in a clinically actionable interval [67].

Differing mutational and genomic signatures between primary and metastatic disease sites complicates validation efforts. The majority of cancers are thought to arise through the accumulation of 3 to 8 "driver" mutations [68]. In theory, these mutations are the initial foundation or "trunk" of the cancer's evolutionary tree. As cancer cells progress and metastasize to different sites in the body, they acquire additional driver mutations that are then unique to those subpopulations. These "branch/leaf" mutations can continue to evolve, especially when selective pressures such as chemotherapy and other therapies are applied.

While ptDNA may in some cases offer a more complete profile of a tumor's mutational profile than tissue biopsy, this picture is also perhaps more error-prone owing to smaller amounts of DNA, which often undergoes additional cycles of PCR amplification prior to sequencing. Tumor heterogeneity also makes results harder to interpret; e.g., if a mutation displays a low clonal allele fraction within a solid tumor, wild-type sequences shed from other tumor and normal cells may reduce the relative fraction of ptDNA for that mutation and may fail to reflect the overall tumor burden [52, 69, 70]. The use of multiple somatic alterations as markers can mitigate some of these concerns. Investigators have also considered the possibility of using stool, urine, and increased volumes of plasma to improve the sensitivity of detecting rare mutations within ptDNA [71, 72].

Residual Disease, Recurrence, and Tumor Dynamics

Given the complexities described above, the story of how ptDNA will fit into cancer detection, monitoring of response to therapies, and clinical decision making continues to evolve.

In 2008, Diehl and colleagues correlated amounts of ptDNA with rising or falling tumor burden using BEAMing for four genes (APC, PIK3CA, TP53, and KRAS) in 18 colorectal cancer patients [10]. Recent studies demonstrate similar results using personalized assays. For example, in a 2016 comparison of 30 women with metastatic breast cancer, ptDNA displayed a greater dynamic range and better correlation with changes in tumor burden during treatment than CTCs and a greater sensitivity than the protein tumor marker CA 15–3 or CTCs (97% vs. 78% and 87%, respectively).

If ptDNA reliably reflects tumor dynamics, this suggests clinical utility at the very least in prognostication, for guiding clinical decision making, defining clinical trial populations enriched for high-risk features, or assisting in patient–physician communication. A recent study risk-stratifying breast cancer patients into ptDNA high, low, or free correlated patient status with DFS and OS [73]. Olsson and colleagues used whole genome sequencing to identify tumor-specific mutations in 20 breast cancer patients and showed that these personalized signatures discriminated between patients with and without recurrence, preceding clinical detection in 86% with an average lead time of 11 months [74]. Pietrasch et al. were recently able to develop sequencing libraries from ptDNA alone—not preceded by sequencing of tissue biopsy—in patients with various stages of pancreatic cancer and were able to correlate detection of ptDNA with poor OS [75].

Tracking multiple tumor-specific mutations over multiple time points appears to have some sensitivity in detecting minimal residual disease, also potentially guiding therapies. Tie et al. used ptDNA to detect MRD in stage II colon cancer patients, a population in which risk-stratification to determine potential benefit from adjuvant chemotherapy has proven challenging. The researchers subjected 231 operative tumor samples to MPS, revealing somatic mutations in 99.6% of cases. Patients positive for ptDNA postoperatively had a reduced recurrence-free survival (RFS) compared with ptDNA-negative patients. The authors went on to perform mutation tracking for ptDNA-positive patients who underwent chemotherapy and were able to demonstrate correlation between mutational load and radiologic recurrence [76].

A similar effort by Garcia Murillas and colleagues used MPS to detect patient-specific mutations in the tumors of 55 patients with early-stage breast cancer who received neoadjuvant therapy. Using PCR assays personalized for each tumor, they tested patient plasma collected prior to therapy and serially after completion of treatment. ptDNA detection at baseline did not predict disease-free survival, but ptDNA found in a single postsurgical sample predicted early relapse (HR 25, C-index 0.78), and mutation tracking (i.e., serial testing) improved relapse prediction. Moreover, ptDNA became detectable at a median of 7.9 months before clinical relapse, implying that the assay could potentially direct intervention prior to macroscopic recurrence, at a time point when salvage and cure might prove feasible [77].

Many hope that ptDNA may offer a more practical substitute for metastatic biopsy. Lebofsky scanned 46 genes and more than 6800 mutations from the COSMIC database using a multiplexed NGS panel on both ptDNA and tissue and was

able to detect mutations in 27 patients; however, matching ptDNA to tumor was impossible in several patients due to inadequate biopsy material [78]. Our group's IMAGE study, mentioned above, while stopped owing to similar problems with lack of biopsy tissue, did demonstrate that mutations in the patients' plasma could be assessed by NGS within 28 days and that tracking mutations was reflective of tumor burden and response to therapies [67].

Resistance Mutations and Agile Treatment Algorithms

One powerful application of ptDNA is to test for the emergence of resistance clones. Indeed, this represents the only area in which US and European regulatory bodies have already approved clinical assays; work in this field is progressing rapidly and may lead to other resistance assays entering standard practice in the near future.

In 2011, Taniguchi and colleagues demonstrated the ability to detect second-site ptDNA T790 M epidermal growth factor receptor (EGFR) mutations in lung cancer patients treated with EGFR tyrosine kinase inhibitors (TKIs); this mutation confers resistance to standard EGFR-targeted medications such as erlotinib and gefitinib and indicates a possible response to third-generation EGFR TKIs such as osimertinib [79]. This resistance mutation was observed in a significant fraction of EGFR-inhibitor naïve patients, suggesting the natural existence of a minority population of cancer cells that subsequent EGFR inhibition may select for. Conversely, Piotrowska and colleagues demonstrated that T790 M may also revert to wild type in response to third-generation therapies and that ptDNA can detect the emergence of these "new" wild-type clones [80]. Other groups have also created ptDNA assays for T790 M mutations [81, 82]; one study was able to show ptDNA emergence of a resistance clone predating clinical progression by months [83]. While most assays rely on digital PCR to detect previously characterized mutations, Chabon et al. recently developed a targeted NGS method (CAPP-Seq) to study emerging changes in ptDNA from patients receiving a third-generation TKI (rociletinib, no longer in clinical development) and were able to identify several novel alterations in EGFR, ERBB2, and most notably MET, all suspected to confer resistance to TKI therapies [84].

Such work presaged the announcement in 2016 of the first Food and Drug Administration (FDA)-approved plasma-based companion diagnostic, the cobas EGFR Mutation Test v2. The approval for the assay relied on testing of specimens from the large EURTAC trial [85] and complemented the FDA's previous approval granted in 2013 for EGFR testing in tissue as a selection for treatment with Tarceva. The test detects EGFR mutations (exon 19 deletion or exon 21 [L858R] substitution mutations) in NSCLC patients' blood samples. If these mutations are not detected in the plasma, then a tumor biopsy should be performed to better indicate who may benefit from Tarceva treatment [86].

Resistance to endocrine therapy in hormone-positive positive breast cancers may prove the next relatively low-hanging fruit for ptDNA validation. An emerging literature demonstrates that mutations in the gene encoding for estrogen receptor-alpha, ESR1, can develop in the tumors of metastatic breast cancer patients whose disease progresses on hormone therapy. Our group was able to retrospectively and prospectively detect ESR1 mutations in plasma of patients where a simultaneous biopsy proved negative for the same mutation, suggesting the emergence of a resistance clone; moreover, in some plasma samples, multiple mutations occurred at measurable allelic frequencies, indicating the presence of parallel development of resistant clonal populations [87]. Schiavon and Turner examined ptDNA from 171 women with advanced breast cancer and uncovered ESR1 mutations exclusively in patients treated with aromatase inhibitors (AIs)—who were also found to have shorter progression-free survival [88]. Other investigations of ptDNA assays for ESR1 mutations have likewise shown utility in predicting AI resistance [89, 90].

Similar studies hold hope for validation of ptDNA assays in other resistance mutations; e.g., two separate studies have reported the use of BEAMing to detect the emergence of KRAS mutations that are known to confer resistance to antibody-mediated EGFR-targeted therapies for colorectal cancer [91, 92], and one recent trial used ptDNA to detect acquired PIK3CA mutations that may contribute to cetuximab resistance in metastatic colon cancer [93]. Siravegna et al. developed a targeted NGS panel to explore colorectal tumor evolution in response to therapies, not only identifying the emergence of potentially novel resistance mutations in EGFR but also demonstrating rise and fall of KRAS mutational burden in response to targeted treatment [94]. The relatively common V600E BRAF mutation has also been detected in ptDNA in melanoma patients and showed promise in monitoring response to BRAF-directed therapy [95].

Trials like these, as well as the emerging acceptance of ptDNA assays for resistance mutations, increasingly bring into focus the prospect of anticipatory therapy change, wherein a clinician substitutes a new treatment when the resistance mutation is detected, before clinical or radiological progression. Whether this would actually extend or improve the lives of patients with metastatic cancer remains an open question; it is a testable hypothesis currently being addressed in clinical trials.

ptDNA in Cancer Screening

A "liquid biopsy" for cancer may conjure a vision where simple blood screening tests can take the place of painful or invasive procedures such as mammography, colonoscopy, and diagnostic biopsy. However, the sparsity of DNA in early detectable disease, as well as the lack of tissue specificity for the most common cancer mutations, likely relegates these hopes to the future. Moreover, knowing that an individual has a mutation in ptDNA is not an absolute indication that this person has a malignancy. Finally, for a screening test to be clinically useful, one must demonstrate that the test reduces overall mortality from the cancer and preferably mortality in general. Such studies will take decades to prove or disprove, though several efforts continue to explore the use of ptDNA for primary screening. For example, Cohen et al. recently demonstrated a 64% sensitivity for early-stage, surgically resectable pancreatic cancer when combining KRAS detection in ptDNA with four protein biomarkers (CEA, CA19–9, HGF, and OPN) [96].

Challenges in Oncologic Applications

Despite the promise that ptDNA holds in a variety of oncologic settings outlined above, the field faces numerous challenges and complexities.

Many studies published to date are retrospective, using plasma gathered as part of clinical trials with a different primary focus. While this research is illuminating, it should be viewed as hypothesis-generating; as always, large, multicenter prospective clinical trials will provide the evidence of real-world clinical benefit. A few such trials are underway; for example, one large French study aims to determine whether patients with ESR1 mutations in ptDNA would benefit from an early switch from AI/palbociclib to fulvestrant/palbociclib (clinicaltrial.gov #NCT03079011).

While the FDA has approved one ptDNA assay as described above, overall the field lacks standardization. Publications describe a wide variety of methods to obtain and process plasma, which, as discussed, may explain some of the disparities in results reported. Moreover, in tissue vs. ptDNA studies, even ground truth has yet to be established, given that ptDNA could reasonably be expected to contain mutations not present in tissue; indeed, at least one rapid autopsy study has addressed this problem by comparing total postmortem metastatic mutational burden to ptDNA present in blood at the time of death [97].

Meanwhile, a host of companies and platforms compete for research dollars in ptDNA processing, all with varying proprietary methods; many are excellent, but again, none are uniformly received as a standard. As methods for detecting and sequencing genetic material encompass an increasing number of platforms and techniques, it will be important to understand strengths and pitfalls of available technology. Moreover, NGS assays must become increasingly nimble to accommodate the need for rapid, inexpensive serial monitoring of ptDNA if clinicians are to respond quickly to changes in mutational burden. Sensitivity must also be improved if we hope to move beyond our dependence on tissue for mutation discovery and advance to direct mutation detection from ptDNA alone.

One emerging issue complicating the study of ptDNA involves the blood dyscrasias that develop in many older patients. When sequencing genetic material from either tumor or plasma, a substantial proportion of DNA can originate from healthy tissue or from white blood cells mixed with plasma or tissue samples; mutations, therefore, are assumed to originate from the tumor under investigation—perhaps in some cases erroneously. A recent case study relates an account of a patient whose sequenced solid tumor revealed a JAK2 mutation, which was initially assumed to represent an unusual but actionable mutation from the solid tumor in question; further inquiry revealed the patient's known history of polycythemia vera. From this, it was further demonstrated that the tumor sample contained enough blood to display a detectable allelic fraction of mutant JAK2 [98].

In the above case, investigators were able to discern the source of the patient's JAK2 mutation because the mutation is associated with polycythemia vera, and the patient had been diagnosed with the disease. But an expanding set of literature describes somatic mutations of unknown significance developing in the blood of older persons [99]. This so-called clonal hematopoiesis of indeterminate potential, or CHIP, threatens to further complicate the task of distinguishing actual solid tumor-related mutations from these poorly characterized hematologic mutations.

This issue is arising in both plasma and tissue samples. The more that a tumor sample exhibits a hemorrhagic component, the more one risks CHIP mutations entering the mix if they are present. Regarding plasma, although double spun plasma can remove the majority of cellular contaminants, the very nature of CHIP suggests that mutations from these cells are also found in ccfDNA, again complicating the mutational profile of ptDNA leading to false-positive mutations. As NGS technologies increase their sensitivity, ever smaller mutant allele fractions may achieve detectability, further compounding the complexity of ptDNA analysis. With awareness and careful analysis of mutational profiles, confusion regarding CHIP mutations can be minimized; other technologies if validated, such as methylation haplotyping, could theoretically help distinguish origin tissue type further mitigating these concerns [100].

Future Directions for ccfDNA

Looking ahead, more potential applications of circulating genetic material appear on the horizon. For example, researchers are exploring the use of alternative circulating nucleic acids to detect residual disease and profile mutations in cancer and other disease states. Specifically, messenger RNA (mRNA) also appears to circulate in human serum, although its cellular origin is less clear than that of ccfDNA. Circulating cell-free mRNA can be detected using microarray technologies or reverse transcription qPCR [101], with various potential applications. For example, circulating mRNA may predict graft rejection in transplant patients using fractions of donor-specific ccfDNA and circulating mRNA [102]. It may also have applications in early diagnosis of diabetic retinopathy and neuropathy: by screening for organ-specific mRNA in the plasma, investigators envision identifying these diabetic complications sooner, with the hope that early intervention could improve outcomes [5, 103]. Finally, microRNAs (miRNAs)—noncoding RNA species that regulate gene expression—have been described in the serum of cancer patients with B cell lymphoma in several large studies, and serum miRNA levels have also been shown to correlate with solid tumor metastases [104, 105].

The ratio of long to short DNA fragments (DNA integrity) is also under investigation as a possible biomarker of tumor presence and tumor burden, which could be broadly applicable for many cancer subtypes and could also improve sensitivities of current assays [106]. In addition, studies examining epigenetic alterations in the plasma of cancer patients, specifically detection of promoter hypermethylation by methylation-specific PCR, have been performed in various cancer subtypes and hold significant promise as another biomarker of cancer burden [107–109].

In addition to cancer and maternal–fetal medicine, other potential uses for ccfDNA in medical diagnostics are under exploration. For example, circulating bacterial DNA fragments can diagnose a causative bacterial organism in culture-negative but clinically septic patients [110]. Cell-free DNA may be helpful in risk-stratifying trauma and burn patients, whose levels of ccfDNA may correlate with severity of injury, outcomes, and length of hospital stay [111–113]. Investigators have looked into whether ccfDNA levels can predict sepsis [114, 115] and whether increasing levels correspond with worsening myocardial damage and/or cardiac outcomes in acute coronary syndromes [116]. Others have explored correlations between classic cardiac ischemic markers such as troponin and creatinine kinase with increasing levels of ccfDNA [3] and the prognostic value of ccfDNA in stroke patients [117]. No doubt as the field matures, more applications in various areas of clinical medicine will emerge.

Conclusions

Researchers in the still-nascent ccfDNA field have made huge strides in developing ccfDNA applications in cancer and other disease states, yet far more remains to be explored. Like other areas of human research, we face the exciting but challenging task of interfacing with the private sector, governmental regulatory bodies, academic institutions, and others, to understand and select rapidly developing technology and platforms. We struggle with the technical problems of detecting a small percentage of variant molecules in a sea of wild-type DNA, all in a setting where currently neither an industry standard nor a uniformly agreed-upon platform exists. As tests increasingly require both high complexity and fast turnaround to be clinically actionable, we will need to increase speed and efficiency. Finally, in order to validate these promising assays, we must prove not only their feasibility but also their clinical utility in the real world. In short, we must work to establish the best techniques to quantify, detect, and monitor ccfDNA and develop appropriate criteria for ccfDNA surveillance through prospective clinical trials. Through standardization and improved detection, this emerging technology promises to produce rapid, noninvasive, sensitive assays that will allow clinicians and patients to make better decisions and improve clinical outcomes.

References

1. Mandel P, Metais P. Not AvailableC R Seances Soc Biol Fil. 1948;142(3–4):241–3.
2. Lam NY, et al. Plasma DNA as a prognostic marker for stroke patients with negative neuroimaging within the first 24 h of symptom onset. Resuscitation. 2006;68(1):71–8.
3. Antonatos D, et al. Cell-free DNA levels as a prognostic marker in acute myocardial infarction. Ann N Y Acad Sci. 2006;1075:278–81.
4. Saukkonen K, et al. Association of cell-free plasma DNA with hospital mortality and organ dysfunction in intensive care unit patients. Intensive Care Med. 2007;33(9):1624–7.
5. Sandhu HS, et al. Measurement of circulating neuron-specific enolase mRNA in diabetes mellitus. Ann N Y Acad Sci. 2008;1137:258–63.
6. Choi JJ, Reich CF 3rd, Pisetsky DS. The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells. Immunology. 2005;115(1):55–62.
7. Stroun M, et al. The origin and mechanism of circulating DNA. Ann N Y Acad Sci. 2000;906:161–8.
8. Jahr S, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659–65.
9. Diehl F, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A. 2005;102(45):16368–73.
10. Diehl F, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14(9):985–90.
11. De Mattos-Arruda L, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun. 2015;6:8839.

12. Pan W, et al. Brain tumor mutations detected in cerebral spinal fluid. Clin Chem. 2015;61(3):514–22.
13. Lo YM, et al. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet. 1999;64(1):218–24.
14. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer--a survey. Biochim Biophys Acta. 2007;1775(1):181–232.
15. Emlen W, Mannik M. Effect of DNA size and strandedness on the in vivo clearance and organ localization of DNA. Clin Exp Immunol. 1984;56(1):185–92.
16. Chang CP, et al. Elevated cell-free serum DNA detected in patients with myocardial infarction. Clin Chim Acta. 2003;327(1–2):95–101.
17. Wimberger P, et al. Impact of platinum-based chemotherapy on circulating nucleic acid levels, protease activities in blood and disseminated tumor cells in bone marrow of ovarian cancer patients. Int J Cancer. 2011;128(11):2572–80.
18. Lo YM, et al. Plasma DNA as a prognostic marker in trauma patients. Clin Chem. 2000;46(3):319–23.
19. Chiu TW, et al. Plasma cell-free DNA as an indicator of severity of injury in burn patients. Clin Chem Lab Med. 2006;44(1):13–7.
20. Rhodes A, et al. Plasma DNA concentration as a predictor of mortality and sepsis in critically ill patients. Crit Care. 2006;10(2):R60.
21. Herzenberg LA, et al. Fetal cells in the blood of pregnant women: detection and enrichment by fluorescence-activated cell sorting. Proc Natl Acad Sci U S A. 1979;76(3):1453–5.
22. Lo YM, et al. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350(9076):485–7.
23. Lo YM, Chiu RW. Prenatal diagnosis: progress through plasma nucleic acids. Nat Rev Genet. 2007;8(1):71–7.
24. Li Y, et al. Cell-free DNA in maternal plasma: is it all a question of size? Ann N Y Acad Sci. 2006;1075:81–7.
25. Lo YM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet. 1998;62(4):768–75.
26. Lun FM, et al. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem. 2008;54(10):1664–72.
27. Chiu RW, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011;342:c7401.
28. Lo YM. Fetal RhD genotyping from maternal plasma. Ann Med. 1999;31(5):308–12.
29. Fan HC, et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A. 2008;105(42):16266–71.
30. Chiu RW, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A. 2008;105(51):20458–63.
31. Chiu RW, Lo YM. Clinical applications of maternal plasma fetal DNA analysis: translating the fruits of 15 years of research. Clin Chem Lab Med. 2013;51(1):197–204.
32. Liao GJ, Gronowski AM, Zhao Z. Non-invasive prenatal testing using cell-free fetal DNA in maternal circulation. Clin Chim Acta. 2014;428:44–50.
33. (ACOG), A.C.o.O.a.G., Cell-free DNA Screening for Fetal Aneuploidy. Society for Maternal-Fetal Medicine, Committee on Genetics. 2015.
34. Lo YM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2(61):61ra91.
35. Leon SA, et al. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37(3):646–50.
36. Allen D, et al. Role of cell-free plasma DNA as a diagnostic marker for prostate cancer. Ann N Y Acad Sci. 2004;1022:76–80.
37. Chun FK, et al. Circulating tumour-associated plasma DNA represents an independent and informative predictor of prostate cancer. BJU Int. 2006;98(3):544–8.
38. Schwarzenbach H, et al. Detection and monitoring of cell-free DNA in blood of patients with colorectal cancer. Ann N Y Acad Sci. 2008;1137:190–6.
39. Li BT, et al. A prospective study of total plasma cell-free DNA as a predictive biomarker for response to systemic therapy in patients with advanced non-small-cell lung cancers. Ann Oncol. 2016;27(1):154–9.
40. Giacona MB, et al. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. Pancreas. 1998;17(1):89–97.
41. Chen X, et al. Detecting tumor-related alterations in plasma or serum DNA of patients diagnosed with breast cancer. Clin Cancer Res. 1999;5(9):2297–303.
42. Heid CA, et al. Real time quantitative PCR. Genome Res. 1996;6(10):986–94.
43. Branford S. Chronic myeloid leukemia: molecular monitoring in clinical practice. Hematology Am Soc Hematol Educ Program. 2007:376–83.
44. Yu M, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;339(6119):580–4.
45. Morley AA. Digital PCR: a brief history. Biomol Detect Quantif. 2014;1(1):1–2.
46. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999;96(16):9236–41.
47. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
48. Forshew T, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
49. Kinde I, et al. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A. 2011;108(23):9530–5.
50. Kennedy SR, et al. Detecting ultralow-frequency mutations by duplex sequencing. Nat Protoc. 2014;9(11):2586–606.
51. Leary RJ, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med. 2012;4(162):162ra154.
52. Castells A, et al. K-ras mutations in DNA extracted from the plasma of patients with pancreatic carcinoma: diagnostic utility and prognostic significance. J Clin Oncol. 1999;17(2):578–84.
53. Kopreski MS, et al. Somatic mutation screening: identification of individuals harboring K-ras mutations with the use of plasma DNA. J Natl Cancer Inst. 2000;92(11):918–23.
54. Dianxu F, et al. A prospective study of detection of pancreatic carcinoma by combined plasma K-ras mutations and serum CA19-9 analysis. Pancreas. 2002;25(4):336–41.
55. Garcia JM, et al. Extracellular tumor DNA in plasma and overall survival in breast cancer patients. Genes Chromosomes Cancer. 2006;45(7):692–701.
56. Boddy JL, et al. Prospective study of quantitation of plasma DNA levels in the diagnosis of malignant versus benign prostate disease. Clin Cancer Res. 2005;11(4):1394–9.
57. Schwarzenbach H, et al. Comparative evaluation of cell-free tumor DNA in blood and disseminated tumor cells in bone marrow of patients with primary breast cancer. Breast Cancer Res. 2009;11(5):R71.
58. Schwarzenbach H, et al. Cell-free tumor DNA in blood plasma as a marker for circulating tumor cells in prostate cancer. Clin Cancer Res. 2009;15(3):1032–8.
59. Toro PV, et al. Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. Clin Biochem. 2015;48:993.

60. Bettegowda C, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
61. Hadano N, et al. Prognostic value of circulating tumour DNA in patients undergoing curative resection for pancreatic cancer. Br J Cancer. 2016;115(1):59–65.
62. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
63. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;467(7319):1114–7.
64. Higgins MJ, et al. Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. Clin Cancer Res. 2012;18(12):3462–9.
65. De Mattos-Arruda L, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Ann Oncol. 2014;25(9):1729–35.
66. Rothé F, Laes J-F, Lambrechts D, Smeets D, Vincent D, Maetens M, Fumagalli D, Michiels S, Stylianos D, Moerman C, Detiffe J-P, Larsimont D, Awada A, Piccart M, Sotiriou C, Ignatiadis M. Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. Ann Oncol. 2014;25:1959.
67. Parsons HA, et al. Individualized molecular analyses guide efforts (IMAGE): a prospective study of molecular profiling of tissue and blood in metastatic triple negative breast cancer. Clin Cancer Res. 2017;23(2):379–86.
68. Tomasetti C, et al. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci U S A. 2015;112(1):118–23.
69. Ryan BM, et al. A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up. Gut. 2003;52(1):101–8.
70. Wang S, et al. Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer. Clin Cancer Res. 2010;16(4):1324–30.
71. Diehl F, et al. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. Gastroenterology. 2008;135(2):489–98.
72. Husain H, et al. Monitoring daily dynamics of early tumor response to targeted therapy by detecting circulating tumor DNA in urine. Clin Cancer Res. 2017;23:4716.
73. Oshiro C, et al. PIK3CA mutations in serum DNA are predictive of recurrence in primary breast cancer patients. Breast Cancer Res Treat. 2015;150(2):299–307.
74. Olsson E, et al. Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. EMBO Mol Med. 2015;**7**(8):1034–47.
75. Pietrasz D, et al. Plasma circulating tumor DNA in pancreatic cancer patients is a prognostic marker. Clin Cancer Res. 2017;23(1):116–23.
76. Tie J, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. Ann Oncol. 2015;26:1715.
77. Garcia-Murillas I, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015;7(302):302ra133.
78. Lebofsky R, et al. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. Mol Oncol. 2015;9(4):783–90.
79. Taniguchi K, et al. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. Clin Cancer Res. 2011;17(24):7808–15.
80. Piotrowska Z, et al. Heterogeneity underlies the emergence of EGFRT790 wild-type clones following treatment of T790M-positive cancers with a third-generation EGFR inhibitor. Cancer Discov. 2015;5(7):713–22.
81. Ishii H, et al. Digital PCR analysis of plasma cell-free DNA for non-invasive detection of drug resistance mechanisms in EGFR mutant NSCLC: correlation with paired tumor samples. Oncotarget. 2015;6(31):30850–8.
82. Que D, et al. EGFR mutation status in plasma and tumor tissues in non-small cell lung cancer serves as a predictor of response to EGFR-TKI treatment. Cancer Biol Ther. 2016;17(3):320–7.
83. Oxnard GR, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin Cancer Res. 2014;20(6):1698–705.
84. Chabon JJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun. 2016;7:11815.
85. Rosell R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012;13(3):239–46.
86. FDA, Editor. 2016. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm504488.htm
87. Chu D, et al. ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients. Clin Cancer Res. 2016;22(4):993–9.
88. Schiavon G, et al. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. Sci Transl Med. 2015;7(313):313ra182.
89. Fribbens C, et al. Plasma ESR1 mutations and the treatment of estrogen receptor-positive advanced breast cancer. J Clin Oncol. 2016;34:2961.
90. Wang P, et al. Sensitive detection of mono- and polyclonal ESR1 mutations in primary tumors, metastatic lesions and cell free DNA of breast cancer patients. Clin Cancer Res. 2016;22(5):1130–7.
91. Diaz LA Jr, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature. 2012;486(7404):537–40.
92. Misale S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012;486(7404):532–6.
93. Xu JM, et al. PIK3CA mutations contribute to acquired cetuximab resistance in patients with metastatic colorectal cancer. Clin Cancer Res. 2017;23:4602.
94. Siravegna G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med. 2015;21(7):795–801.
95. Shinozaki M, et al. Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy. Clin Cancer Res. 2007;13(7):2068–74.
96. Cohen JD, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci U S A. 2017;114(38):10202–7.
97. Desmedt C, Brown DN, Szekely B, Smeets D, Szasz MA, Adnet P, Rothé F, Nagy Z, Farago Z, Tokes A, Zardavas D, Zoppoli G, Ignatiadis M, Pusztai L, Piccart M, Larsimont D, Lambrechts D, Kulka J, Sotiriou C. Unraveling breast cancer progression through geographical and temporal sequencing, in AACR 2014. San Diego; 2014.
98. Lee J, et al. A polycythemia vera JAK2 mutation masquerading as a duodenal cancer mutation. J Natl Compr Cancer Netw. 2016;14(12):1495–8.

99. Steensma DP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015;126(1):9–16.
100. Guo S, et al. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet. 2017;49(4):635–42.
101. O'Driscoll L, et al. Feasibility and relevance of global expression profiling of gene transcripts in serum from breast cancer patients using whole genome microarrays and quantitative RT-PCR. Cancer Genomics Proteomics. 2008;5(2):94–104.
102. Schutz E, et al. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: a prospective, observational, multicenter cohort study. PLoS Med. 2017;14(4):e1002286.
103. Simo-Servat O, Simo R, Hernandez C. Circulating biomarkers of diabetic retinopathy: an overview based on physiopathology. J Diabetes Res. 2016;2016:5263798.
104. Bradshaw G, et al. Dysregulated MicroRNA expression profiles and potential cellular, circulating and polymorphic biomarkers in Non-Hodgkin Lymphoma. Genes (Basel). 2016;7(12):130.
105. Sapp RM, et al. Circulating microRNAs in acute and chronic exercise: more than mere biomarkers. J Appl Physiol (1985). 2017;122(3):702–17.
106. Zonta E, Nizard P, Taly V. Assessment of DNA integrity, applications for cancer research. Adv Clin Chem. 2015;70:197–246.
107. Visvanathan K, et al. Monitoring of serum DNA methylation as an early independent marker of response and survival in metastatic breast cancer: TBCRC 005 prospective biomarker study. J Clin Oncol. 2017;35(7):751–8.
108. Tang Y, et al. Promoter DNA methylation analysis reveals a combined diagnosis of CpG-based biomarker for prostate cancer. Oncotarget. 2017;8(35):58199–209.
109. Hagrass HA, Pasha HF, Ali AM. Estrogen receptor alpha (ERalpha) promoter methylation status in tumor and serum DNA in Egyptian breast cancer patients. Gene. 2014;552(1):81–6.
110. Grumaz S, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. Genome Med. 2016;8(1):73.
111. Shoham Y, et al. Admission cell free DNA as a prognostic factor in burns: quantification by use of a direct rapid fluorometric technique. Biomed Res Int. 2014;2014:306580.
112. Lam NY, et al. Time course of early and late changes in plasma DNA in trauma patients. Clin Chem. 2003;49(8):1286–91.
113. Hu Q, et al. Elevated levels of plasma mitochondrial DNA are associated with clinical outcome in intra-abdominal infections caused by severe trauma. Surg Infect. 2017;18:610.
114. Long Y, et al. Diagnosis of sepsis with cell-free DNA by next-generation sequencing technology in ICU patients. Arch Med Res. 2016;47(5):365–71.
115. Hou YQ, et al. Branched DNA-based Alu quantitative assay for cell-free plasma DNA levels in patients with sepsis or systemic inflammatory response syndrome. J Crit Care. 2016;31(1):90–5.
116. Rainer TH, et al. Plasma beta-globin DNA as a prognostic marker in chest pain patients. Clin Chim Acta. 2006;368(1–2):110–3.
117. O'Connell GC, et al. Circulating extracellular DNA levels are acutely elevated in ischaemic stroke and associated with innate immune system activation. Brain Inj. 2017:1–7.

Part II

Practice-Related Aspects of Clinical Genomics

8 Genomic Pathology: Training for New Technology

Richard L. Haspel

Introduction

Sequencing of the first human genome took over 10 years and cost more than $2 billion [1]. Today, physicians routinely order testing utilizing massively parallel next-generation sequencing (NGS) methods [2]. Pathologists, as the directors of clinical laboratories, have the expertise to effectively translate genomic technology to patient care. To play this important role, pathologists must be trained in genomic methods and result interpretation. This chapter provides evidence demonstrating the need for genomic pathology education, addresses the progress to date of several educational initiatives, and suggests possible ways to improve future training.

Much of molecular pathology involves testing for single-gene variants (e.g., *BRCA*). For the purpose of this chapter, "genomics" refers to analysis of large portions of the genome with a single "test." Aside from the whole genome, only gene-coding regions (exome) or expressed genes (transcriptome) can be sequenced. Chip-based testing, as well as other approaches, can be utilized in the analysis of hundreds of genes, millions of single nucleotide polymorphisms (SNPs), or copy number variation across the genome.

Integration of Genomic Testing into Clinical Care

Genomic testing using NGS technology is being incorporated into almost all areas of medicine. In oncology, genomic analysis of tumors has paved the way for personalized therapeutic approaches. For example, as early as 2007, a patient was diagnosed with an oral adenocarcinoma [3] and developed lung metastasis despite excision and adjuvant radiation. Although treated with an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (based on increased EGFR immunohistochemical (IHC) staining), the tumor continued to grow. Lacking additional chemotherapeutic options, whole-genome and whole-transcriptome sequencing were performed on a lung biopsy specimen. This analysis demonstrated upregulation of the *RET* oncogene, and this finding was confirmed with fluorescence in situ hybridization (FISH) analysis as well as IHC staining. Subsequent treatment with a RET tyrosine kinase inhibitor led to stabilization of disease for 4 months. Upon disease progression, sequencing of a new biopsy specimen revealed mutations that could bypass the RET inhibition.

In a second illustrative oncology case, a patient developed what appeared, by morphology, to be acute promyelocytic leukemia (APML) [4]. The corresponding *PML-RAR* rearrangement, however, could not be detected using a standard FISH assay. Using NGS methods, a cytogenetically cryptic *PML-RAR* fusion was identified, and this result was subsequently confirmed by FISH and polymerase chain reaction (PCR) assays. Determination of the genetic basis of the disease led to appropriate treatment with all-trans-retinoic acid. The entire diagnostic process was completed in approximately 7 weeks.

Although the above case reports are informative, there is now data on the utility of genomic oncology testing beyond single patient examples. In a 2015 study of 102 children with refractory cancer, whole-exome and whole-transcriptome analysis led to treatment modification in 14 patients with 9 having ongoing partial or complete clinical remissions [5]. In a trial involving 1640 patients, molecular profiling of tumor samples led to 89 being enrolled in genotype-guided clinical trials [6]. The response rate of those on genotype-matched trials was higher than those on non-genotype-matched trials (19% vs. 9%). In addition to large-scale sequencing, gene panels are becoming more commonly used in a variety of cancers. In a 2014 study of 1007 patients with lung cancer, use of a 10-gene panel of oncogenic drivers allowed 28% to be treated with targeted therapy [7]. These

R. L. Haspel (✉)
Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA
e-mail: rhaspel@bidmc.harvard.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_8

patients survived a median of 3.5 years compared to 2.4 years for those who did not receive targeted treatment. For breast cancer, both a 21-gene and a 70-gene assay performed on tumor samples are commercially available to provide information regarding risk of recurrence and possible need for chemotherapy [8]. Gene panels have also been developed to help determine appropriate management of cytologically indeterminate thyroid nodules [9]. A 13-gene panel of oncogenes has been used to guide pharmacologic management of cancer patients. In a prospective study of salivary duct carcinoma cases, the assay influenced treatment decisions in six of eight patients tested [10].

Genomic technology has also been applied to nonneoplastic diseases. A 15-month-old presented with intractable inflammatory bowel disease (IBD) requiring multiple surgical interventions [11]. Whole-exome sequencing led to the discovery of a variant in the X-linked inhibitor of apoptosis (XIAP) gene. Deficiency of this gene can lead to hemophagocytic histolymphiocytosis (HLH). Although not typically associated with colitis, given the prognostic implications of HLH, the patient underwent a bone marrow transplant, and the procedure appears to have cured the patient's IBD. Genomic testing has also revealed the genetic cause of other rare diseases. In a study involving 250 probands, whole-exome sequencing revealed a molecular cause in 25% of cases [12]. Results may not always have an immediate direct effect on patient care but can lead to insights into the disease pathophysiology. Demonstrating the rapidly changing testing landscape, a quarter of the patients' gene-disease correlations were based on discoveries made in the 2 years prior to the study.

Genomics applications are not limited to disease state and are increasingly being offered to healthy individuals. As an example, for women planning a pregnancy, there is a commercially available single test assessing risk for over 100 genetic diseases [13]. An NGS-based test to detect fetal trisomies 13, 18, and 21 in maternal peripheral blood is also commercially available [14]. Using a sample of the mother's blood, cell-free DNA is isolated and sequenced, and the amount of representation from each chromosome is quantified. In such testing, an excess of chromosome 21 DNA is consistent with Down syndrome in the fetus.

Other areas of pathology have been impacted by genomic testing. In transfusion medicine, high-throughput assays have been developed to determine the blood group antigen genotypes of healthy donors [15]. This genotyping will allow better donor-recipient matching and identification of donors with rare variants for which classic serologic methods are of limited utility. NGS methods are also being applied to microbiologic testing. During the 2011 *E. coli* outbreak in Europe, the entire sequence of the causative organism was determined in less than a week [16]. During a tuberculosis outbreak in Canada, whole-genome sequencing of 32 isolates led to the determination of the outbreak epidemiology when traditional methods failed [17]. In regard to direct patient care, when standard testing did not identify a cause, a child with chronic meningitis was diagnosed with leptospirosis based on NGS of cerebrospinal fluid [18]. In a proof-of-principal study, all samples sent to an academic microbiology laboratory in a single day underwent genomic sequencing. Sequencing and analysis took between 2 and 3 days. They were able to identify the organisms by sequencing for 115 of 130 samples (88.5%) [19].

Are Physicians Prepared for the Genomic Era?

The application of genomic testing to patient care will only continue to increase. As such, physicians must be prepared to understand appropriate ordering practice and the interpretation of these new assays. Unfortunately, there is evidence that many physicians do not understand single-gene molecular testing, let alone genomic analysis.

A study from 1997 examined physician practice in testing for the *APC* gene variant associated with familial adenomatous polyposis [20]. In approximately 20% of cases, an inappropriate strategy for pre-symptomatic testing was used, and 32% of the results were misinterpreted by ordering physicians. Two decades later, physician's ability to appropriately utilize and interpret genetic tests does not appear to have significantly improved [21–25]. A number of studies have examined genetic counselor screening of molecular laboratory testing requests with review leading to revision of a remarkable 8–26% of orders [21, 22]. In many cases, the tests are cancelled demonstrating a lack of physician comprehension leading to a waste of resources and the potential for patient worry due to variants of uncertain significance. Physicians are aware of their need for acquiring additional knowledge of genetic testing. In one survey-based study of over 200 internists, while 65% stated that they have counseled a patient on a genetic issue and 44% had ordered a genetic test in the past 6 months, 74% rated their knowledge of genetics as "somewhat poor" or "very poor," and approximately 80% indicated a need for additional training [23]. Similar results have been found for neurologists and psychiatrists [24]. In another study involving 401 family physicians, 55% reported that they had no awareness of the Genetic Information Nondiscrimination Act (GINA) [25].

Current Physician Training in Genomics

Clearly, many of today's practicing physicians have difficulty interpreting single-gene testing and are not prepared for the genomic era. It would also seem logical that such training should begin in medical school to provide a foundation for further learning. While within US and

Canadian medical schools there is a trend toward increasing genetics and genomics education, a recent survey-based study of genetics course directors found that 40% of respondents still do not cover key emerging genomic medicine topics in their curriculum [26]. In addition, this study found that the overwhelming majority of the material that is taught is presented in the first 2 years (preclinical). Such an approach bypasses the time in medical students' education in which they would take key genomic concepts, such as targeted cancer therapies, ordering genomic tests and explaining test results to patients, from theory to clinical practice. As almost all physicians will be involved to some degree in patient genetic and genomic testing, medical training must include practical education in this area.

Others have recognized this gap in medical education, and there have been numerous publications calling for greater training in genomic medicine [27–30]. There have also been several single site examples of implementation and evaluation of curricula. These include student self-testing, testing of cadavers during a gross anatomy course, and incorporating genomics as a "horizontal" strand during a longitudinal curriculum for the first and second years of medical school [31–35].

There are also several national efforts to provide resources for genomics education. In the United States, the Intersociety Coordinating Committee for Practitioner Education in Genomics (ISCC), a group under the auspices of the National Human Genome Research Institute (NHGRI) with representation from medical professional societies, has published a list of competencies for physicians [27]. G2C2, also under the NHGRI, provides a clearinghouse of educational resources for genetics and genomics [36]. Resources are submitted by groups or individuals, vetted by an editorial board, and are also mapped to ISCC physician competencies and/or competencies for other health professionals. The Association of Professors of Human and Medical Genetics (APHMG) has also developed a list of genetics and genomics competencies for a medical student curriculum [37]. In the UK, there are ongoing national efforts through the National Genetics and Genomics Education Centre and Health Education England (HEE) Genomics Education Programme to make resources available such as e-learning modules and curricula [30]. Taken together, however, while these initiatives in the USA and UK to develop competencies and resources are important, these curricula and teaching tools have not been broadly implemented and rigorously evaluated on a national scale.

The Important Role of Pathologists in the Genomic Era

Pathologists are in a unique position to assist in translating genomic technology to clinical care. In many large medical centers, pathologists with subspecialty expertise and certification direct laboratories offering single-gene and gene panel testing and have the expertise in ensuring accurate and precise results. In addition, almost all specimens used for genomic testing will pass through the pathology laboratory.

In anatomic pathology, a pathologist must first determine that there is a malignant process before sending for assays that determine prognosis or potential targeted therapies for a given neoplastic disease. The pathologist must also ensure that an appropriate sample is sent. Determining the type of processing (fresh versus frozen or formalin-fixed) and the portion of the specimen to analyze is crucial in providing accurate results [38]. In a 2011 study, in-house HER2 breast cancer testing was compared to the results from a commercial prognostic gene panel assay [39]. Such testing has important implications for treatment with specific HER2-targeted therapy. Of the 843 cases, there were 14 in which the in-house positive result conflicted with a negative gene panel result. The authors hypothesized that the issue may have been the selection of tissue blocks with limited tumor. That is, in these instances, the company may have been primarily testing normal tissue. In addition to sample selection, pathologists also assist other clinicians through integration of genomic findings into histology reports and providing important insight at genomic tumor boards [40]. In clinical pathology, whether in the blood bank, microbiology, hematology, or molecular pathology laboratories, pathologists have access to samples for genomic analysis. Furthermore, pathologists are already versed in incorporating genetic data into pathology reports that enable other clinicians to understand the results and act appropriately.

Given the experience and training of pathologists in sample preparation, assay validation, and quality control, one can argue that without pathologists overseeing genomic testing, there is the potential for patient harm. As evidence of such potential danger, in 2009, a direct to consumer (DTC) genomic testing company mixed up samples, leading to clients receiving incorrect results suggesting risk for a variety of diseases [41]. Illustrating the importance of standardizing testing, in a study where identical samples from five individuals were sent to two different DTC genomic testing companies, the results were discordant 33% of the time (e.g., an individual received a report of an increased risk for a disease from one company and average or decreased risk for the same disease from the other) [42]. Pathologists, along with geneticists, also manage the laboratories that perform clinical genomic testing and have played a role in stopping unnecessary genetic testing at academic medical centers [43]. There is also a need to determine whether genomic testing is better than traditional techniques. As an example, a commercial 21-gene expression panel was compared to standard immunohistochemical testing (HER2, ER, and PR) in regard to predicting disease recurrence [44]. The standard technique performed as well as the expensive genomic

testing and had a shorter turnaround time. Pathologists are uniquely positioned to evaluate the utility of novel testing with the idea that a genetic approach is not always an improvement.

Pathologists are also well versed in many of the statistical issues that arise in the setting of genomic testing. For example, a test with very high specificity may still have a low positive predictive value (PPV) if the prevalence of the disease in the population tested is very low. An example of this issue relates to genomic testing for trisomy 21 through NGS methods. While high sensitivity and specificity were reported for these new assays, concern was raised regarding the actual PPV in a more general as opposed to high-risk population [45]. Subsequent studies did show better results with the NGS method over traditional screening [46, 47]. Another statistical issue, as most genomic assays are made up of many individual "tests" (e.g., a multigene panel), is the increased risk of false-positive results [48]. In the genomic era, pathologists' familiarity with issues related to statistics, accuracy, precision, result reporting, and quality control is vitally important.

Given the above, pathologists need to be centrally involved in translating genomic methods to patient care. As genomic testing will affect all areas of medicine, however, pathologists will need to collaborate with other specialists such as genetic counselors and medical geneticists. In a pilot study, pathologists developed a workflow for tumor analysis consisting of sample processing, sequencing, and result validation [40]. They also described the formation of a "genomic" or "molecular" tumor board consisting of oncologists, medical geneticists, ethicists, as well as pathologists. Similar approaches are being implemented at other institutions.

It is important to note, however, that there are currently approximately 6000 medical geneticists and genetic counselors in the USA [49, 50]. In contrast, there are approximately 18,000 board-certified pathologists [51]. As such, pathologists have not only the expertise but also the workforce needed to translate genomic testing to patient care, even if only a subset of pathologists will specialize in this area.

Single-Program Approaches to Genomic Pathology Training

As can be surmised from the above discussion, there is a strong case for training pathologists in genomics. In 2010, a group of representatives from leading pathology organizations, insurance consortiums, industry, the National Institutes of Health (NIH), and the military met at the Banbury Conference Center at Cold Spring Harbor Laboratory to discuss the future of genomic pathology. Recommendations from the meeting listed seven "Blue Dot" projects to help ensure that pathologists play a significant role in applying genomic technology to patient care [52]. One of these projects (Blue Dot project #1) had the goal "to ensure that every Accreditation Council for Graduate Medical Education (ACGME)-approved residency in pathology in North America includes a mandatory curriculum in genomics and personalized medicine." An editorial published in the same issue of the *American Journal of Clinical Pathology* as the Banbury Conference recommendations stated that "although all seven projects certainly have merit and are important to pathologists … project 1 is, without doubt, a 'no-brainer'" and "the need to introduce NGS and whole-genome technology topics into medical student and pathology resident education is mandatory" [53].

Given current medical school training, individuals entering pathology residency would be expected to have a limited background in genetics and genomics. The Accreditation Council of Graduate Medical Education (ACGME) Pathology Program Requirements simply mandate residency training in "molecular pathology" [54]. Milestones, developed by the ACGME to provide some additional context to core requirements, specify training in "personalized medicine" and "precision diagnostics" but do not provide specific objectives or curricula [55].

In 2016, to provide a framework for pathology resident education in molecular genetics and genomics, the Association for Molecular Pathology released "A Suggested Molecular Pathology Curriculum for Residents" [56]. This document provides specific objectives related to all aspects of molecular pathology including genomic technology and informatics. The objectives are divided into "prerequisites," "required," and "recommended." In 2012, a College of American Pathologists (CAP) work group published a list of 32 high-priority genomics competencies for practicing pathologists (Table 8.1) [57]. The list was developed using a structured approach utilizing a modified Delphi survey method. While the above documents are valuable resources for guiding content for pathologist education, they do not provide tools for implementation of a comprehensive curriculum.

Prior to the development of the AMP curriculum and CAP competencies and in the absence of available teaching tools, several pathology residency programs established genomic pathology curricula, and two programs published their approach. In 2009, the faculty at Beth Israel Deaconess Medical Center (BIDMC) developed a mandatory resident curriculum in genomic pathology [58]. Knowledge-, affective-, and performance-based objectives were included [59]. First, residents attended three lectures. An introductory lecture provided an overview of genomics and the important role a pathologist is expected to play in genomic testing. The second focused on genomic testing methods such as NGS. Recognizing the need for inter-specialty collaboration, the third lecture focused on communicating genetic and

Table 8.1 CAP Genomics Working Group's highest priority competencies for genomics education[a]

Competency subject area	Competency statement
Sample acquisition	Ensure effective communication between the primary clinical team and the pathologist to facilitate appropriate specimen collection including the types, quantities, and handling requirements (including sample preservation)
	Specify the criteria for determining if cells of diagnostic interest are present in a specimen in sufficient quantities (particularly for oncology genomic testing)
Testing and interpretation	Specify clinical contexts in which genomic testing may be of diagnostic, prognostic, or reproductive planning utility
	Evaluate whether the genomic test or interpretation method ordered for a patient sample is appropriate for the clinical context of the patient
Patient management and reporting	Effectively integrate genomic results with clinical information and into other laboratory and pathology reports
	Communicate effectively with genetic counselors and referring physicians in an interdisciplinary team setting on pre- and posttest counseling
Basic genomics concepts	Define targeted gene panel, whole-genome, whole-exome, and whole-transcriptome sequencing

[a]Based on a previously published material from Laudadio et al. [57], with permission from the College of American Pathologists

genomic test results to patients and was given by BIDMC genetic counselors.

The three lectures provided a strong knowledge base for the other components of the curriculum. To demonstrate the ability to apply this knowledge, residents were asked to select a paper on a disease of their choice that used genomic methods. With a faculty advisor, the resident reviewed the paper and delivered a 15-min presentation to his/her peers describing the findings. A wide variety of conditions including both malignant (e.g., melanoma) and nonmalignant (e.g., macular degeneration) were discussed. Demonstrating the thoroughness of the literature review, in the first year of administering the genomic pathology curriculum at BIDMC, two residents found an error on a DTC company website in listing the risk for multiple sclerosis associated with a specific single nucleotide polymorphism [60].

The final component of the curriculum, offering residents free-of-charge DTC genomic testing, allowed participants to appreciate affective issues related to genomic testing (i.e., the impact testing has on patients). A company was selected that utilized SNP analysis and GWAS to determine risk for 40 conditions and also provided genetic counselors to help answer questions. The testing was completely voluntary, not required to participate in the curriculum, and results were only seen by the resident who ordered the testing. In addition, the curriculum was scheduled over several months (while residents were on other rotations), so participants were able to hear the lectures and have an adequate knowledge base before deciding on the testing. Each year testing was offered, over 70% of residents participated. In an anonymous survey, no residents felt coerced in participating, and several commented that the testing added to their understanding of genomic pathology. In addition, a key driver of adult learning is relevance and the "need to know" [61]. Several residents used their testing results to decide on the topic for their presentation. While there has been some debate on the utility and ethics of educational DNA testing, self-testing on a smaller scale is not a new concept in clinical pathology training [62, 63]. At some programs, for example, a resident may perform laboratory testing on their own blood sample. Several medical school courses have implemented genetic self-testing as a curricular component [32, 34].

The BIDMC curriculum has been published, and key components including the lectures and resident presentations were made available online [58, 60]. As described below, the curriculum was used as a starting point for the Training Residents in Genomics (TRIG) Working Group.

Pathology faculty at Stanford University have published the genomic pathology curriculum offered to pathology residents at their site. This mandatory series of ten core lectures was started in 2010 and made available online in 2012 [64]. The first lecture provides an introduction to methods for measuring and manipulating nucleic acids and includes a discussion of polymerase chain reaction and sequencing technology. The following three lectures provide a background on types of genetic variation as well as microarray and NGS methods. The subsequent five lectures cover specific clinical applications of genomics in areas including inherited disorders, solid tumors, pharmacogenomics, HLA genetics, and hematopoietic cancers. The final lecture addresses ethical, regulatory, and economic issues in genomic pathology. In 2011, the Stanford University Pathology Department also began offering an advanced genomic medicine elective for residents, faculty, and fellows who "plan to work actively with genomic data." This elective is taught in a small-group interactive environment and includes additional instruction in NGS, genetic variation, and sequence analysis.

A National Approach to Genomic Pathology Education

In 2010, a survey was distributed to members of the Pathology Residency Program Directors Section (PRODS) of the Association of Pathology Chairs (APC) in order to obtain a better assessment of current national practice in regard to

genomics education [65]. Of 185 programs surveyed, 42 (23%) responded. While 93% of programs provided training in molecular pathology, only 31% had any training in genomic pathology-related topics such as NGS and DTC genetic testing. And, whereas 91% of programs without training wanted to have a curriculum, lack of faculty expertise (52%) and time in the resident schedule (76%) were cited as major barriers. Due to these issues, 74% of programs did not plan on initiating training in the following year. Respondents rated availability of online modules as the most helpful tool in implementing a new curriculum or for improving an existing curriculum in genomic medicine.

The survey results prompted the creation of a PRODS committee to facilitate integration of genomic pathology training into residency programs. The Training Residents in Genomics (TRIG) Working Group is made up of experts in medical education and molecular genetic pathology as well as members of leading pathology organizations. The American Society for Clinical Pathology (ASCP) provides administrative support. The TRIG Working Group initially included three past presidents of the Association for Molecular Pathology (AMP), a past editor in chief of *The Journal of Molecular Diagnostics*, and the former chief of the molecular pathology section of the National Cancer Institute. Recognizing the need for collaboration across specialties, the National Society of Genetic Counselors (NSGC) and the American College of Medical Genetics and Genomics (ACMG) have also appointed representatives. Given the well-thought-out approach, the working group chose to begin by revising the BIDMC curriculum leading to the development of four lectures with notes originally posted on the Intersociety Council for Pathology Information (ICPI) website and distributed in booklet form and presented at the 2012 USCAP Annual Meeting [66]. In 2012, based on this preliminary work, the chair of the TRIG Working Group was awarded an R25 grant from the National Cancer Institute (NCI) which allowed the development of additional effective tools for pathology resident genomics education.

Subsequent goals of TRIG included (1) further development of the TRIG curriculum to develop workshops at national pathology meetings, (2) an instructor handbook and tool kit to allow individual residency programs to implement their own workshops, and, (3) for programs lacking faculty expertise and/or time, online modules that could translate the workshop experience into a virtual environment. Through a structured survey-based approach, conference calls, and face-to-face meetings, four team-based learning (TBL)/flipped classroom exercises were developed: single-gene testing, prognostic gene panel testing, cancer gene panels, and whole-exome sequencing. As an important innovation in medical education, TBL emphasizes learner preparation outside of class and knowledge application inside the classroom [67]. Each exercise includes an instructor-delivered 15–30-min PowerPoint lecture introducing key concepts, a 60-min activity consisting of teams of 3–6 residents answering a series of questions with some necessitating the use of online genomic tools, and an instructor-delivered 15–30-min PowerPoint lecture presenting answers to the questions and incorporating a discussion of team responses.

The first workshop based on the exercises was held at the 2013 ASCP Annual Meeting. There have since been over 20 pathology-related workshops/courses with over 450 participants held internationally to educate both pathology trainees and practicing physicians in genomic pathology. These have ranged from covering all four exercises in 8 h to covering one or two exercises in 2–3 h. The data on the first three workshops was published as an "Educational Innovation" article in the *Journal of Graduate Medical Education* [68]. TBL had not been previously implemented and evaluated at a national meeting, and the positive results demonstrated the utility of this approach. There have also been eight train-the-trainer sessions in which participants discuss educational methodology and then work in teams as "students" to directly experience and learn how to implement the TBL approach.

To further assist others in implementing local teaching sessions, based on the successful workshops, an instructor handbook and tool kit have been developed. The over 80-page handbook not only contains workshop questions and answers but also detailed information on teaching using the flipped classroom and TBL format, a preparation checklist and tips on implementation. The tool kit consists of all the necessary handouts and PowerPoint lectures. For sites with limited faculty to teach genomic pathology, online modules have also been developed. The modules translate the TBL workshop experience to a virtual environment and include simulations to teach how to effectively access genomics website.

In December 2014, the handbook and tool kit were made available free of charge on the TRIG website, and in September 2016, the online modules were added [66]. Since the release of the handbook and tool kit, more than 750 individuals from over 40 different countries have registered to download TRIG material. In an August 2016 survey of handbook and tool kit downloaders, 31 of 67 respondents (46%) reported using at least some portion the materials with 776 medical students, 214 residents, 76 fellows, and 60 laboratory technologists. In a 2016 PRODS survey, approximately 30% of residency program directors report using the TRIG materials with over 120 trainees.

Aside from development, implementation, and evaluation of teaching materials, another major goal of the TRIG Working Group is to nationally assess the degree of resident training and knowledge in genomic pathology. Administered by the ASCP, the yearly pathology resident in-service exam (RISE) is taken by almost all residents in the USA (approxi-

mately 2500). Residents receive scores and percentile equivalents in comparison to their peer group postgraduate year (PGY) for each of the ten content areas. Scores on the exam allow residents to gauge their progress and have been correlated with board exam performance [69].

Since 2013, survey and knowledge questions related to genomic medicine have been included on the RISE [70]. Of note, from the initial PRODS survey suggesting that only 30% of residency programs had training in genomic medicine, the result has now climbed to almost 80% based on responses for the 2017 RISE. The RISE survey and knowledge questions created by the TRIG Working Group have provided valuable data on the degree and efficacy of resident training in genomic pathology on a national scale. There are few published examples of the use of an assessment tool with the scope of the RISE to study curricular improvement.

Future Directions

The TRIG Working Group has had considerable success in developing, implementing, and evaluating a genomics curriculum for pathology residents. Future goals will include updating the materials with integration of the published CAP genomics competencies for pathologists and recently developed AMP molecular pathology resident curriculum [56, 57]. To further encourage implementation of genomic training and with the new pathology milestones related to personalized medicine, the American Board of Pathology should continue to include relevant questions on the anatomic and clinical pathology certification exams [71]. The need to know the material to become board-certified incentivizes residents to learn genomic pathology. Continuing medical education programs also need to be established to teach practicing pathologists. Training beyond residency could build on the teaching tools created by individual programs, the TRIG Working Group, as well as resources developed for physicians in other specialties. Assessment tools to determine efficacy should be established for these educational programs.

There is also an ongoing work to adapt the TRIG model to other learners. Training in genomic pathology needs to begin in medical school. There have already been some innovative single institution approaches in undergraduate genomics education, but pathologists, because they are already playing a major role in teaching during the first 2 years of medical school, should take an active part in incorporating genomic pathology instruction into coursework. In 2017, the Undergraduate Medical Educators Section (UMEDS) of the APC formed the Undergraduate Training in Genomics (UTRIG) Working Group. This group is made up of pathology course directors, medical geneticists, and genetic counselors with representation from major pathology organizations, APHMG, ACMG, and NSGC. Using the TRIG curriculum as a starting point, the group hopes to develop a medical student curriculum and tools for implementation.

The utility of the TRIG approach has also been recognized by the ISCC. Following a review of TRIG progress at an ISCC face-to-face meeting, an ISCC Innovative Approaches Working Group was formed. With funding from the NHGRI, this group created four "universal modules" of plug-and-play TBL exercises; genes and diseases can be entered to create specialty-specific educational sessions. The modules have been used for workshops at the annual meetings of the American Heart Association (AHA), American Academy of Ophthalmology, and American Academy of Neurology (AAN) [72]. Similar to TRIG, an instructor handbook and tool kit have been released on the TRIG website to promote genomics education in other specialties.

Conclusions

As the director of the National Human Genome Research Institute wrote in 2011, "It is time to get serious about genomics education for all healthcare professionals" [28]. Based on published data, there is a need to improve education in the utility and interpretation of genomic testing for medical students, graduate trainees, and practicing physicians. Pathologists, with their access to tissue samples and expertise in laboratory testing, must play a leading role in ensuring the safe application of genomics to patient care.

Further work is needed to educate pathologists in genomics and should build on the resources created by individual programs, pathology organizations, as well as the TRIG Working Group. The latter provides a collaborative and structured model for curriculum design and assessment, not only for genomics education but, in other novel technologies, for other learners beyond residents and other specialties beyond pathology.

Acknowledgment This work was supported by the National Institutes of Health (1R25CA168544).

Reference

1. National Human Genome Research Institute. The human genome project completion: Frequently asked questions. 2017. https://www.genome.gov/11006943.%20accessed%205%20feb%202013/. Accessed 6-1-2017.
2. Rangachari D, VanderLaan PA, Le X, Folch E, Kent MS, Gangadharan SP, et al. Experience with targeted next generation sequencing for the care of lung cancer: insights into promises and limitations of genomic oncology in day-to-day practice. Cancer Treat Commun. 2015;4:174–81.

3. Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. Genome Biol. 2010;11(8):R82.
4. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011;305(15):1577–84.
5. Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. JAMA. 2015;314(9):913–25.
6. Stockley TL, Oza AM, Berman HK, Leighl NB, Knox JJ, Shepherd FA, et al. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. Genome Med. 2016;8(1):109.
7. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA. 2014;311(19):1998–2006.
8. Azim HA Jr, Michiels S, Zagouri F, Delaloge S, Filipits M, Namer M, et al. Utility of prognostic genomic tests in breast cancer practice: the IMPAKT 2012 working group consensus statement. Ann Oncol. 2013;24(3):647–54.
9. Zhang X. Value of molecular tests in cytologically indeterminate lesions of thyroid. Arch Pathol Lab Med. 2015;139(12):1484–90.
10. Nardi V, Sadow PM, Juric D, Zhao D, Cosper AK, Bergethon K, et al. Detection of novel actionable genetic changes in salivary duct carcinoma helps direct patient treatment. Clin Cancer Res. 2013;19(2):480–90.
11. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13(3):255–62.
12. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502–11.
13. Counsyl. Family Prep Screen. 2017. https://www.counsyl.com/services/family-prep-screen/2017. Accessed 6-2-2017.
14. Sequenom. MaterniT 21 Plus. 2017. https://www.sequenom.com/tests/reproductive-health/maternit21-plus. Accessed 6-2-2017.
15. St-Louis M. Molecular blood grouping of donors. Transfus Apher Sci. 2014;50(2):175–82.
16. Bio-IT. German Teams, BGI and Life Technologies Identify Deadly European E.coli Strain. 2017. http://www.bio-itworld.com/news/06/02/2011/German-teams-BGI-Life-Technologies-Identify-E-coli-strain.html. Accessed 6-2-2017.
17. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med. 2011;364(8):730–9.
18. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med. 2014;370(25):2408–17.
19. Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, et al. A genomic day in the life of a clinical microbiology laboratory. J Clin Microbiol. 2013;51(4):1272–7.
20. Giardiello FM, Brensinger JD, Petersen GM, Luce MC, Hylind LM, Bacon JA, et al. The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. N Engl J Med. 1997;336(12):823–7.
21. Kotzer KE, Riley JD, Conta JH, Anderson CM, Schahl KA, Goodenberger ML. Genetic testing utilization and the role of the laboratory genetic counselor. Clin Chim Acta. 2014;427:193–5.
22. Miller CE, Krautscheid P, Baldwin EE, Tvrdik T, Openshaw AS, Hart K, et al. Genetic counselor review of genetic test orders in a reference laboratory reduces unnecessary testing. Am J Med Genet A. 2014;164A(5):1094–101.
23. Klitzman R, Chung W, Marder K, Shanmugham A, Chin LJ, Stark M, et al. Attitudes and practices among internists concerning genetic testing. J Genet Couns. 2013;22(1):90–100.
24. Salm M, Abbate K, Appelbaum P, Ottman R, Chung W, Marder K, et al. Use of genetic tests among neurologists and psychiatrists: knowledge, attitudes, behaviors, and needs for training. J Genet Couns. 2014;23(2):156–63.
25. Laedtke AL, O'Neill SM, Rubinstein WS, Vogel KJ. Family physicians' awareness and knowledge of the genetic information nondiscrimination act (GINA). J Genet Couns. 2012;21(2):345–52.
26. Plunkett-Rondeau J, Hyland K, Dasgupta S. Training future physicians in the era of genomic medicine: trends in undergraduate medical genetics education. Genet Med. 2015;17(11):927–34.
27. Korf BR, Berry AB, Limson M, Marian AJ, Murray MF, O'Rourke PP, et al. Framework for development of physician competencies in genomic medicine: report of the Competencies Working Group of the Inter-Society Coordinating Committee for Physician Education in Genomics. Genet Med. 2014;16(11):804–9.
28. Feero WG, Green ED. Genomics education for health care professionals in the 21st century. JAMA. 2011;306(9):989–90.
29. Demmer LA, Waggoner DJ. Professional medical education and genomics. Annu Rev Genomics Hum Genet. 2014;15:507–16.
30. Slade I, Burton H. Preparing clinicians for genomic medicine. Postgrad Med J. 2016;92(1089):369–71.
31. Wiener CM, Thomas PA, Goodspeed E, Valle D, Nichols DG. "Genes to society" – the logic and process of the new curriculum for the Johns Hopkins University School of Medicine. Acad Med. 2010;85(3):498–506.
32. Salari K, Pizzo PA, Prober CG. Commentary: to genotype or not to genotype? Addressing the debate through the development of a genomics and personalized medicine curriculum. Acad Med. 2011;86(8):925–7.
33. Dhar SU, Alford RL, Nelson EA, Potocki L. Enhancing exposure to genetics and genomics through an innovative medical school curriculum. Genet Med. 2012;14(1):163–7.
34. Walt DR, Kuhlik A, Epstein SK, Demmer LA, Knight M, Chelmow D, et al. Lessons learned from the introduction of personalized genotyping into a medical school curriculum. Genet Med. 2011;13(1):63–6.
35. Gerhard GS, Jin Q, Paynton BV, Popoff SN. The anatomy to genomics (ATG) start genetics medical school initiative: incorporating exome sequencing data from cadavers used for anatomy instruction into the first year curriculum. BMC Med Genet. 2016;9(1):62.
36. G2C2. G2C2 Genetics/Genomics Competency Center. 2017. http://genomicseducation.net/. Accessed 6-2-2017.
37. Hyland KM, Dasgupta S, Garber KB, Gold J-A, Toriello H, Weissbecker K, et al. APHMG Medical School Core Curriculum in Genetics. 2017. http://media.wix.com/ugd/3a7b87_7064376a9eb346cfa1b85bc2f137c48f.pdf. Accessed 6-2-2017.
38. Goetz L, Bethel K, Topol EJ. Rebooting cancer tissue handling in the sequencing era: toward routine use of frozen tumor tissue. JAMA. 2013;309(1):37–8.
39. Dabbs DJ, Klein ME, Mohsin SK, Tubbs RR, Shuai Y, Bhargava R. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the Oncotype DX test: an independent quality assurance study. J Clin Oncol. 2011;29(32):4279–85.
40. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med. 2011;3(111):111ra121.
41. MacArthur D. Sample swaps at 23andMe: a cautionary tale. 2017. https://www.wired.com/2010/06/sample-swaps-at-23andme-a-cautionary-tale/6-7-2010. Accessed 6-2-2017.
42. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature. 2009;461(7265):724–6.
43. Mathias PC, Conta JH, Konnick EQ, Sternen DL, Stasi SM, Cole BL, et al. Preventing genetic testing order errors with a laboratory utilization management program. Am J Clin Pathol. 2016;146(2):221–6.
44. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone

receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol. 2011;29(32):4273–8.
45. Morain S, Greene MF, Mello MM. A new era in noninvasive prenatal testing. N Engl J Med. 2013;369(6):499–501.
46. Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. N Engl J Med. 2014;370(9):799–808.
47. Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, et al. Cell-free DNA analysis for noninvasive examination of trisomy. N Engl J Med. 2015;372(17):1589–97.
48. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA. 2006;296(2):212–5.
49. Vassy JL, Korf BR, Green RC. How to know when physicians are ready for genomic medicine. Sci Transl Med. 2015;7(287):287fs19.
50. National Society of Genetic Counselors. 2017. Who are genetic counselors? http://www.nsgc.org/page/whoaregcs. Accessed 6-12-2017.
51. Intersociety Council for Pathology Information. Career Opportunities in Pathology. 2017. http://www.pathologytraining.org/trainees/documents/recruit.ppt. Accessed 6-02-2017.
52. Tonellato PJ, Crawford JM, Boguski MS, Saffitz JE. A national agenda for the future of pathology in personalized medicine: report of the proceedings of a meeting at the Banbury Conference Center on genome-era pathology, precision diagnostics, and preemptive care: a stakeholder summit. Am J Clin Pathol. 2011;135(5):668–72.
53. Ross JS. Next-generation pathology. Am J Clin Pathol. 2011;135(5):663–5.
54. Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Anatomic Pathology and Clinical Pathology. 2017. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/300_pathology_2016.pdf. Accessed 6-2-2017.
55. Accrediation Council for Graduate Medical Education, American Board of Pathology. The Pathology Milestone Project. 2017. http://www.acgme.org/Portals/0/PDFs/Milestones/PathologyMilestones.pdf. Accessed 6-2-2017.
56. Aisner DL, Berry A, Dawson DB, Hayden RT, Joseph L, Hill CE. A suggested molecular pathology curriculum for residents: a report of the association for molecular pathology. J Mol Diagn. 2016;18(2):153–62.
57. Laudadio J, McNeal JL, Boyd SD, Le LP, Lockwood C, McCloskey CB, et al. Design of a Genomics Curriculum: competencies for practicing pathologists. Arch Pathol Lab Med. 2015;139(7):894–900.
58. Haspel RL, Arnaout R, Briere L, Kantarci S, Marchand K, Tonellato P, et al. A call to action: training pathology residents in genomics and personalized medicine. Am J Clin Pathol. 2010;133(6):832–4.
59. Kern DE, Thomas PA, Hughes MT. Curriculum development for medical education: a six-step approach. 2nd ed. Baltimore, MD: The Johns Hopkins University Press; 2009.
60. Haspel RL, Olsen RJ, Berry A, Hill CE, Pfeifer JD, Schrijver I, et al. Progress and potential: training in genomic pathology. Arch Pathol Lab Med. 2014;138(4):498–504.
61. Knowles MS, Holton EF, Swanson RA. The Adult Learner. Elsevier, Burlington, MA. 6th ed; 2005.
62. Callier SL. Swabbing students: should universities be allowed to facilitate educational DNA testing? Am J Bioeth. 2012;12(4):32–40.
63. Genzen JR, Krasowski MD. Resident training in clinical chemistry. Clin Lab Med. 2007;27(2):343–58.
64. Schrijver I, Natkunam Y, Galli S, Boyd SD. Integration of genomic medicine into pathology residency training: the Stanford open curriculum. J Mol Diagn. 2013;15(2):141–8.
65. Haspel RL, Atkinson JB, Barr FG, Kaul KL, Leonard DGB, O'Daniel J, et al. TRIG on TRACK: educating pathology residents in genomic medicine. Pers Med. 2012;9:287–93.
66. Training Residents in Genomics (TRIG) Working Group. Training Residents in Genomics (TRIG) Website. 2017. http://pathology-learning.org/trig. Accessed 6-2-2017.
67. Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. Med Teach. 2012;34(5):e275–87.
68. Haspel RL, Ali AM, Huang GC. Using a team-based learning approach at National Meetings to teach residents genomic pathology. J Grad Med Educ. 2016;8(1):80–4.
69. Rinder HM, Grimes MM, Wagner J, Bennett BD. Senior pathology resident in-service examination scores correlate with outcomes of the American Board of Pathology certifying examinations. Am J Clin Pathol. 2011;136(4):499–506.
70. Haspel RL, Rinder HM, Frank KM, Wagner J, Ali AM, Fisher PB, et al. The current state of resident training in genomic pathology: a comprehensive analysis using the resident in-service examination. Am J Clin Pathol. 2014;142(4):445–51.
71. American Board of Pathology. Taking an examination: Primary examinations. 2017. http://www.abpath.org/index.php/taking-an-examination/primary-certificate-requirements. Accessed 10-13-2017.
72. Musunuru K, Haspel RL. Improving genomic literacy among cardiovascular practitioners via a flipped-classroom workshop at a National Meeting. Circ Cardiovasc Genet. 2016;9(3):287–90.

Clinical Implementation of Next-Generation Sequencing (NGS) Assays

9

Joshua L. Deignan

Introduction

Though genomic sequencing assays originated in research laboratories, they have now made their way into clinical molecular diagnostic laboratories where they are increasingly adopted for clinical laboratory testing. The regulatory and technical requirements for a clinical laboratory test whose results will be communicated back to an ordering clinician for patient management are obviously more stringent than those of research assays, the results of which should not legally be used for patient care. Therefore, the task of implementing a genomic sequencing assay in the clinical environment poses a challenge regardless of whether the assay involves a gene panel, exome sequencing (ES), or genome sequencing (GS). As more clinical laboratories attempt to incorporate next-generation sequencing (NGS) technologies into their molecular diagnostic toolbox, there will be a continual need for appropriate NGS standards and guidelines from professional organizations, and several NGS guidelines have already emerged in the literature [1–3]. The purpose of this chapter is to give the reader an overview of the various issues that should be considered when a clinical laboratory director makes the decision to evaluate genomic sequencing as a potential method for clinical testing in his/her laboratory.

What Is a CLIA-Certified/CAP-Accredited Laboratory?

Clinical laboratory directors are likely already familiar with the Clinical Laboratory Improvement Amendments (CLIA), which provide quality standards and guidance on how clinical laboratories are required to operate. These regulations apply to every laboratory that performs testing on human specimens for clinical purposes (which by definition does not include research laboratories), and such laboratories are required by law to obtain CLIA certification in order to offer these tests. Many clinical laboratories are also accredited by the College of American Pathologists (CAP), which is one of the accrediting organizations that is approved to enforce the CLIA regulations through onsite inspections, the administration of proficiency testing surveys, and the provision of other educational resources. The CAP currently issues a "Molecular Pathology" checklist that specifies various items that a clinical molecular diagnostic laboratory is required to implement in order to obtain/maintain CAP accreditation, and the CAP checklists are the foundation for onsite CAP inspections. The CAP has recently published a checklist of items (as a section in the larger "Molecular Pathology" checklist) that are specific for laboratories performing clinical NGS. Finally, in addition to the specialty-specific checklist, molecular diagnostic laboratories are also required to fulfill the set of requirements listed in the "Laboratory General" and "All Commons" checklists, which pertain to all clinical laboratories.

Equipment and Reagents

One of the first decisions a clinical laboratory needs to make is what type of NGS platform to purchase. Currently, there are several commercially available platforms, some with more widespread use than others and each with its own unique technical characteristics and attributes that need to be considered. Platforms may differ with regard to acquisition cost (ranging from the hundreds of thousands of dollars to nearly a million dollars), length of sequence reads (short versus long), overall total sequencing capacity (low versus high), instrument size/footprint, sequencing time (hours to

J. L. Deignan (✉)
Department of Pathology and Laboratory Medicine,
David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
e-mail: jdeignan@mednet.ucla.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_9

days), and ease of use for the laboratory. The reader is referred to several existing reviews on this topic as well as other chapters in this textbook, as the technical specifications of the various platforms currently in use have previously been covered in much detail [4, 5]. Laboratories are also encouraged to communicate with colleagues at other institutions who have practical experience (which can be research lab or clinical lab experience) using the various types of platforms before making their own financial investment. Given that the technical specifications and cost of NGS platforms are continuously changing, it is often difficult to weigh the decision to financially invest now or wait for an upgraded or a completely new platform later which may be much more attainable. Laboratories may ultimately decide to outsource the technical sequencing for their assay to another laboratory with existing instrumentation while retaining their own professional interpretation capabilities in-house. However, note that for those choosing this route, the laboratory and personnel performing the technical sequencing still need to hold appropriate certification/licensure (more on that later).

As part of deciding what platform to invest in, laboratories will also need to decide what type of assay they want to implement. They may desire to launch a gene panel, covering a known set of clinically relevant genes for a defined condition (e.g., cardiomyopathy or deafness); they may desire to launch an ES test, covering the majority of all protein-coding regions in order to assist with the diagnosis of Mendelian disorders; or they may desire to launch a GS test, assessing as much of the entire genomic sequence as is technically feasible, in order not only to sequence the protein-coding regions but also to find pathogenic intronic variants and copy number alterations (gains and losses) which may be causal in undiagnosed individuals. Small benchtop sequencers tend to have a lower total sequencing capacity, which may be sufficient for a limited gene panel where the lab only wishes to sequence a handful of specific regions at a much greater depth, whereas larger sequencers with a higher total sequencing capacity may be more optimal for analyzing variants at a lower depth but across the entire exome or genome. Newer benchtop sequencers, which have the potential to perform ES and GS with a smaller footprint, are also emerging. If a laboratory decides to only sequence a set of genes/regions (such as a gene panel or the exome), the laboratory is encouraged to evaluate as many of the various available technologies and chemistries as possible in order to determine the one that works best for the intended purpose of the test [4]. This may involve the use of existing commercially available reagents, or it may involve working with a manufacturer to design and develop a custom reagent which can then be validated in the laboratory.

Personnel

Whereas personnel who perform NGS assays in research laboratories can be anyone at the undergraduate, graduate, or postdoctoral levels, current CLIA regulations require that only individuals who are appropriately trained/licensed/certified perform clinical laboratory testing. For all clinical molecular diagnostic testing, including NGS, all of the qualification requirements for personnel working in laboratories performing high-complexity testing will apply, and under CLIA regulations, individuals must have at least an associate degree (or the equivalent) with a major in a laboratory science in order to perform high-complexity testing. Though individuals are not specifically required to be licensed/certified in some states, it can be more challenging to find employment without some form of licensure/certification. In most states, documentation of relevant education as well as a satisfactory score on the American Society for Clinical Pathology (ASCP) Molecular Biology Examination is sufficient in order to perform clinical molecular diagnostic testing. Other states, like New York and California, have additional specific requirements. For example, technologists working in clinical molecular diagnostic laboratories in California are required to obtain one of two licenses, either a Clinical Laboratory Scientist license, allowing them to perform any type of clinical laboratory testing, or a Clinical Genetic Molecular Biologist Scientist license. The latter will only allow personnel to perform clinical molecular diagnostic testing but not any other types of clinical laboratory testing. Both require relevant educational experience, 1 year of full-time training in an approved training program, and a satisfactory score on an approved exam.

However, for NGS, many molecular biology certification and training programs do not currently address the complex challenges associated with these types of assays. Generalist training programs in California may tend to focus more time on other areas of the clinical laboratory system, such as chemistry, hematology, and microbiology, than on molecular diagnostic testing. Nevertheless, a technologist who obtains a generalist license is legally allowed to perform clinical NGS testing even without much experience. Some of these shortcomings are currently being addressed by various professional groups including the Association for Molecular Pathology (AMP). As NGS assays involve both a "wet lab" and a "dry" bioinformatics component, finding additional appropriately licensed individuals with a strong bioinformatics background also poses a challenge, as most of those individuals are likely to have acquired that experience as part of their graduate program and may not desire a future career as a clinical laboratory technologist. However, because the

bioinformatics analysis is also a part of the analytical component of any NGS clinical test, those individuals should ideally be appropriately licensed/certified as well.

Test Ordering, Clinical Information, and Informed Consent

Like any other clinical laboratory tests, NGS assays require an appropriate test order (on paper or electronically) placed by a clinician. For gene panels, laboratories may need to decide whether to list the analyzed genes directly on the order, which may include up to 50–100 genes, or publish that information elsewhere. Laboratories may also want to be able to offer clinicians the option of selecting which specific genes to analyze, though this would require the laboratory to have a bioinformatics-based mechanism for masking the results from specific genes during the analysis, interpretation, and reporting steps. However, it would likely be impossible to list all of the relevant regions covered as part of an exome or a genome sequencing test on a lab order. One possibility would be to refer the ordering clinician to a website, where information about specific gene and exon coverage depths could be provided in more detail. This would also allow the ordering clinician to determine whether a given clinical NGS test will be of benefit to their particular patient in question before ordering the test.

Depending on the intended purpose of the test, acquiring sufficient clinical information from the ordering clinician will be a critical component for proper interpretation of the results. Phenotypic keywords, suspected diagnoses, and any information pertaining to the patient's family history should be provided, and the laboratory will also need to have a process in place to address any NGS test requests for which the clinical information was not initially provided. Clinical samples, such as peripheral blood, have a limited stability, and sample quality may be compromised if they are stored for an extended period of time, while the ordering clinician is contacted for additional information. Having a genetic counselor on staff as part of the NGS laboratory may be necessary to allow for immediate contact with the ordering clinician in order to discuss and clarify cases for which the clinical indications are unclear. Given the high reagent cost for this type of clinical testing, the NGS laboratory is less likely to perform testing if there is uncertainty as to whether this type of testing is appropriate for a given patient.

Due to the complexity of NGS assays, informed consent from each patient should ideally also be obtained by the ordering clinician prior to ordering the test. The consent document should convey the purpose of the specific NGS test, its limitations, possible unintended consequences such as unexpected consanguineous familial relationships, and the types of samples to be obtained. If the laboratory intends to use either the remnant sample such as genomic DNA or the patient sequence data for future research after the test has been performed and reported, this requires human subjects' research consent as well as Institutional Review Board (IRB) approval. As part of this consent process, a mechanism by which the patient can opt for his or her sample not to be used in such a manner and be discarded following test completion should be provided as well. Consultation with the appropriate IRB and institutional or laboratory legal counsel is recommended during the process of creating any informed consent documentation.

Specimen Selection

Most clinical molecular diagnostic testing is done on blood and tissue specimens. Tissue samples are either formalin fixed and paraffin embedded (known as FFPE tissue) or snap frozen in liquid nitrogen immediately after a clinical procedure and stored at −80 °C until the laboratory performs the test. They may also be submitted as cell cultures (such as fibroblasts). Laboratories wishing to set up an NGS assay will need to decide which type(s) of sample(s) they are going to accept, based on the intended purpose(s) of the test. For germline analyses looking for variants implicated in Mendelian disorders, blood is usually preferred. Laboratories may also wish to accept pre-extracted genomic DNA in order to facilitate international requests. However, in such circumstances, it is recommended that the final report contain a disclaimer stating that the genomic DNA used for the analysis was extracted outside the clinical laboratory and was tested at the request of the ordering clinician and the accuracy of the identifying information provided with the specimen regarding its patient of origin cannot be independently confirmed by the clinical laboratory. This may help protect the clinical laboratory from potential liability in case there had been a sample mix-up in the laboratory where the DNA extraction was initially performed. Laboratories that elect to accept FFPE tissue for various NGS oncology assays should be aware of inherent limitations associated with FFPE tissue use. Formalin fixation leads to DNA cross-linking that often results in fewer long intact DNA fragments compared to those obtained from fresh frozen tissue samples. Therefore, assays utilizing FFPE tissue-derived DNA need to be designed for amplification/capture of only short genomic sequence fragments (100–200 bp). As previously mentioned, fresh frozen tissue obtained after biopsy will usually result in

a higher quality and more intact DNA but requires special arrangements during transportation to the molecular laboratory, including transport on wet or dry ice, storage in a −80 °C freezer, and additional biohazard precautions. FFPE and fresh frozen tissue (as well as cell cultures) may also be requested by the ordering clinician to assess potential mosaicism in typically germline conditions.

Reporting of Results and Variant Interpretation

For molecular diagnostic tests targeting a single clinically relevant variant, laboratories typically report findings as positive, indicating that the specified variant was observed, or negative, indicating that the specified variant was not observed. For larger variant panels, such as those indicated for conditions such as cystic fibrosis, laboratories typically report findings as (a) positive, when one or more variant(s) was/were observed, or (b) negative, indicating that no variants were observed. For full-gene sequencing assays (such as those for *BRCA1/2*-associated breast cancer risk), laboratories may choose to only report pathogenic variants, or they may choose to report any and all variants they observe, whether pathogenic or benign. However, for NGS assays, there will likely be too many variants observed to include all of them in the report, even if the lab wanted to do so. Therefore, labs will need to decide on appropriate reporting criteria in order to clearly convey relevant or potentially relevant information back to the ordering clinician.

Guidelines currently exist on how to compose a clinical NGS report [3]. Similar to any basic laboratory report, laboratories should ideally provide an overview statement at the top of the report regarding whether any clinically relevant finding was identified, especially in relation to the clinical indication for ordering the test. Reporting issues unique to laboratories performing ES and GS assays emerge from the fact that such tests may also find clinically relevant variants in genes that are unrelated to the primary clinical concerns in the tested patient, usually referred to as "incidental" or "secondary" findings, as well as many variants which are already known to be benign, as they are very common in the general population [6]. Incidental/secondary findings have previously been addressed in detail by professional guidance documents [7–10]. The decision regarding which categories of variants to report (and which categories not to report) ultimately resides with the laboratory director, but this information should be clearly conveyed to both the ordering clinician and the patient prior to performing the test during the informed consent process.

In addition to reporting and attempting to interpret clinically relevant variants, another equally important component of a clinical report is the limitations in the laboratory's ability to fully assess the clinical question based on the method of testing. For certain tests, this may be as simple as stating that only specific genes are covered in a given gene panel assay and listing the anticipated clinical sensitivity and specificity of only testing for variants in those genes. For others, such as an ES test, the limitation disclaimer statement may require a description of the sequence capture method used indicating that only a certain percentage of the clinically relevant variants are expected to be sufficiently sequenced by this test. Some laboratories may elect to only interpret and report single-nucleotide variants (SNVs), whereas others may also wish to interpret and report small insertions and deletions. Other laboratories may also choose to report larger copy number alterations. While some molecular laboratories may perform confirmation of clinically relevant variants using Sanger sequencing or other appropriate methods, others may forego confirmation by an alternative methodology altogether [11]. Regardless of the decision made by an individual laboratory, laboratories should attempt to clearly communicate this type of information to the ordering clinician so they fully understand the test they ordered.

In addition to deciding which variants to report and providing statements regarding the limitations of the assay, the laboratory should utilize as many resources as possible when providing its interpretation of individual variants. Such resources include the American College of Medical Genetics and Genomics (ACMG) sequencing guidelines, among others [12, 13]. In addition to utilizing existing variant databases such as ClinVar, molecular laboratories should also closely assess family history and variant population frequency (using databases like ExAC), because many of the variants observed in NGS assays, especially ES and GS, may have never been previously reported in association with a similar phenotype. Rare variants are more likely to be pathogenic than common ones, but the presence of a rare potentially pathogenic variant in a clinically relevant gene does not prove causality.

Bioinformatics Requirements and Data Storage

Many molecular diagnostic tests are primarily comprised of a "wet lab" component with minimal interpretation/calculation required prior to finalizing a result, e.g., calculation of the size of an *FMR1* CGG allele for fragile X syndrome. For some FDA-approved assays, such as some commercially available assays for *KRAS* and *BRAF* mutations in colorectal adenocarcinoma and melanoma, respectively, no interpretation/calculation is required. Preset software programs, which are part of the FDA test submission and approval of such assays, handle any interpretation/calculation in the background prior to generating a result.

In contrast, as previously mentioned, NGS assays require a more substantial bioinformatics/interpretation component. At present, there are only a handful of commercially available software packages to assist with NGS assay result interpretations [14]. However, many laboratories offering this type of clinical testing have assembled their own in-house-developed bioinformatics pipelines consisting of various algorithms and databases [15, 16]. Ideally, a laboratory should employ several qualified bioinformaticists who are familiar with both bioinformatics and molecular biology. However, bioinformaticists who are also trained in the requirements needed in a CLIA-certified clinical laboratory may be difficult to find, and formal educational pathways may be needed in this area in the future. Given that much of the downstream interpretation and reporting will be based on the results of the bioinformatics analysis, it will be important for the laboratory to allocate sufficient resources toward ensuring the appropriateness of its pipeline.

Laboratories that plan to offer NGS assays should arrange for a marked increase in data storage requirements. The information produced by NGS platforms far exceeds the capacity of a typical desktop computer hard drive or DVD. Therefore, laboratories should be prepared for a financially significant investment in this aspect of the testing, which may be equivalent to the financial investment required for acquiring the NGS platform itself. Specific data files (such as BAM, FASTQ, and/or VCF files) will also need to be stored for extended periods of time in order to allow for reproduction of the original results as well as reanalyses if desired.

Test Validation

Most clinical laboratory tests involving NGS are currently considered to be laboratory-developed tests (LDTs) and require a full validation, as opposed to the usually more limited verification required for FDA-cleared or FDA-approved products. The main elements of required assessment as stipulated by CLIA regulations pertain to accuracy (how well do the results match what they should be), precision (would a laboratory get the same result from the same sample any day or time), analytical sensitivity (what percentage of a variant allele could a laboratory detect among a background of reference alleles), analytical specificity (what is the effect of neighboring variation or interfering substances on the test results), reference range (what type of results does the laboratory consider to be normal), and reportable range (what type of results does the laboratory consider to be abnormal) [17, 18]. The reader is referred to other recent publications that contain detailed suggestions for the validation of NGS-based testing including AMP- and CAP-issued guidelines on NGS assays [1–3]. Suitable reference materials for validation (such as those available from NIST) are now commercially available for use by laboratories [19].

Proficiency Testing

The requirement for proficiency testing is defined in the CLIA regulations and applies to all molecular tests. Laboratories must participate in proficiency testing at least semiannually using one of several available mechanisms. Typical proficiency testing surveys, such as those administered by the CAP, consist of three blinded samples sent out to every laboratory performing a given clinical laboratory test (e.g., a cystic fibrosis carrier screening mutation panel) twice per year. Laboratories are expected to treat these samples in the same manner as they would treat real patient samples. Once testing is performed on the proficiency testing samples, final results and interpretations are submitted back to the organization administering the proficiency survey. The laboratory is subsequently graded as "acceptable" or "unacceptable" based on either existing values which are known to be the "true" values or based on consensus findings among the majority of participating laboratories.

Proficiency testing for ES/GS assays presents several unique challenges. First, many clinical laboratories may find it to be cost prohibitive to test three additional samples, twice yearly, at the cost of hundreds to thousands of dollars per sample. Second, there are only a limited number of samples with known genotypes throughout the genome (such as J. Craig Venter's genomic DNA) [20] that would be ideal for proficiency testing for ES and GS assays. Third, as already described, laboratories vary in terms of the genomic regions they interrogate and which types of variants they interpret, so it may be challenging to determine which variants in the genome to use for grading purposes if a sample is submitted to multiple laboratories. Thankfully, the CAP has already created a formal, method-based proficiency testing survey (known as the NGS survey). However, if a laboratory chooses not to subscribe to the NGS survey, they would be encouraged to establish interlaboratory sample exchanges with other clinical laboratories performing similar clinical NGS testing. It should be noted that there are also two other available CAP sequencing proficiency surveys (known as SEC and SEC1). These sequencing surveys can serve to assess a laboratory's ability to analyze and interpret variants of interest according to Human Genome Variation Society (HGVS) nomenclature using provided or newly created Sanger sequencing traces, respectively [21]. The SEC and SEC1 surveys also assess a laboratory's ability to appropriately interpret variants (as pathogenic, VUS, or benign), whereas the NGS survey currently does not.

Conclusion

Clinical laboratories need to be aware of the challenges associated with implementing NGS assays before deciding whether or not to undertake the investment, and a laboratory may choose to postpone offering this type of testing until the instrumentation cost and bioinformatics investment become less financially prohibitive. This technology is not likely to disappear from the clinical arena, so clinical laboratories, as well as practicing and training clinicians, should remain continually aware of the improvements that are occurring with the various platforms, databases, and bioinformatics programs, as these technologies are already impacting the medical management of patients and will likely continue to do so in the foreseeable future.

References

1. Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Greiner TC, Grody WW, Hambuch T, Kalman L, Kant JA, Klein RD, Leonard DG, Lubin IM, Mao R, Nagan N, Pratt VM, Sobel ME, Voelkerding KV, Gibson JS. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. J Mol Diagn. 2012;14(6):525–40. https://doi.org/10.1016/j.jmoldx.2012.04.006. Epub 2012 Aug 20.
2. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, Lu F, Lyon E, Voelkerding KV, Zehnbauer BA, Agarwala R, Bennett SF, Chen B, Chin EL, Compton JG, Das S, Farkas DH, Ferber MJ, Funke BH, Furtado MR, Ganova-Raeva LM, Geigenmüller U, Gunselman SJ, Hegde MR, Johnson PL, Kasarskis A, Kulkarni S, Lenk T, Liu CS, Manion M, Manolio TA, Mardis ER, Merker JD, Rajeevan MS, Reese MG, Rehm HL, Simen BB, Yeakley JM, Zook JM, Lubin IM. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30(11):1033–6.
3. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15(9):733–47.
4. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol. 2011;29(10):908–14.
5. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11(5):759–69.
6. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA. 2006;296(2):212–5.
7. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG, American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565–74.
8. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19(2):249–55.
9. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG. CORRIGENDUM: ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2017;19(5):606.
10. Hehir-Kwa JY, Claustres M, Hastings RJ, van Ravenswaaij-Arts C, Christenhusz G, Genuardi M, Melegh B, Cambon-Thomsen A, Patsalis P, Vermeesch J, Cornel MC, Searle B, Palotie A, Capoluongo E, Peterlin B, Estivill X, Robinson PN. Towards a European consensus for reporting incidental findings during clinical NGS testing. Eur J Hum Genet. 2015;23(12):1601–6.
11. Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, Deignan JL. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. Genet Med. 2014;16:510. https://doi.org/10.1038/gim.2013.183.
12. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.
13. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017;100(2):267–80.
14. Reinert K, Langmead B, Weese D, Evers DJ. Alignment of next-generation sequencing reads. Annu Rev Genomics Hum Genet. 2015;16:133–51.
15. Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. Brief Bioinform. 2011;12(5):489–97.
16. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11(5):473–83.
17. Jennings L, van Deerlin VM, Gulley ML, College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med. 2009;133(5):743–55.
18. Mattocks CJ, Morris MA, Matthijs G, Swinnen E, Corveleyn A, Dequeker E, Müller CR, Pratt V, Wallace A, EuroGentest Validation Group. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet. 2010;18(12):1276–88.
19. Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A, Weng Z, Liu Y, Mason CE, Alexander N, Henaff E, McIntyre AB, Chandramohan D, Chen F, Jaeger E, Moshrefi A, Pham K, Stedman W, Liang T, Saghbini M, Dzakula Z, Hastie A, Cao H, Deikus G, Schadt E, Sebra R, Bashir A, Truty RM, Chang CC, Gulbahce N, Zhao K, Ghosh S, Hyland F, Fu Y, Chaisson M, Xiao C, Trow J, Sherry ST, Zaranek AW, Ball M, Bobe J, Estep P, Church GM, Marks P, Kyriazopoulou-Panagiotopoulou S, Zheng GX, Schnall-Levin M, Ordonez HS, Mudivarti PA, Giorda K, Sheng Y, Rypdal KB, Salit M. Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data. 2016;3:160025.
20. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. PLoS Biol. 2007;5(10):e254.
21. Richards CS, Palomaki GE, Lacbawan FL, Lyon E, Feldman GL. Three-year experience of a CAP/ACMG methods-based external proficiency testing program for laboratories offering DNA sequencing for rare inherited disorders. Genet Med. 2013;16:25–32.

Regulatory and Reimbursement Issues Related to Genomic Testing Services

10

Jan A. Nowak and Anthony Sireci

Introduction

The development of advanced genetic sequencing technologies and genomic testing services challenges the existing regulatory framework for clinical laboratory testing. These challenges will demand a refinement and adaptation on the part of laboratories, professional and accrediting organizations, vendors and manufacturers, and regulatory agencies of existing standards and practices to accommodate novel genomic technologies and clinical applications. The value of existing concepts of laboratory-developed tests and companion diagnostics will need revision in order to accommodate genomic sequencing assays. In some circumstances, these novel technologies will challenge established definitions of disease and the foundations of medical practice.

Similarly, reimbursement for genomic sequencing assays will demand a reassessment of the traditional understanding of laboratory testing, moving away from simple chemical analyses to syndromically defined genetic inquiries necessitating increasing amounts of professional work and involvement. The existing CPT (current procedural terminology) structure can be modified to accommodate these new realities but will also demand re-evaluation of existing concepts of laboratory testing and professional services related to test interpretation and usage.

Laboratory Regulations

Traditionally, novel technologies have been introduced into medical laboratories and clinical usage through fairly defined routes. Technically superior analytical methodologies or entirely novel tests would first emerge from research laboratories into clinical settings in select centers with specific interests in a particular analyte. These initial experiences would typically be reported and vetted at professional society meetings and in peer-reviewed publications where they might attract the attention of other professionals who would incorporate these technical advancements into their own laboratories' offerings and assays. Eventually such advancements would supplant extant methodologies and ultimately become standard of care. Assays, methods, and associated instrumentation which offer investment opportunities might be developed commercially and marketed, further disseminating a particular technology and promoting further standardization of methodology and clinical usage. This route of introduction was made possible by the established ability of individual laboratories to modify and develop analytical assays for clinical use as authorized by the Clinical Laboratory Improvement Act (CLIA) of 1988.

Under CLIA, laboratories are authorized to implement their analytic procedures for clinical use if they adhere to these basic requirements of laboratory-developed tests (LDTs) and Food and Drug Administration (FDA) cleared or approved tests [1]:

> *42 CFR 493.1253 – Standard: Establishment and verification of performance specifications*
>
> (1) Verification of performance specifications. Each laboratory that introduces an unmodified, FDA-cleared or approved test system must do the following before reporting patient test results:
>
> (i) Demonstrate that it can obtain performance specifications comparable to those established by the manufacturer for the following performance characteristics:
>
> (A) Accuracy.
> (B) Precision.
> (C) Reportable range of test results for the test system.
>
> (ii) Verify that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population.

J. A. Nowak (✉)
Department of Pathology and Laboratory Medicine,
Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA
e-mail: Jan.nowak@roswellpark.org

A. Sireci
Pathology and Cell Biology, Columbia University Medical Center,
New York, NY, USA
e-mail: ans2133@cumc.columbia.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_10

(2) Establishment of performance specifications. Each laboratory that modifies an FDA-cleared or approved test system, or introduces a test system not subject to FDA clearance or approval (including methods developed in-house and standardized methods such as text book procedures, Gram stain, or potassium hydroxide preparations), or uses a test system in which performance specifications are not provided by the manufacturer must, before reporting patient test results, establish for each test system the performance specifications for the following performance characteristics, as applicable:

(i) Accuracy.
(ii) Precision.
(iii) Analytical sensitivity.
(iv) Analytical specificity to include interfering substances.
(v) Reportable range of test results for the test system.
(vi) Reference intervals (normal values).
(vii) Any other performance characteristic required for test performance.

(3) Determination of calibration and control procedures. The laboratory must determine the test system's calibration procedures and control procedures based upon the performance specifications verified or established under paragraph (b)(1) or (b)(2) of this section.

(c) Documentation. The laboratory must document all activities specified in this section.

Sec. 493.1254 Standard: Maintenance and function checks

(a) Unmodified manufacturer's equipment, instruments, or test systems. The laboratory must perform and document the following:

(1) Maintenance as defined by the manufacturer and with at least the frequency specified by the manufacturer.
(2) Function checks as defined by the manufacturer and with at least the frequency specified by the manufacturer. Function checks must be within the manufacturer's established limits before patient testing is conducted.

(b) Equipment, instruments, or test systems developed in-house, commercially available and modified by the laboratory, or maintenance and function check protocols are not provided by the manufacturer. The laboratory must do the following:

(1) (i) Establish a maintenance protocol that ensures equipment, instrument, and test system performance that is necessary for accurate and reliable test results and test result reporting.
(ii) Perform and document the maintenance activities specified in paragraph (b)(1)(i) of this section.

(2) (i) Define a function check protocol that ensures equipment, instrument, and test system performance that is necessary for accurate and reliable test results and test result reporting.
(ii) Perform and document the function checks, including background or baseline checks, specified in paragraph (b)(2)(i) of this section. Function checks must be within the laboratory's established limits before patient testing is conducted.

Laboratory Regulations for Genomic Testing

Establishment of performance specifications for a single analyte is often challenging, but addressing these seemingly simple concepts for genomic breadth test procedures is particularly daunting. Genomic sequencing assays are technically complex and capable of addressing multiple "analytes" on a scale not previously seen. Consequently, the usual methods and standards for addressing the analytical performance characteristics required by CLIA are impractical, and alternative approaches to assuring compliance will need to be developed. Furthermore, the technical, chemical, and interpretive complexity of these genomic technologies requires that commercial developments play a much more significant role, and at a much earlier stage in test evolution, in bringing these assays to reality. Finally, the complexity of the information achievable by genomic sequencing assays, not to mention the volume of information, makes clinical validation and clinical utility difficult to define in the traditional sense. While the popular press continues to highlight the coming revolution of genomic medicine, professional societies struggle to conform to traditional practices for assuring quality and uniformity in testing, and vendors and manufacturers tread cautiously in a rarified atmosphere of regulatory guidance, with the regulatory agencies themselves uncertain of how to best perform their missions. In the interim, some entrepreneurial laboratories have proceeded to offer genomic testing, with their own interpretations of CLIA requirements and of what constitutes clinical validity. Some laboratories have offered such testing directly to patients, or "consumers," obviating any allegiance to CLIA altogether. The issue being challenged is not simply the correct mechanism of new test evolution, but much more fundamentally what constitutes the practice of clinical laboratory medicine and to what extent genomic information is a part of that practice.

Although some might argue that a revolutionary approach to regulatory oversight of genomic testing is in order, it is reasonable to expect that the foundation of laboratory medicine practice as defined by CLIA continue and that we will more likely see an evolution of established practices to accommodate genome-based testing. Understanding that, it is worthwhile exploring several current regulatory concepts that will impinge on the clinical adaptation of genomic testing methodologies.

Laboratory-Developed Procedures

The ability of laboratories to develop and implement their own assays (laboratory-developed procedures or tests, LDTs) has been a cornerstone of medical advancement whereby improvements in laboratory analyses could be introduced into clinical usage in a controlled and regulated fashion. With increasing acceptance, such laboratory-developed assays eventually overtake and replace extant methodologies. The vast majority of clinical assays in use today had their origins in LDTs. The responsibility for assuring LDT assay credibility rests with those entities authorized by CLIA to inspect such laboratories for compliance with the regulations, as well as with the professional standards of those entrusted with administering those laboratories, typically pathologists,

in their roles as laboratory medical directors. In recent years, this concept of the LDT or laboratory-modified test has been challenged from at least two quarters. The emergence of highly technically complex testing systems has fostered the establishment of large, high-volume corporate laboratories often devoted to a single, frequently proprietary, assay. Test complexities, as well as IP restrictions, typically preclude adoption of such tests by other laboratories. Whereas CLIA regulations are generally respected through internal studies and validations, the assurances that accompany widespread use by multiple laboratories and institutions, affecting proficiency testing, peer-to-peer debate and review, and ultimately quality improvement, would seem to be diminished by this adaptation of the sole-provider LDT paradigm.

The traditional laboratory-developed or laboratory-modified test has also been challenged by manufacturers who have endeavored to develop specific "companion diagnostic tests," a concept promoted by the FDA to ensure that appropriate tests and devices are available to support specific therapeutic claims. The requirement that specific tests be developed, validated, and approved by the FDA in concert with specific therapeutic drugs necessitates considerable expense on the part of the developer, costs that are not realized by the enterprising LDT developer. Consequently, the FDA has been petitioned by some test manufacturers to enforce its claimed jurisdiction over LDTs and require that all LDTs be subject to formal review by FDA before being placed into clinical service. Whereas the notion of a companion diagnostic has some legitimacy for the agency in ensuring that validated and approved testing methodologies are available for all approved drugs that require them, extension of the concept to mean that the approved assay methodologies are the best available and are the only tests that can be used in concert with a particular drug fails terribly in practice when one envisages the impracticality that a "one drug–one test–one clinical condition" constraint creates. Furthermore, having ordained a particular assay as the test of choice, the FDA effectively disincentivizes any further test improvement (and in that, may be franchising certain "harms" to patients). The concept is also oblivious to already established LDT-based tests. For some analytes, LDTs have been the standard of laboratory practice for multiple years, with documented proficiency, often with performance characteristics exceeding those of a more recently approved companion diagnostic. None of this historic information is reviewed or acknowledged whenever the FDA touts the approval of yet another "companion diagnostic" as the latest major step in the advancement of personalized medicine.

A more realistic approach would be to recognize that certain analytes, and not tests, have bearing on the performance of particular drugs. This notion is rooted in our understanding of biology, which is ultimately the basis of molecular medicine. Tests which are designed to evaluate these specific "companion analytes" must demonstrate certain performance characteristics in order to be valid. The necessary performance characteristics are defined in the peer-reviewed literature and are constantly subject to review and revision. A specific assay may be assigned "companion analyte test" status if it meets certain performance requirements. Laboratory assays used in clinical trials (clinical trial assays, CTAs) may deserve some special recognition in that they frequently, but not always, help define the necessary performance characteristics for the biological analyte. Formal approval of CTAs as in vitro diagnostic products (IVDs) could substantiate their utility without introducing the negative consequences that accompany "companion diagnostic" designation.

Movement away from the "companion diagnostic" terminology is also desirable in underscoring that laboratory test results, in themselves, are rarely diagnostic. Every laboratory test result needs to be interpreted and understood in the context of a specific patient by the patient's physician. The laboratory test result, however informative, never dictates the course of treatment. That decision needs to be made individually for each patient. To think otherwise is the ultimate depersonalization of medicine.

The FDA and Medical Practice in Clinical Laboratories

Whereas the practice of laboratory medicine is under the purview of CLIA, the FDA also claims significant jurisdiction over clinical testing in that such tests, and the instruments and reagents used to perform them, are medical devices, specifically IVDs. The FDA includes LDTs in this category and considers laboratories which use LDTs to be medical device manufacturers. The history of "enforcement discretion" by the FDA in administering traditional LDTs suggests that there is clear understanding within the agency of the difference between a commercially developed product intended for distribution and sale and a locally developed biochemical assay or modification of an approved assay to better serve a limited population. The notion that a corporate laboratory can offer a menu of proprietary LDTs has presented a challenge to FDA's traditional practice in overseeing LDTs. The FDA has been developing a refinement of its LDT oversight policy for several years but has yet to release further guidance.

The FDA has proposed a three-tier classification system for laboratory tests based on "risk" of potential harm to a patient. The lowest tier includes tests which generate results which are typically interpreted in the context of other clinical and laboratory test values and which, in themselves, are unlikely to cause harm to a patient should they be "incorrect." The high-risk category includes tests which the FDA holds to be "determinative" for a particular drug or course of therapy. Examples include tests for targeted agents (*HER2* fluorescence in situ hybridization (FISH) for Herceptin in the treatment of breast cancer, *EGFR* mutation testing for the use of Tarceva in the treatment of non-small cell lung

cancer (NSCLC), and *BRAF* mutation testing for the treatment of metastatic malignant melanoma with vemurafenib. The FDA has advocated that any tests in the Tier 3 risk group must be reviewed and approved by the FDA prior to clinical implementation, a requirement that is currently not imposed on LDTs. This line of thinking would create a conundrum for many LDT performing laboratories, both in terms of the work required for submission to FDA for test approval and in terms of practicality because the same analyte may be evaluated for different purposes in different circumstances. The *BRAF* p.Val600Glu (commonly known as p.V600E) mutation may be predictive of sensitivity to vemurafenib therapy in metastatic malignant melanoma but can be used as a corroborative diagnostic marker in evaluation of FNA (fine needle aspirate) specimens of thyroid nodules, and as a surrogate for *MLH1* hypermethylation in cases of microsatellite-unstable colorectal cancers, bearing on the likelihood of Lynch syndrome. The *BRAF* mutation also has prognostic significance for colorectal cancer. Thus, in different circumstances the same analyte (and the same analytical assay) may be assigned to different risk tiers. The concept that laboratory tests can be neatly categorized further deteriorates when one considers that many laboratory tests can become "determinative" in some clinical situations, e.g., a single blood glucose measurement, a single serum potassium determination, or an antibiotic sensitivity profile for a bacterial pathogen. Whether performed with an FDA-approved IVD or with an LDT, it must be acknowledged that no laboratory test is perfect (or diagnostic, as discussed earlier) and that an understanding of its performance characteristics and its application in any specific clinical scenario demands professional knowledge and judgment. The FDA is creating a conundrum in pursuing too literal an interpretation of "risk" in applying this categorization scheme. The implication is that the FDA is the ultimate arbiter of how tests should be performed, how their results should be interpreted, and how the results should be used clinically, actions which arguably exceed FDA's mission.

While awaiting FDA promised guidance on LDTs, several professional organizations have issued their own proposals. Most of these have grappled with the risk concept put forward by the FDA. The College of American Pathologists (CAP) reserved the highest-risk category to tests that utilize a nontransparent algorithm and are not subject to the usual checks and balances afforded by more straightforward tests performed by most laboratories, whose performance can be assessed through traditional proficiency testing programs and peer inspections. CAP acknowledged that there may be some tests that deserve greater scrutiny, i.e., those that are associated with targeted therapies but which are more appropriately placed in a lower-risk tier. The quality performance of such tests, performed as LDTs, could be assured through a stepped-up inspection system that would include peer review of test validation data. The Association for Molecular Pathology (AMP) has issued a similar proposal.

The current complex and convoluted regulatory oversight environment will continue to be challenged by the genomic sequencing assays as will our society's fundamental understanding and concepts of health, disease, and the practice of medicine. Such challenges should be welcome if they strengthen the basic principles and purposes for the rules and agencies that were established to assure high-quality clinical laboratory testing. Conversely, established oversight mechanisms will need to be adapted to accommodate the novel characteristics of genomic sequencing assays. It should be clear that oversight of clinical laboratory tests and testing does not reside in any one law or agency. This responsibility is shared by a wide array of government agencies, accrediting organizations, professional societies, payers, clinicians, and medical specialists, each of whom focuses attention on part of the complex fabric that we consider oversight. Consequently, we should expect the "new" rules for oversight of genomic testing to emerge from a variety of quarters that will address issues of safety, efficacy, quality assurance, analytical and clinical validity, and clinical utility.

Reimbursement for Genomic Tests

Reimbursement for pathology and laboratory services is intrinsically linked to the American Medical Association (AMA)'s CPT system for accurately describing medical, surgical, and diagnostic services. Once defined and codified, specific services are assigned values by the Centers for Medicare and Medicaid Services (CMS) or private payers through a number of different mechanisms, which typically endeavor to accommodate the various technical and professional components necessary for providing those services in specific clinical situations.

The vast majority of laboratory test codes are analyte specific (e.g., 84295 sodium; serum, plasma, or whole blood) making it readily apparent to payers what test was performed, and with linkage to specific ICD10 (International Classification of Diseases, Tenth Revision) codes, in what clinical setting, allowing for some degree of confidence of appropriate test usage. Such transparency was lacking for the molecular pathology codes available through 2012 which were descriptive of test component methodologies (e.g., 83896 nucleic acid probe each) and demanded the application of multiple codes for any one analyte, further obscuring what service was being provided and making linkage to specific clinical scenarios impossible. Consequently, in 2009 the AMA CPT Editorial Panel commissioned a working group to develop a revised coding scheme in order to recognize molecular pathology tests in a manner consistent with other laboratory and pathology services. These new molecular codes were first introduced in 2012 and fully implemented in 2013 with the retirement of the older methodology-based codes.

Designed with input from a variety of stakeholders including public and private payers, test vendors, laboratories, and

trade and professional organizations, the new molecular pathology CPT codes are exquisitely analyte specific, frequently with descriptors that suggest the clinical scenario for test usage (e.g., 81241 F5 (coagulation factor V) (e.g., hereditary hypercoagulability) gene analysis, Leiden variant). Organized into two levels, the Tier 1 codes accommodate the most commonly performed molecular diagnostic tests. The rationale for establishing this grouping of tests was based on an acknowledgment that the methodologies for performing those assays, while perhaps still varied, had matured sufficiently that a single reimbursement value could fairly accommodate all laboratory practice settings. The Tier 2 grouping was intended to accommodate less common tests for which methodologies had not sufficiently evolved to define common practice parameters, either in terms of test design, instrumentation, or professional work. These tests were grouped into levels based on an acknowledged level of test complexity using existing methodologies, with traditional Sanger sequencing as the prototype methodology used by most laboratories. Categorization in Tier 2 is not expected to be permanent. With time, as methodologies become more standardized, a Tier 2 test could eventually be assigned a specific Tier 1 code. In recent years the AMA CPT Editorial Panel has moved many of the Tier 2-coded tests into Tier 1 through its ongoing code set maintenance efforts. The resemblance of the Tier 2 scheme to the surgical pathology code levels (88300–88309) is not happenstance, recognizing the functionality of the surgical pathology codes. A significant difference from the surgical pathology codes is that the Tier 2 codes cannot be self-assigned. Unassigned tests are relegated to the unlisted procedure code (81479, unlisted molecular pathology procedure).

Like the Tier 1 codes, the Tier 2 codes endeavor to be analyte specific, with the vast majority defining specific genetic variations that oftentimes define a specific inherited syndrome. Perusal of the Tier 2 codes shows that the majority are essentially syndromically defined, i.e., this is the molecular inquiry needed to evaluate one specific (genetic) clinical question. This evolving nature of the molecular pathology codes distinguishes these tests from traditional laboratory tests. These complex molecular assays move beyond simply analytical detection of a biochemical analyte to a more involved evaluation of genetic complexity related to a specific clinical scenario. In that sense, it is more appropriate to refer to these evaluations as "services" rather than "tests," acknowledging the significant professional knowledge and judgment necessary to appropriately perform, understand, and interpret such tests.

From a reimbursement perspective, Tier 1 and Tier 2 were recommended by the AMA Editorial Panel to be valued on the physician fee schedule (PFS), as opposed to the clinical laboratory fee schedule (CLFS), where the majority of clinical laboratory tests are valued. Placement on the PFS would have recognized that molecular diagnostic procedures involve a significant amount of interpretative, professional effort, similar to surgical pathology procedures. The codes had been vetted and valued by the RVS Update Committee (RUC). Nevertheless, the Centers for Medicare and Medicaid Services (CMS) placed the codes on the CLFS.

This is the reimbursement setting in which clinical laboratories are performing assays that utilize next-generation sequencing (NGS). Anticipating the issue of how to recognize NGS-based testing within the CPT coding system, the AMP Economic Affairs Committee initiated discussions in early 2012 that would, hopefully, enlighten the topic. AMP released its proposed coding scheme in March 2013 [2]. The AMA CPT Editorial Panel, working through its Molecular Pathology Advisory Group (MPAG) scheduled stakeholder meetings that would lead to a functional coding system for 2014.

A new group of codes known as genomic sequencing procedures (GSP) were adopted and added to the code set in 2015. These codes were designed to describe disease-based gene panels as well as larger whole-exome and whole-genome assays that were becoming more commonly used. Additionally, there were codes introduced to describe panel testing in the settings of cancer and certain inherited conditions. As an example, separate CPT codes were designated for panel-based testing for cardiomyopathy, intellectual disability, and cancer panels on solid tumors between 5 and 50 genes. Importantly, while many of these panels were expected to be performed on NGS platforms, the codes were deliberately left method agnostic to mirror the spirit of the remainder of the molecular pathology code set.

Another type of molecular pathology CPT code, termed the multianalyte assays with algorithmic analyses (MAAAs), was added along with GSP codes. MAAAs describe DNA, RNA, or protein-based assays which are performed and then analyzed by specific algorithms to arrive at a clinically useful result. Because of the algorithmic nature of these assays, they are generally (but not always) proprietary assays offered by a single laboratory provider. MAAA codes for assays which have established clinical utility and widespread adoption are Category I codes. A separate category of MAAA code, the administrative MAAA, was established to identify for tracking purposes MAAA assays which do not satisfy the requirements for Category I placement.

The development of GSPs and MAAA codes to describe NGS tests brings to the forefront questions that go far beyond simple reimbursement and call for serious introspection and debate about how and why any test should be used, ethical considerations in reviewing and making available unsolicited genetic information, and even questions regarding the basic principles of medical practice. The challenge in addressing NGS testing is to recognize unique laboratory and pathology services in a manner that is consistent with existing pathology-related services and codes. Consideration of the existing Tier 1 and Tier 2 molecular pathology codes emphasizes three elements that need to be accommodated: transparency, clinical utility, and professional work. In the context of NGS

tests coded by GSPs and MAAAs, all three of these elements depart from their traditional meanings in laboratory testing.

Whereas transparency is readily achieved for the simple Tier 1 "analyte-specific" codes, the definition blurs with the Tier 2 codes where the analyses are of many genetic alterations in a single gene whose commonality is an association, often to varying degrees, with specific clinical syndromes and scenarios. There may be no one specific "analyte," and the service that is requested is more appropriately described as a syndromically related genetic evaluation. This distinction takes on even greater import with GSPs where the genetic evaluation can be readily extended to involve multiple relevant genes for their contribution to specific clinical scenarios. The "analyte" is now more a clinical question than a biochemical entity.

The issue of clinical utility for simple tests, molecular or biochemical, is generally acknowledged in the descriptor, if not by the fact that a CPT code has been assigned to a specific assay. The clinical descriptor takes on more importance in the Tier 2 codes where, in many circumstances, the test itself is essentially defined by the clinical syndrome. GSPs are, in many circumstances, utilized in a similar manner to provide multigene evaluations relevant to specific clinical questions.

The unique capability of NGS methodologies to interrogate multiple target sequences simultaneously challenges the current mode of clinical laboratory testing. Typically, a single analytical test is ordered, the test is performed, a result is reported, a bill is generated, and the laboratory or physician is compensated for the work performed. The technical output of an NGS assay could generate data far in excess of what is needed to address a specific clinical question. The opportunity to "re-query" an existing NGS dataset defines a new kind of medical service that markedly deviates from traditional laboratory tests. A "re-query" could reflect a series of iterative inquiries to evaluate a complex phenotype, for example, or could be unrelated to the primary clinical question. An example of the latter might be an evaluation of drug-metabolizing gene variants for the purpose of selecting therapy and dosing some months after primary evaluation of a malignant tumor for diagnosis and classification. The technical work for the "re-query" would be different than for the initial sequencing, although costs of data storage, retrieval, editing, and quality assurance (QA) could still be substantial. The predominant work for the re-query would be professional evaluation of the data and interpretation in the context of the question being addressed. New codes have been added to the GSP code set to allow for re-evaluation of data from a clinical exome and genome.

With these considerations in mind, the AMP proposal on which the NGS codes are based sought to categorize potential uses of NGS tests, identifying usages that could complement or replace older methodologies and highlighting those unique features of NGS tests that herald novel medical services which will need novel CPT codes to adequately recognize the technical and professional services provided. Examples put forward in the AMP proposal are listed in Table 10.1. Some of these usages reflect some technical advantage or efficiency in addressing clinical problems already defined by the Tier 2 codes. Unless the newer technology generates novel clinically useful information, the existing Tier 2 codes would appear sufficient. Other usages of NGS, however, are categorically new tests. For some applications, the clinical question will define the extent of inquiry, and the services can be defined in those terms. Undoubtedly, broader genomic inquiries will emerge as NGS becomes more common and feasible. For those inquiries which have demonstrated clinical utility, a CPT coding scheme which follows the aforementioned principles will be able to accommodate their introduction into clinical service.

Table 10.1 Examples of clinical genomic sequencing assays (excerpted from [2])

Aneuploidy detection in circulating cell-free fetal DNA (chromosome 21 only or 21, 18, and 13)
Disorder-specific multigene evaluations for heritable disorders
Identification of rare genetic defects in individual patients
Multigene evaluation of a neoplasm for diagnostic, prognostic, and/or therapeutic decision-making
Clonality assessment in lymphoma
Whole-exome and whole-genome analysis of a neoplasm
Microbiome evaluations

PAMA

The **Protecting Access to Medicare Act of 2014** (H.R. 4302; Pub.L. 113–93) (PAMA) introduced changes into how all laboratory tests can be assigned codes and how CMS must determine reimbursement values for those services. This legislation defines a novel category of laboratory test, the Advanced Diagnostic Laboratory Test (ADLT) as a test *"that is offered and furnished only by a single laboratory and not sold for use by a laboratory other than the original developing laboratory (or a successor owner) and meets one of the following criteria:*

A. *The test is an analysis of multiple biomarkers of DNA, RNA, or proteins combined with a unique algorithm to yield a single patient-specific result.*
B. *The test is cleared or approved by the Food and Drug Administration.*
C. *The test meets other similar criteria established by the Secretary."*

ADLTs are thus distinguished from existing clinical laboratory diagnostic tests (CLDT). For tracking purposes, PAMA also requires the assignment of unique HCPCS codes for ADLTs and existing CDLTs that are cleared or approved

by the FDA if they have not already been assigned a unique HCPCS code and to publicly report the payment rate for the test. To accommodate the need for unique HCPCS codes as required by PAMA, AMA CPT created a separate code set to recognize proprietary laboratory assays, the PLA code set. Applications for PLA codes are reviewed quarterly with panel action also taking place on that same schedule with codes being published for use within several months of application.

The PLA codes are intended solely for tracking purposes and consequently are not reviewed by the CPT Editorial Panel with the same rigor as are applications for Category I CPT codes where issues for clinical validity, clinical utility, and widespread adoption are significant considerations in assigning a code. Because of this and the proprietary nature of the PLA-coded tests, developing coverage and reimbursement policies for these assays has been problematic for payers, since each PLA-coded test must be individually evaluated.

PAMA has significantly changed how CMS determines the payment for all laboratory testing services. As stated in the 2016 Final Rule, "the payment amount for a test on the CLFS furnished on or after January 1, 2018, will be equal to the weighted median of private payor rates determined for the test, based on the data of applicable laboratories that is collected during a specified data collection period and reported to CMS during a specified data reporting period. A subset of tests on the CLFS -- advanced diagnostic laboratory tests (ADLTs) -- will have different data collection, reporting, and payment policies associated with them as required by the statute." Once implemented, reductions in payment cannot be more than 10% per year for the first 3 years and no more than 15% in the subsequent 3 years.

A key variable in the implementation of PAMA has been the definition of what constitutes an "applicable laboratory." PAMA defines applicable laboratories as having the majority of their Medicare revenues paid under the CLFS or the physician fee schedule (PFS). CMS has determined that applicable laboratories include those identified by their National Provider Identifier (NPI), as having the majority of their Medicare revenues paid under the CLFS or the physician fee schedule (PFS). Responding to concerns about the administrative burden the reporting requirements would place on smaller laboratories, CMS also exempted from reporting low-volume laboratories with Medicare revenues of $12,500 or less under the CLFS during a data collection period. While it is unclear how much of an effect this exclusion will have on the "weighted median" benchmark determination, there remains considerable anxiety about the fiscal stability of molecular diagnostic laboratories which are generally smaller, academic- and hospital-based operations with relatively higher operating costs.

The first required PAMA reporting period took place in 2017 with proposed CLFS pricing issued in 2017. Subsequently, in late 2017 the American Clinical Laboratory Association (ACLA) brought suit against CMS claiming that CMS ignored congressional intent and instituted a highly flawed data reporting process which excluded more than 99% of laboratories from submitting relevant data.

Conclusions

The implementation of genomic testing in medicine is a major milestone in a journey that had its beginnings more than a half century ago with the discovery that DNA serves not only as the basis of heredity but is also the regulatory medium for cellular differentiation and regulation. It is tempting to regard such a technological advancement as revolutionary, but it is important to remember that the genetic underpinnings and our molecular understanding of disease processes are only one domain in the practice of medicine and our approach to the individual patient. It is likewise important to remember that current regulatory and reimbursement mechanisms have evolved to complement the practice of medicine. Although imperfect in some details, our current oversight mechanisms have functioned well in ensuring the quality and availability of laboratory testing and evaluation and, just as importantly, have served to promote constant innovation and improvement.

Progress on this front will demand a response from all parties including laboratories, professional and accrediting organizations, vendors and manufacturers, and regulatory agencies in order to adapt existing standards and practices to novel genomic technologies and clinical applications.

Reimbursement for genomic sequencing assays will demand a reassessment of the traditional understanding of laboratory testing, moving away from simple chemical analyses to syndromically defined genetic inquiries necessitating increasing amounts of professional work and involvement. Existing CPT structure can be modified to accommodate these new realities but will demand re-evaluation of existing concepts of laboratory testing and professional services related to genomic test interpretation and clinical usage. These challenges offer an opportunity to focus on those elements that have served us well and identify those which have not and to evolve new systems to accommodate these new technologies in a way that promotes better clinical care of each patient.

References

1. 42 CFR 493.1253 – Standard: Establishment and verification of performance specifications, and 42 CFR 493.1254 – Standard: Maintenance and function checks.
2. Proposal to address CPT coding for Genomic Sequencing Procedures, Association for Molecular Pathology Economic Affairs Committee, March 2013.

Patents and Proprietary Assays

11

Roger D. Klein

Introduction

In the diagnostic realm, patents on relationships between human gene variants and clinical phenotypes, as well as on the underlying genetic sequences themselves, have proven to be extremely controversial. Pathologists, geneticists, other laboratory professionals, as well as some patient and consumer organizations have criticized such patents for increasing test costs, decreasing innovation, reducing patient access, restricting patients' choices of providers and their access to second opinions, inhibiting clinical and basic research, and fostering the development of proprietary databases of medically significant genetic findings [1–3].

Author Michael Crichton joined the chorus of critics in his 2006 novel *Next*, going as far as to include an appendix to the book that exposed the "evils" of gene patents and advocated a ban on them, views he also expressed in *The New York Times* column [4]. In February 2007, Congressmen Xavier Becerra (D-Calif.) and David Weldon (R-Fla.) introduced "The Genomic Research and Accessibility Act" (HR 977), a bill that would have banned future patents on all nucleic acid sequences.

Conversely, proponents of gene patents have argued that these patents incentivize gene discovery, as well as investments in and commercialization of genetic tests. Gene patents, as it has been argued, benefit patients by encouraging discoveries of genetic relationships and the development and introduction of new assays that in the absence of patents would not have been brought to fruition.

This chapter chronicles the history of human gene patents, discusses arguments for and against gene patents, and presents key legal cases that impact on or directly address the validity and permissible scope of such patents. Finally, the implications of these recent legal developments for diagnostic testing are discussed.

R. D. Klein (✉)
Faculty Fellow, Center for Law, Science and Innovation, State University, Sandra Day O'Connor College of Law, Arizona State University, Tempe, AZ, USA
e-mail: roger@rogerdklein.com

Patent Overview

A US utility patent confers upon the patent holder the right to exclude others from making, using, selling, offering to sell, or importing an invention or a product made by a patented process, for 20 years from the filing date [5]. The basis for the US patent system is found in the Constitution, which in Article I, Section 8, Clause 8 states, "The Congress shall have Power…To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries;…" Patent exclusivity has historically been justified by the incentives it generates for inventors to create, commercialize, and disclose new inventions, the benefits from which will accrue to society at large.

The Congress enacted the first US patent laws in 1790. The Patent Act of 1790 was repealed and replaced in 1793, and the patent laws have subsequently been modified on numerous occasions. The basic structure of the current Patent Act was established in 1952, when the patent laws were reenacted in their entirety. Since passage of the 1952 Act, the patent laws have been amended several times, recently undergoing significant revisions by way of the America Invents Act of 2011.

Under US patent law, patentable inventions must be novel, nonobvious, and useful [6]. In addition, under "written description" and "enablement" requirements, a patent must describe the patented invention in what is termed its "specification," "in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same." Moreover, the specification must set forth the "best mode," in the mind of the inventor, of practicing the invention [7]. Within the specification, patent "claims" define the invention's features, establishing the boundaries of what is claimed, much as a real estate deed delineates the boundaries of a plot of land.

Patent applications are submitted to the United States Patent and Trademark Office (USPTO) where they are

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_11

rejected or allowed and issued. "Processes, machines, manufactures, and compositions of matter" can be patented [8], but patents may not be obtained on products of nature or, under the "natural phenomenon doctrine," "laws of nature, natural phenomena, and abstract ideas" [9].

Patent infringement, which encompasses the making, using, selling, offering to sell, or importing of a patented product or a product made by a patented process, can occur through direct infringement of the patent, [10] inducement of others to infringe the patent [11], or contributing to another's infringement of the patent [12]. For example, prior to recent US Supreme Court decisions, a laboratory could have been found to have directly infringed a gene patent if it tested for mutations in a patented gene, or variants claimed in a patented genotype–phenotype association.

In order to be found liable for inducing another to infringe a patent, a party must have actively, intentionally, and knowingly solicited or assisted another to infringe the patent, with the solicited individual or entity itself having directly infringed the patent. Thus, a laboratory that used educational materials to promote an offered test for a patented genetic association to physicians who then ordered the test, received the results, and thought about the association during the management of their patients could until recently have been found to have induced the direct infringement of the patent by the ordering physicians.

Finally, sale of a material component of a patented invention that has no substantial use other than as a component of the invention denotes contributory infringement. Applying this definition, the laboratory in the preceding example could also have been found liable for contributory infringement for providing testing for the patented genetic association.

History of Gene Patents

The legitimization of gene patents in the USA appears to have been an outgrowth of legal and political changes that were initiated in response to the economic dislocations of the late 1970s and early 1980s. During this period, the country was plagued by high unemployment, high inflation, and a decline in economic confidence. In response, the Congress took a number of steps to encourage the growth of domestic technology industries. Among the most significant of these were changes to the US patent system.

To maximize the economic value derived from our substantial federal investments in basic science research, the Congress in 1980 passed the Bayh–Dole Act, which encouraged universities to patent, and thereby commercialize, inventions arising out of government-sponsored research grants [13]. In the years subsequent to the passage of Bayh–Dole, federal financial commitments dedicated to biomedical research dramatically increased. National Institutes of Health funding of biomedical research ballooned from approximately $5 billion in the late 1970s to $26 billion in 2003 [14]. Because of these governmental actions, the number of patents assigned to universities increased from 264 in 1979 to 3291 in 2002 [15, 16].

In another important event, in 1980 the US Supreme Court ruled in *Diamond v. Chakrabarty* [17] that man-made, living organisms could be patented. In its decision, the Supreme Court urged a broad interpretation of patent eligibility, holding that "anything under the sun that is made by man," including living organisms, can be patented. Finally, in an effort to provide national uniformity and add greater certainty and expertise to the application of patent law, in 1982 the Congress created the Court of Appeals for the Federal Circuit (CAFC), with exclusive appellate jurisdiction for patent cases [18].

Since its inception, Federal Circuit decisions have affected the biotech sector significantly by generally expanding patent-eligible subject matter and strengthening the rights of patent holders relative to potential infringers. Many patents have since been issued on a range of biotech inventions, from transgenic mice and leukemia-derived cell lines to recombinant drugs and vaccines. Thousands of patents have also been awarded on human gene sequences, genetic variants, and, more recently, genotype–phenotype correlations [19].

The coalescence of the preceding events set the stage for the enormous growth of the US biotech industry. For example, from 1982 to 2002, the US Food and Drug Administration (FDA) approvals for biotech drugs and vaccines grew from 2 to 35. The number of US biotech companies expanded from 225 in 1977 to 1457 in 2001. Biotech employment mushroomed from 700 in 1980 to 191,000 in 2001. In addition, the industry's growth has created hundreds of thousands of jobs in related industries [20, 21].

It has been argued that in awarding gene patents, the US Patent and Trademark Office and the CAFC merely followed the Supreme Court's instruction in *Chakrabarty* to interpret patent eligibility broadly [22]. Importantly, post-Chakrabarty our patent system looked into chemical law precedents as a basis for awarding gene patents and treated DNA itself as a chemical despite its dual roles as a physical substance and a store of biological information. In *Amgen v. Chugai Pharmaceutical Co*., the CAFC wrote, "A gene is a chemical compound, albeit a complex one" [23].

Prior precedents in chemical law upheld the patenting of isolated, purified compounds such as aspirin, epinephrine, vitamin B12, and prostaglandins [24–27]. The Patent Office applied these legal precedents to isolated DNA sequences. This direct superimposition of chemical law precedents to DNA permitted circumvention of the "product of nature," doctrine's long-standing prohibition against patenting natural substances, and allowed for the issuance of patents on isolated, purified human genes.

Evidence for and Against

To practitioners in the field, it appears obvious that gene patents have significantly inhibited the provision of genetic testing services [28]. Many providers have discontinued or have been prevented from providing molecular genetic testing for inherited breast and ovarian cancer, Duchenne muscular dystrophy, spinocerebellar ataxias, genes causing long QT syndrome, as well as the *FLT3* internal tandem duplication in patients with intermediate-risk acute myelogenous leukemia (AML), the *JAK2* p.Val617Phe (better known as V617F) variant in myeloproliferative neoplasms, and many others.

Intuitively, one would expect that monopolistic behavior would lead to increased prices and decreased patient access to testing. Although there is some support for this contention, true markets do not exist for health-care services in the USA because of the roles of third-party insurance and government as major payers. Further, prices are difficult to obtain, which makes comparisons of actual charges difficult [29]. However, for single gene discoveries and their subsequent introduction into clinical testing, the notion that gene patents have been a necessary stimulus seems dubious. In general, rather than encouraging the introduction of new tests, gene patents have tended to cause laboratories to discontinue tests they had already been performing.

Most human genes on which clinical testing has been performed have been discovered by university faculty members. For these professors, publication and solicitation of grants based upon their discoveries are necessary for academic promotion and even professional "survival," rendering patents a superfluous incentive. Inherited diseases are rare, offering very limited market potential. Yet many such genes have been discovered despite an apparent lack of significant commercial or monetary potential because of the research interests of the investigator.

Lastly, it is usually relatively inexpensive to design, develop, validate, and perform genetic tests using justifiably patented tools and techniques. This is in contrast to pharmaceuticals, which require costly, extensive periods of discovery and testing and must undergo an expensive approval process, features that support the need for robust patent protection [30].

Although the preceding discussion regarding the adverse effects of gene patents on the introduction of new molecular genetic assays holds true for most assays, the relative impact of gene-related patents on some tests based on multi-analyte gene expression profiling seems less clear. A central feature of these assays is a reliance on proprietary mathematical algorithms that proponents claim allow for correlation of the expression patterns of, for example, multiple mRNAs, sometimes in combination with other parameters, with relevant clinical characteristics such as diagnosis, prognosis, or response to drug therapy. A variety of such tests are oncology oriented.

Implementation of these types of expression assays typically requires prolonged and potentially expensive periods of study in order to establish sufficient clinical validity and utility to justify their use. At the time of this writing, few such assays have crossed this threshold and are supported by high-level evidence of this nature. In addition, expression profiling tests may in the future require FDA approval or clearance, increasing development costs. Therefore, exclusivity may be necessary to attract sufficient funding to advance those assays that ultimately prove worthy in clinical care. Arguably, some inventive work has occurred in such assays through establishment of the gene "signature." Moreover, although the assays rely on natural, biological associations, they also can generally be "invented around" and therefore pose less risk of tying up essential natural phenomena. Interestingly, although patent protection may be essential to bring assays of this type to market, individual gene or nucleic acid patents could otherwise obstruct their development by restricting use of the genes available for inclusion in the test.

Key Legal Cases

Bilski v. Kappos

In *Bilski v. Kappos*, the USPTO rejected a patent application in which the submitted claims covered a process of hedging commodities against price fluctuations. The method involved contracting to purchase commodities at fixed prices from sellers who wanted to hedge against a fall in prices while contracting to sell commodities at fixed prices to consumers who were hedging against a rise in prices. On appeal the CAFC upheld the USPTO's decision denying a patent [31].

Prior to *Bilski*, the rule at the CAFC was that patentable processes had to produce a "useful, concrete, and tangible result." In *Bilski*, the CAFC articulated a new standard, its "machine or transformation test." The CAFC sitting as the entire court, termed *en banc*, held that patentable processes must be tied to a particular machine or apparatus or must transform a particular article into a different state or thing and that this transformation must be central to the purpose of the process. Bilski's hedging process, the CAFC ruled, failed to meet the machine or transformation test and therefore was ineligible to receive a patent [31].

The US Supreme Court affirmed the lack of patent eligibility of the claimed hedging process but refined the CAFC's reasoning [32]. Although the machine or transformation test may be a "useful and important clue or investigative tool" for deciding whether some processes are patent-eligible inventions under 35 USC Section 101 of the Patent Act, the Supreme Court held that it is not the sole test of patent eligibility by which such processes are to be evaluated.

Some gene patent claims that assert ownership over genotype–phenotype associations have framed these natural laws as a series of steps, thus characterizing them as processes. The *Bilski* decision influences the framework under which the patent eligibility of process claims is evaluated. Therefore, although it was narrowly crafted to the specific set of business facts before the Supreme Court in the case, *Bilski* has relevance for the assessment of the patent eligibility of process claims involving human genes.

KSR Int'l Co. v. Teleflex Inc.

In the 2007 case of *KSR Int'l Co. v. Teleflex Inc*., a unanimous US Supreme Court relaxed the legal standards for determining patent obviousness under Section 103 of the Patent Act [33]. KSR added a sensor to one of its previously designed automobile throttle pedals. Teleflex then sued KSR for infringement of a patent that claimed the combination of an adjustable automobile accelerator pedal and an electronic sensor. In response, KSR argued that the patent was invalid because its subject matter was obvious. The district court agreed with KSR and ruled that the accelerator–sensor combination was obvious. The CAFC reversed the lower court decision.

In upholding the patent, the CAFC applied what was termed its "teaching, suggestion, or motivation" test (TSM test) for obvious determinations. Under this test, a patent claim could only be found obvious if there was "some motivation or suggestion to combine the prior art teachings" present in the previous body of knowledge in the field, the nature of the problem the solution sought to solve, or the knowledge of a person who possessed ordinary skill in the field. That an approach was "obvious to try," the CAFC wrote, had under previous precedents long been irrelevant.

The Supreme Court rejected the CAFC's rigid, formalistic, and narrow process for obviousness determination in favor of a more "expansive and flexible approach," ruling that the throttle pedal–sensor combination had been obvious at the time of the patent application. Importantly, the Supreme Court held that obviousness to try a problem-solving approach can in fact render a patent obvious under circumstances in which there is a demonstrated need for a discovery and a finite number of identified, predictable solutions to the problem. The Court wrote:

> When there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to try might show that it was obvious under §103. [33]

Many patented genes were initially mapped to a chromosomal region prior to their discovery. In addition, many medically important genes are involved in sequential biochemical pathways, in which disease-related perturbations were known before the identification of particular genetic associations. Therefore, it would have been obvious to look for variants in these genes among a finite number of genes during genetic studies of the relevant disorder. Finally, cDNA sequences are directly derived from the exon sequences of native genes and can also be deduced from the amino acid sequences of the proteins for which they encode, likely rendering significant numbers of patent claims on cDNA obvious. In light of the preceding, the Supreme Court's decision in *KSR* potentially affects the validity of many gene-related patents.

In Re Kubin

In 2009, the case of *In Re Kubin* provided the CAFC with an early opportunity to apply the obviousness paradigm the Supreme Court set forth in *KSR* [34]. In *Kubin*, the USPTO refused to award a patent on the full gene and cDNA sequences of the Natural Killer Cell Activation Inducing Ligand (NAIL), a natural killer (NK) cell surface receptor that plays a role in cellular activation. The Patent Office rejected the application both on obviousness grounds under 35 USC Section 103 and for inadequate written description under 35 USC Section 112.

On appeal, the CAFC affirmed the Patent Office's decision, agreeing that delineation of the NAIL gene sequences was obvious in light of the prior art, which included knowledge of the existence of the NAIL protein, but not its protein sequence. Citing the case of *Graham v. John Deere Co*. [35], the CAFC reviewed the factual inquiries necessary for a legal finding of obviousness. These, the CAFC wrote, include "(1) the scope and content of the prior art; (2) the differences between the prior art and the claims at issue; (3) the level of ordinary skill in the art at the time the invention was made; and (4) objective evidence of non-obviousness, if any." Under the aforesaid criteria, the CAFC found that the NAIL gene sequences were obvious.

Applying the Supreme Court's *KSR* decision, the CAFC reversed one of its previous DNA cases, *In re Deuel*, in which it had held that "obvious to try" an approach was an inappropriate test for obviousness [36]. In *In re Deuel*, the CAFC had reversed the Patent Office's conclusion that the existence of a prior art reference describing a method of gene cloning together with the partial amino acid sequence of the protein rendered the underlying cDNA sequence obvious. Instead, the *In re Deuel* Court found that knowledge of the protein sequence was itself insufficient to generate the sequence of the underlying cDNA and, therefore, that the sequence was

nonobvious. Further, the CAFC eliminated "obviousness to try" as a potential determinant of obviousness. The Court wrote:

> [T]he existence of a general method of isolating cDNA or DNA molecules is essentially irrelevant to the question whether the specific molecules themselves would have been obvious, in the absence of other prior art that suggests the claimed DNAs… 'Obvious to try' has long been held not to constitute obviousness. A general incentive does not make obvious a particular result, nor does the existence of techniques by which those efforts can be carried out.

In light of the Supreme Court's prior rejection of the CAFC's "obvious to try" doctrine in *KSR*, the CAFC in *Kubin* found that the NAIL cDNA and full gene sequences were obvious to try and therefore obvious under Section 103. The CAFC stated:

> In light of the concrete, specific teachings of Sambrook and Valinente, artisans in this field, as found by the Board in its expertise, had every motivation to seek and every reasonable expectation of success in achieving the sequence of the claimed invention. In that sense, the claimed invention was reasonably expected in light of the prior art and 'obvious to try.'

Mayo v. Prometheus

In *Mayo Collaborative Services v. Prometheus Laboratories, Inc*. [37], Prometheus Labs sued Mayo Clinic in the District Court for the Southern District of California for infringement of two patents covering the post-administration correlation of blood levels of the thiopurine metabolites 6-methyl mercaptopurine and 6-thioguanine with thiopurine efficacy and related side effects. Both patents were written in the form of stepwise processes, the relevant claims of which included the generic steps of (1) administering the drug, (2) measuring the metabolite levels, and (3) describing the metabolite concentrations above and below which are associated with an increased likelihood of toxicities or lack of efficacy, respectively, and then informing the ordering physician of the potential need to decrease or increase the drug dose. Thus, the patent in effect claims the reference range for thiopurine drugs. Mayo Clinic had been utilizing Prometheus' test but in 2004 announced that it was going to offer its own internally developed test for metabolites. Prometheus sued Mayo for patent infringement. Mayo Clinic argued that Prometheus' patents covered unpatentable natural phenomena and were therefore invalid as a matter of law under Section 101 of the Patent Act. The District Court agreed with Mayo and ruled that Prometheus' patents were invalid. The CAFC reversed the District Court, instead holding that the patents claimed methods of treatment. Moreover, the CAFC held, the in vivo metabolism of thiopurine agents constituted transformations of matter under that Court's "machine or transformation test," a test which was discussed earlier in this chapter in connection with *Bilski v. Kappos*. In *Bilski*, the Supreme Court clarified that although the "machine or transformation test" is an important and useful clue to patent eligibility, it is not a definitive test for it.

Mayo appealed to the Supreme Court, which, following its decision in *Bilski*, accepted *Mayo v. Prometheus* and immediately returned it to the CAFC for reconsideration. On remand, the CAFC reaffirmed its earlier decision reversing the District Court's determination that Prometheus' patents were invalid.

Mayo again appealed to the Supreme Court, and the Court accepted the case. In a nine-to-zero decision, the Supreme Court held that the processes claimed in Prometheus' patents were not patent eligible. The Court recognized that an unpatentable biological correlation lay at the center of Prometheus' patents. In order to receive a process patent that purports to claim an application of a natural law, the Court noted, sufficient inventive effort must be added to the natural law so as to ensure that the patent is "significantly more than a patent upon the natural law itself." Moreover, the Court emphasized that the addition of routine steps cannot convert the natural law into a patentable process. As the Court explained, "If a law of nature is not patentable, then neither is a process of reciting a law of nature, unless that process has additional features that provide practical assurance that the process is more than a drafting effort designed to monopolize the law of nature itself." The Court succinctly summarized: "[T]o transform an unpatentable law of nature into a patent-eligible application of such a law, one must do more than simply state the law of nature while adding the words 'apply it'."

The unanimity, clarity, and strength of the Supreme Court's opinion in support of this ruling standing alone imply that analogous patents covering genotype–phenotype associations are also invalid. This conclusion is bolstered by the Court of Appeals' affirmance of the District Court finding of invalidity of Myriad Genetics' sequence comparison claims in *Association for Molecular Pathology v. Myriad Genetics* discussed subsequently [38] and is reinforced and further strengthened by the Supreme Court's decision finding human DNA patent ineligible in the *Myriad* case. *Mayo v. Prometheus* and *AMP v. Myriad* have important implications for genomic analyses performed using next-generation sequencing and for genetic testing as a whole.

Association for Molecular Pathology v. Myriad Genetics, Inc.

Finally, in *Association for Molecular Pathology v. Myriad Genetics, Inc*. [38], a lawsuit sponsored by the American Civil Liberties Union, various medical and professional societies,

health-care providers, and breast cancer patients sued Myriad Genetics, the University of Utah Research Foundation, and the USPTO seeking to invalidate key composition of matter and process claims of patents covering the wild-type and mutated sequences of the *BRCA1* and *BRCA2* genes, as well as correlations between variants in those sequences and the predisposition to breast and ovarian cancer.

In total, the plaintiffs challenged 15 claims contained in 7 patents. They argued that these patent claims were invalid under Section 101 of the Patent Act of 1952, and unconstitutional under Article I, Section 8, Clause 8 and the First and Fourteenth Amendments, because they asserted ownership of natural products, natural laws, natural phenomena, abstract ideas, and basic human knowledge or thought. In response, Myriad argued that its patents claimed DNA sequences that were identical to those in the human body, but because the sequences were isolated from the body they constituted human inventions.

Myriad also asserted that its patented associations between variants in *BRCA1* and *BRCA2* and the hereditary predisposition to breast and ovarian cancers were actually diagnostic methods involving sequence comparisons, not patents on the biological relationships themselves. The District Court distilled the lawsuit into a single fundamental question, "Are isolated human genes and the comparison of their sequences patentable?"

The Judge, Robert W. Sweet, emphasized the centrality of knowledge of molecular biology to the proper disposition of the case, as well as the importance of any potentially relevant additional inventive steps [39]. On page 27 of its opinion, Judge Sweet wrote: "An understanding of the basics of molecular biology is required to resolve the issues presented and to provide the requisite insight into the fundamentals of the genome, that is, the nature which is at the heart of the dispute between the parties…." The Court devoted the next 19 pages of the opinion to a thorough review of generally accepted principles of molecular biology. It concluded the section with the recognition that some inventive work was involved in the initial sequencing of the *BRCA1* and *BRCA2* genes stating: "However, because sequencing requires knowledge of the sequence of a portion of the target sequence, some ingenuity and effort is required for the initial sequencing of a target DNA." Expert declarations by Mark A. Kay, M.D., Ph.D., and this chapter's author helped the Court sort out the extent, significance, and relevance of this work to the validity of the claims at issue.

In the pertinent sections of their dueling declarations, Kay attempted to emphasize the inventive aspects of sequencing a newly discovered product, while Klein delineated the breadth of the patents and the natural products and laws they claimed; the routine, insubstantial, and non-transformative steps involved in performing genetic testing; and the relationship of genetic testing to other forms of medical diagnosis. In paragraph 183 of his declaration, Dr. Kay described the steps involved in sequencing a newly identified product:

> To sequence a particular target, at least part of the target sequence must be known to design a suitable primer. The initial sequencing of a target sequence requires ingenuity far beyond the mere application of routine laboratory techniques and usually involves a significant amount of trial and error. A primer is used to initiate the sequencing reaction at the desired location of a target sequence. A primer is an artificial DNA fragment, usually between 15 and 30 nucleotides long, that binds specifically to the target nucleotide sequence. The nucleotide sequence of the primer is complementary to the target sequence such that the bases of the primer and the bases of the target sequence bind to each other.

By contrast, in paragraphs 32–34, Dr. Klein wrote:

> The claims at issue in this case do not cover diagnostic tools or actual methods used in genetic testing. Nor are they analogous to patents on medical instruments. Rather they claim DNA sequences which are themselves the subject of medical inquiry. Further, they incorporate generic steps in an effort to describe the biological relationships between mutations in *BRCA1* and *BRCA2* and the predisposition to cancer in the abstract patent language of a 'process.' However, the key steps in genetic testing, DNA extraction, amplification, and sequencing can now be performed using routine, automated methods. Nevertheless, the defendants claim the exclusive right to read and compare *BRCA1* and *BRCA2* sequences irrespective of the method used, whether that method is in existence now or will be invented in the future. Correlating a patient's gene sequence with the predisposition to disease is simply another form of medical diagnosis, similar to correlating elevations in blood glucose with diabetes, a heart murmur with mitral stenosis, or the patterns on a pathology slide with a particular type of tumor and its optimal therapy. Automated sequencers reveal the sequence of the nucleotides visually in what is called a chromatogram. That chromatogram is then "read" (by software and visual inspection) to determine a patient's gene sequence. DNA extraction and sequencing are not transformative activities. Rather extraction is a routine, non-substantial preparatory step that allows for PCR amplification and sequencing. Sequencing is an automated procedure. DNA extraction, PCR, and sequencing do not involve transformations that are central to the purpose of the process of reading a patient's gene sequence. Unlike "tanning, dyeing, making waterproof cloth, vulcanizing India rubber, or smelting ores," which are performed for the purpose of physically transforming substances so as to create what are essentially new materials for their own sake, the purpose of genetic testing is solely to read the sequence of the DNA, not to transform it into something else. Only in this way can the patient and her physician learn whether a medically relevant mutation is present in her body.

On March 29, 2010, in a landmark decision, the District Court held that the composition of matter claims on the *BRCA1* and *BRCA2* gene sequences and their cDNAs and the process claims covering the correlations between mutations in *BRCA1* and *BRCA2* and the predisposition to breast cancer and ovarian cancer are invalid as a matter of law. In evaluating the composition of matter claims on the isolated gene sequences, the Court emphasized the unique informa-

tional characteristics contained in the DNA sequence, and the preservation of that native sequence in isolated DNA, stating "Because the claimed isolated DNA is not markedly different from native DNA as it exists in nature it constitutes unpatentable subject matter under 35 U.S.C. section 101." Similarly, the Court found comparison claims of known wild-type and patient sequences for diagnosis, claims that in effect asserted ownership over the biological relationships between *BRCA1* and *BRCA2* mutations and the predisposition to breast cancer, invalid as merely claiming abstract mental processes.

On appeal, the CAFC on July 29, 2011, in a two-to-one decision reversed the District Court, holding that isolated human gene sequences are patent eligible. However, the CAFC upheld the lower court's ruling that Myriad's sequence comparison claims were invalid. The plaintiffs appealed the case to the Supreme Court.

Immediately after deciding *Mayo v. Prometheus*, the Supreme Court accepted *AMP v. Myriad*, threw out the CAFC's decision, and sent the case back to the Court of Appeals for further consideration in light of its decision in *Mayo*. After rebriefing of the case and a second round of oral arguments, the CAFC again held two-to-one decision that isolated human genes are patent eligible on the grounds that they represent new compositions of matter that do not exist in nature.

As in the CAFC's previous decision in the case, each judge wrote a separate opinion. All three judges agreed that *BRCA1* and *BRCA2* cDNA should be patent eligible, reasoning that cDNA is not naturally occurring and is made by man. The central disagreement among these judges was whether separating human DNA from its chromosome and other cellular constituents renders it a patent-eligible invention.

The two judges who determined that human DNA is patent eligible came to the same conclusion using different reasoning. One judge, who authored what was nominally the primary opinion for the Court, opined that because separating a gene from its chromosome involves breaking covalent bonds, a DNA sequence removed from its natural environment is a new chemical. Another judge relied at least in part on the past practice of the USPTO in granting such patents, and the reliance of companies and inventors on that practice. This judge said she may have voted differently had the question come before her on a "blank canvas."

The dissenting judge wrote that the breaking of covalent bonds alone did not create a new molecule and was not determinative of patent eligibility. Rather, he concluded that the genes' DNA sequences are identical whether the genes are within or outside the body, and because of this that these DNAs are fundamentally the same molecule, irrespective of location. For the dissenting judge, the importance of the sequence of nucleotides in the DNA molecules substantially outweighed the importance of any chemical differences between the DNA in the body and DNA isolated from it.

However, the CAFC ultimately chose to disregard the constancy of the gene's most fundamental and relevant property, its coding sequence. On behalf of the Court Judge Alan Lourie wrote: "The isolated DNA molecules before us are not found in nature. They are obtained in the laboratory and are man-made, the product of human ingenuity." Judge Lourie maintained that native and isolated gene sequences have distinct chemical structures and identities because the native genes have been separated from associated proteins and the chromosomes on which they naturally reside, either through the cleaving of covalent bonds or by synthesis. In addition, the CAFC again held that Myriad's sequence comparison claims were invalid. The plaintiffs once more appealed the case to the Supreme Court.

On June 13, 2013, in an historic nine-to-zero decision authored by Justice Clarence Thomas, the Supreme Court held that naturally occurring DNA sequences are "products of nature" that are not patent eligible. The court acknowledged Myriad's contribution to the field but noted that its discoveries were limited to identifying the precise location and sequence of the *BRCA1* and *BRCA2* genes. The Court stated: "In this case…Myriad did not create anything. To be sure, it found an important and useful gene, but separating that gene from its surrounding genetic material is not an act of invention." Moreover, the Court referred back to Myriad's patent claims, which themselves confirmed that the fundamental essence of DNA lies in its information content.

> "Myriad's claims," the Court wrote, "are simply not expressed in terms of chemical composition, nor do they rely in any way on the on the chemical changes that result from the isolation of a particular section of DNA. Instead, the claims understandably focus on the genetic information encoded in the *BRCA1* and *BRCA2* genes. If the patents depended upon the creation of a unique molecule, then a would-be infringer could arguably avoid at least Myriad's patent claims on entire genes…by isolating a DNA sequence that included both the *BRCA1* or *BRCA2* gene and one additional nucleotide pair. Such a molecule would not be chemically identical to the molecule 'invented' by Myriad. But Myriad obviously would resist that outcome because its claim is concerned primarily with the information contained in the genetic sequence, not with the specific chemical composition of a particular molecule."

Finally, the Court did rule that cDNA is patent eligible because it is not naturally occurring. However, patent eligibility, as the Court pointed out in a footnote, does not necessarily equate to patentability under other sections of the Patent Act that this decision did not address. Moreover, because cDNA is not essential for the performance of most genetic testing, the ruling that cDNA is patent eligible is unlikely to have a significant impact on molecular genetic testing going forward.

Implication of Recent Court Decisions for Genetic Testing

In two recent decisions relevant to genetic testing, both unanimous, the Supreme Court reaffirmed its long-standing prohibitions on patenting natural laws and products of nature. In *Mayo*, the Court was clear that characterizing a biological association as a process does not, without adding a truly inventive step, convert the association into a patent-eligible application of a natural law. *Mayo* was an extremely important decision, which seemingly means that method patents that attempt to claim associations between genetic variants and clinical phenotypes are invalid. In *Association for Molecular Pathology*, the Supreme Court found that naturally occurring human DNA sequences are not patentable, rendering patents on human genes invalid. When read together these two cases appear to have removed the intellectual property barriers associated with testing for genetic mutations and relationships to clinical phenotypes, whether testing is for identification of predisposition to disease, therapeutic responsiveness, medicinal side effects, or tumor behavior. Thus, the Supreme Court has helped facilitate the introduction of large-scale sequencing into clinical practice and has thereby encouraged the advancement, development, and implementation of personalized medicine.

References

1. Klein RD. Gene patents and genetic testing in the United States. Nat Biotechnol. 2007;25:989–90.
2. Klein RD. Legal developments and practical implications of gene patenting on targeted drug discovery and development. Clin Pharmacol Ther. 2010;87:633–5.
3. Cook-Deegan R, Conley JM, Evans JP, Vorhaus D. The next controversy in genetic testing: clinical data as trade secrets? Eur J Hum Genet. 2013;21:585–8.
4. Michael C. Patenting Life, N.Y. TIMES, 13 Feb 2007, at A23. http://www.nytimes.com/2007/02/13/opinion/13crichton.html.
5. 35 U.S.C. § 154(a) (2012).
6. 35 U.S.C. §§ 101–103 (2012).
7. 35 U.S.C. § 112 (2012).
8. 35 U.S.C. § 101 (2012).
9. Diamond v. Diehr, 450 U.S. 175, 185 (1981).
10. 35 U.S.C. § 271(a) (2012).
11. 35 U.S.C. § 271(b) (2012).
12. 35 U.S.C. § 271(c) (2012).
13. 35 USC §§ 200–212 (2012).
14. Moses H, Dorsey ER, Matheson DHM, Their SO. Financial anatomy of biomedical research. JAMA. 2005;294:1333–42.
15. Rai AK, Eisenberg RS. Bayh-Dole reform and the progress of medicine. Law Contemp Probs. 2003;66:289–314.
16. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ/asgn/table_1_2005.htm.
17. Diamond v. Chakrabarty, 447 US 303 (1980).
18. Federal Courts Improvement Act of 1982, 96 Stat. 25, codified as 28 USC 1295 (2012).
19. Caulfield T, Cook-Deegan RM, Kieff FS, Walsh JP. Nat Biotechnol. 2006;24:1091–4.
20. http://www.bio.org/ataglance/bio/.
21. Lee SB, Wolfe LB. Biotechnology industry. In: Encyclopaedia of occupational health and safety. 4th ed: International Labour Organization. http://iloencyclopaedia.org/
22. Brief for Respondent, Association for Molecular Pathology v. Myriad Genetics, No. 12–398, 569 US: (7 Mar 2013); U.S. patent and trademark office utility guidelines, 66 Fed. Reg. 1092 (5 Jan 2001).
23. Amgen v. Chugai Pharmaceutical Co., 927 F.2d 1200 (1990), cert. denied, 502 US 856 (1991).
24. Kuehmsted v. Farbenfabriken, 179 F. 701 (7th Cir. 1910), cert. denied, 220 US 622 (1911) (acetyl salicylic acid).
25. Parke-Davis & Co. v. H.K. Mulford & Co., 189 F. 95 (SDNY 1911), aff'd, 196 F. 496 (2d Cir. 1912) (epinephrine).
26. Merck & Co. v. Olin Mathieson Chemical Corp., 253 F.2d 156 (4th Cir. 1958) (Vitamin B12).
27. In re Bergstrom, 427 F.2d 1394 (CCPA 1970) (PGE, PGF).
28. Cho MK, Illangasekare S, Weaver MA, Leonard DGB, Merz JF. Effects of patents and licenses on the provision of clinical genetic testing services. J Mol Diagn. 2003;5:3–8.
29. Secretary's Advisory Committee on Genetics, Health, and Society. Gene patents and licensing practices and their impact on patient access to genetic tests. 2010. http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_patents_report_2010.pdf.
30. Bessen J, Meurer MJ. Patent failure: how judges, bureaucrats, and lawyers Put innovators at risk. Princeton, NJ: Princeton University Press; 2008.
31. Bilski v. Kappos, 545 F.3d 943 (Fed. Cir 2008) (en banc).
32. Bilski v. Kappos, 130 S.Ct. 3218 (2010).
33. KSR Int'l Co. v. Teleflex Inc., 550 US 398 (2007).
34. In re Kubin, 561 F.3d 1351 (Fed. Cir. 2009).
35. Graham v. John Deere Co., 383 U.S. 1 (1966).
36. In Re Deuel, 51 F.3d 1552 (Fed. Cir. 1995).
37. Mayo Collaborative Services v. Prometheus Laboratories, Inc., 132 S.Ct. 1289 (2012).
38. Association for Molecular Pathology v. Myriad Genetics, Inc., No. 12-398, 569 US (2013).
39. Association for Molecular Pathology v. United States Patent and Trademark Office, No. 09-4515 (SDNY filed 12 May 2009).

Ethical Issues in Clinical Genetics and Genomics

12

Henry T. Greely

Introduction

Genetic testing has been clinically available for over 50 years and has generated ethical and legal questions for at least that long. The methods for genetic testing have changed dramatically over the years, and the details of the ethical and legal issues have changed dramatically as well, but the basic problems have not [1–3]. Today we are on the edge of clinical whole-genome sequencing ("WGS") technologies. As we move from "retail" or targeted tests, few in number and done only a few at a time, to "wholesale" or broadband tests, where hundreds of different important results may appear from one test, the old types of ethical and legal issues will arise, but their implications will change enormously. These implications are just beginning to be discussed [4–11].

The ethical and legal issues that have arisen with contemporary genetic testing can be grouped into five rough categories: the decision to test, accuracy, communicating results, direct-to-consumer ("DTC") testing, and "other concerns." This chapter will look first at those issues under current targeted testing. After a brief section on the existing beginnings of broadband genomic testing, the chapter will then reexamine those five areas as they are likely to appear in the coming whole-genome sequence world where today's students will spend the vast majority of their professional careers—and their lives as patients, family members, and citizens.

"Targeted" Genetic Testing: Ethical and Legal Issues

Most existing genetic tests have been "targeted." The tests are for one or a handful of genetic or chromosomal variants known to be strongly associated with particular diseases. Sometimes the patients suggest the genetic test to their doctors, but typically the doctors, listening to their patients' concerns and family histories and after completion of a clinical evaluation and examination, recommend the test. Either way, concrete risks lead to specific tests for particular genetic variants.

This is still largely the world of clinical genetics. Pregnant couples seek information from obstetricians about Down syndrome, about genetic diseases known to run in their families, or about conditions that carrier screening has shown they might pass down to their children. Parents whose children have various abnormal conditions ask pediatricians or geneticists to confirm tentative diagnoses through genetic tests. Adults with symptoms of genetic diseases, or who learn of strong family genetic risks, seek testing. In these cases, the genetic tests have in the past examined only one or a handful of genetic or chromosomal variants. These tests give rise to difficult ethical and legal problems, today as in the past five decades.

Deciding to Test: Medical Appropriateness and Informed Consent

The first ethical question is whether the test should be done at all. Medically inappropriate or unconsented tests are generally unethical. In the targeted context, these decisions are made about a specific test, for one or a few genes linked to one or a handful of conditions.

A test might be inappropriate because it is a good test but not for this patient. Thus, a woman with an extensive family history that does not include breast or ovarian cancer normally should not be offered *BRCA* testing. Sometimes the test might be inappropriate, because it does not work—or, perhaps, has not been proven to work. Normally one would not want to order a genetic test based on, say, one small and unreplicated study showing a weak association between a genetic variant and susceptibility to a disease.

Some argue that tests are different—they are "only" information and do not have the possible harms and costs

H. T. Greely (✉)
Center for Law and the Biosciences, Stanford University, Stanford, CA, USA
e-mail: hgreely@stanford.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_12

of "actual" interventions. This is clearly wrong. For the patients, inaccurate test results, either because of a bad test or an inaccurate result from an otherwise good test, will cause unnecessary anxiety or false reassurance, as well as the possibility of a cascade of subsequent tests and interventions. For the medical system, the costs of unnecessary tests, as well as unnecessary or inappropriate follow-ups, lead to the waste of resources that might have been put to better use. Inappropriate tests have *only* risks and costs with no balancing benefits.

The fact that a test is medically appropriate, however, does not guarantee that it should be ordered. Although there are a few exceptions, tests, like other medical interventions, usually require the informed consent of the patient.

One public health exception in genetics applies to newborns. Every state provides some neonatal genetic screening (and, where appropriate, follow-up testing); only two states require parental consent [12, 13]. Neonatal screening looks for serious disease where early detection can make a huge difference in the child's life. Phenylketonuria (PKU) is the canonical example. The roughly 1 in 50,000 American children born with this autosomal recessive genetic disorder *will* suffer from severe intellectual disabilities, unless their condition is detected early and they are put on a stringent diet, in which case their development is close to normal. If the issue is saving the brains of infants, the public health imperative trumps the need for informed consent. American states began to require neonatal genetic testing in the 1970s, starting with PKU and a handful of similar diseases. Those requirements expanded dramatically in the mid-2000s, and now most states require mandatory screening for 30–50 genetic conditions (not all, perhaps, appropriately) [12, 13].

Apart from these kinds of exceptional cases, genetic testing, like other medical interventions, requires informed consent. In general the law requires that the patient be told enough about the risks and benefits of the proposed intervention—and its reasonable alternatives—to be able to make an informed and intelligent decision. In determining what information suffices, American states take two different approaches. In most states, healthcare providers are required to provide the information that similarly situated, reasonable providers would give in those circumstances. This physician-centered consent standard is similar to the general standard for medical malpractice. The other approach centers on patients. It requires providers to give the patient the information that might make a difference to a reasonable patient. Neither of these standards is easy to apply to real cases, although guidelines or consensus statements from professional organizations can be powerful evidence for what a reasonable physician would do.

In genetic testing, the issues of informed consent revolve around information about the accuracy of the test (both analytic validity and clinical validity) as well as the potential benefits and risks to the patient. It is important to remember that, legally and ethically, informed consent is not a signature on a form, but a *process* of informing the patient about the procedure, its benefits, and risks, giving the patient a chance to ask questions, and ascertaining the patient's decision. A signature is some evidence that the process took place, but performing an adequate process is crucial.

Test Accuracy

An inaccurate test is both an ethical and legal problem. Regulatory questions around accuracy are addressed in other chapters of this book, but a few points might be usefully made here.

Unites States Federal Food, Drug, and Cosmetic Act defines medical tests as medical devices. If a test is required to be shown safe and effective, the FDA looks not just at its analytic validity but also at its clinical validity. The FDA, however, has not generally regulated laboratory tests done in licensed clinical laboratories. It takes the position that it has the authority to do so but that it is exercising its discretion not to regulate these "laboratory-developed tests" ("LDTs"). The FDA says it has not regulated clinical laboratories because they are under physicians' control and are regulated by the states and by the federal Clinical Laboratory Improvements Amendments Act ("CLIA"), with the College of American Pathologists playing a substantial role through its accreditation procedures [14]. These regulatory schemes, however, look mainly at the analytic validity of the test (how well it was done?) not its clinical validity (is it clinically accurate?). That second, but very important, question has rarely subject to regulatory review, by the FDA or anyone else. Two exceptions are worth noting. First, the FDA does regulate tests that are sold as "kits," either directly to consumers or to physicians who in turn use the kit. Second, the state regulatory agencies may have the power to examine clinical utility, which at least one state, New York, has asserted.

Recent years have brought the signs of possible change. On June 4, 2013, the American Clinical Laboratory Association (ACLA) filed a "citizen's petition" with the FDA arguing that the agency does not, in fact, have statutory authority to regulate LDTs, which it calls "the practice of medicine." [14, 15]. In October 2014, the FDA released the "draft guidances," suggesting ways that it would like to regulate LDTs and calling for some assessment of the clinical validity of each of them, phased in over 9 years. (At the same time, it announced it was rejecting the citizen petition.) The clinical laboratory industry pushed back against this proposal, and its allies began to work with the Congress to prepare legislation to prevent it. The US presidential election in November 2016 appeared to end the threat of an expansion of regulation, and, in fact, the FDA shortly thereafter withdrew the

draft guidances and announced that it was moving the issue to the proverbial "back burner" [14]. More recently, though, the Trump Administration's FDA Commissioner, Scott Gottlieb, declared that he does want to revisit LDTs at some point [16].

Communicating Results

In traditional testing the laboratory returns test results to the physician who ordered them. It is then the physician's responsibility to make sure that those results are conveyed to patients in ways that will allow the patient to understand their implications. For practitioners who are medical geneticists, communicating test results for a particular genetic disease is a task well within their expertise. Other physicians may need to make a special study of genetic testing in order to convey results usefully. Some will choose to use genetic counselors in this role. Genetic counselors, as a result of licensure and reimbursement issues, almost always work under the supervision of a physician.

DTC Testing

Traditionally, genetic tests were ordered, processed, and returned like other medical tests—through ordering physicians. Physicians would make the decision that a patient should have a genetic test, would supervise the collection and dispatch of the sample, would receive directly from the laboratory the results, and would communicate those results to the patient. Another approach to genetic tests arose in the last decade. Instead of being ordered by, and returned through, a physician, DTC tests are ordered by the now consumer (not patient), and their results returned directly to the consumer.

At least three firms were reported to have been offering DTC genetic testing as early as 2003, but the real rise to prominence for this field started in 2007, when three highly publicized DTC genetic firms started operations, 23andMe, Navigenics, and deCODEme, an offshoot of a preexisting Icelandic company. All three firms used single nucleotide polymorphism (SNP) chips to provide a wide range of information to their customers and provided SNP chip results directly to consumers for a wide range of genetic susceptibilities. Their focus varied. 23andMe, at least initially, focused on "genetic entertainment," fun facts about ancestry, and nonmedical traits, such as dry- or wet-type earwax. Navigenics and deCODEme, on the other hand, focused on health traits from the beginning, an approach increasingly adopted by 23andMe. Each looked at hundreds of thousands of SNPs to provide information about scores of traits. For prices that initially ranged from just under $1000 (23andMe) to about $3000 (Navigenics), the companies would provide their analysis of a customer's genetic susceptibilities. At the same time, other firms began to offer DTC genetic tests for single genes or for particular traits or risks, some medical and some not. DTC testing is advertised, for example, for athletic ability or for romantic compatibility.

For many DTC tests, observers doubt whether DTC genetic test results have any proven value, and while a physician might be able to choose useful tests, a consumer may fall prey to inappropriate tests or even quackery. At the other end of the test process, commentators have worried that consumers who directly receive their own genetic results might misinterpret them, either overreacting to apparent bad news by taking inappropriate actions (up to and including suicide) or, perhaps even more worrisome, overreacting to good news by avoiding or ceasing good health practices or medical behaviors, such as regular breast cancer screening [17].

Concerns over the spread of DTC genetic testing came to a head in May 2010. That month another DTC firm, Pathway Genomics, announced that it was going to sell an SNP chip-based DTC genomic product in collaboration with the Walgreens drug store chain. The product would have been very similar to that offered over the Internet by 23andMe, Navigenics, and deCODEme. Instead of ordering online, getting a tube in the mail, spitting in it, and returning it in a mailer to the company, the consumer would buy a kit at Walgreens that included the tube, spit in it, and mail it.

The FDA decided to react, sending warning letters to Pathway Genomics (and Walgreens). In the aftermath, the previous DTC firms, whose long-running marketing had been Internet-based, also received similar letters. In the following 3 years, the FDA announced its intent to apply some form of regulation to genetic testing and held several public hearings to get advice on how to regulate. But for several years, no new regulatory scheme emerged.

At about the same time, several states took an increased interest in DTC genetic testing. The New York State Department of Public Health ordered the various DTC companies to stop marketing to people in New York because they had not demonstrated that the offered tests were clinically valid. California issued a similar "show cause" order to some of the companies but, unlike New York, quickly settled in return for some fairly minor concessions, none requiring any proof of clinical validity.

While the FDA's stalled focus has expanded past the DTC test that sparked its public action, DTC genetic tests remain available. Ironically, however, two of the three original DTC SNP chip companies, Navigenics and deCODEme, have been acquired by other firms, which have stopped offering DTC services. 23andMe's focus appears to be shifting from earning money directly from consumers to earning money from pharmaceutical and biotech firms interested in the data it had collected from consumers who were interested in their results.

23andMe hit a roadblock in November 2013, when the FDA ordered it to stop providing the 254 items of "health information" it was then offering. It initially responded by giving its customers (in the USA, customers in Canada and a few other countries could continue to receive the health results) only genealogical information and access to their raw SNP data—but it resumed talking with the FDA [18].

Over time, 23andMe has received the FDA approval to market to consumers about 50 specific genetic tests. It first was given permission to do about 35 tests for carrier screening of consumers concerned about whether their future children were at risk for genetic diseases from them [19]. These tests did not speak directly to the health of the person tested and were considered lower risk. It then got permission to market ten tests for well-known and well-understood genetic risks [20]. Most recently, in March 2018, it was allowed to market a test for three (of the over 1000) mutations in the BRCA1 and BRCA2 genes that greatly increase a woman's risk for breast and ovarian cancer [21]. In each case, though, 23andMe had to demonstrate to the FDA that its test had the medical implications it claimed—and that consumers, without the assistance of a doctor or genetic counselor (usually), would understand the results well enough for them to be potentially useful and not harmful. Several other companies built large consumer markets for genetic testing for ancestry and genealogical purposes; whether they will follow 23andMe into health indications remains unclear.

Other Issues

Genetic testing, even in its "targeted" version, raises other questions. This section will discuss six specific concerns: confidentiality, discrimination, testing children, family relationships, updating test results, and the relationship of clinical genetic testing to research.

When genetic testing is part of the medical process, the results are health information, protected under state law and the Health Insurance Portability and Accountability Act (HIPAA) regulations just as much, and as little, as any other health information. All health information is subject to both unauthorized and authorized breaches of confidentiality. Unauthorized breaches can come from hackers, misbehaving insiders, or, most frequently, lost laptops or other electronic devices. None of this leakage can be completely prevented. It can only be limited, through appropriate detection and punishment as well as well-designed systems.

Authorized disclosures include disclosures to other health personnel, in many cases to insurers or other payors, to law enforcement, and to those with court orders. Although the law allows these disclosures, patients may not expect them and might be upset about them. For example, law enforcement officials seeking to determine whether a suspect's DNA matches that found at a crime scene might seek genetic information about a suspect from health providers. In one famous case involving a serial killer from Kansas, the police sought and received from a university clinic part of a Pap smear from the daughter of a suspect. After it was genotyped for the 13 "CODIS" markers, the short tandem repeats used for forensic identification in the USA, the investigators were able to determine that the crime scene DNA could have come from her father. They arrested the father and took a DNA sample from him that matched the crime scene DNA, and he pleaded guilty. What the daughter felt about this is not known [22]. The targeted nature of traditional genetic testing affects the incentives around breaching confidentiality in two conflicting ways. If a patient's records only contain the results of a genetic test for one condition, they will not contain much, if any, information about other conditions. Instead of a disclosure leading to information about many genes and many disease risks or traits, it can only lead to information about one. On the other hand, targeted testing is usually only done for people at increased risk of carrying some deleterious genetic variant. Thus, for example, the chance that someone tested specifically for a *BRCA1* mutation actually has one is greater than the chance that a random person has one.

Genetic discrimination has been discussed for almost as long as genetic testing has been used. Concern over this kind of discrimination seemed widespread even though there was very little evidence that insurers or employers had, in fact, used genetic test results to discriminate [23]. Perhaps some of the problem was that physicians and genetic counselors felt compelled to tell patients of the risk of genetic discrimination, whether or not there was strong evidence of its actual existence.

During the 1990s and through the 2000s, the USA adopted more and more protections against genetic discrimination in health insurance and employment. When the Congress adopted HIPAA in 1996, it took two important but little-noticed steps concerning genetic discrimination. First, it banned the use of genetic risks as preexisting conditions for most health coverage. Second, it forbade employers or unions from applying any medical underwriting, based on genetics or otherwise, to those with employment-provided coverage. The vast majority of Americans with private health coverage obtain it as a result of employment; this ban on medical underwriting meant that none of their risks, genetic or otherwise, could be considered. Americans who needed individually underwritten coverage increasingly were protected by state laws banning genetic discrimination in health coverage. By 2008, over 45 states had some bans on the use of genetic information in health coverage; some states also banned employment discrimination.

In 2008 the Congress passed the Genetic Information Nondiscrimination Act (GINA), banning, as a matter of federal law, discrimination in health insurance or in employment

that was based on "genetic information," defined broadly enough to cover not just the results of tests on DNA, RNA, or proteins but also on family history. GINA applies to employment and health coverage; it does not apply to life insurance, disability insurance, or long-term care insurance, although many states have now passed statutes extending protection to those or other areas. In recent years there has been increasing discussion of expanding GINA to include those other possible sources of discrimination.

Clinical genetics had reached a consensus that testing of children should be limited to only those conditions for which the test results would make a medical difference before they become adults [24]. Testing for sickle cell disease can affect early childhood health care; testing for Tay-Sachs disease can affect, sadly, childhood prognosis. Testing children for *BRCA1* mutations cannot change their medical management during childhood, though there is some fear that the results might change, and in negative ways, how their parents treat them. Children therefore have generally not been tested for adult-onset diseases. This consensus has begun to fray [24]. In September 2013 NIH created several pilot projects to do whole-genome sequencing of newborns, informing their parents of both their near-term and longer-term genetic risks [25].

Families share genetic variants, which means family relationships complicate genetic testing, but in two different ways. On the one hand, one family member's test result provides some information about another family member's genome. If a patient has a dominant genetic variant conferring high risk for a disease, unless it is a new mutation, one of the patient's parents must have had that variant and hence that risk. The patient's sibs and children will also have a 50% chance of carrying the variant. Patients are currently encouraged to inform their possibly affected relatives about these risks; the obligations of healthcare providers if the patients refuse are still unclear. On the other hand, genetic testing may provide unexpected and unwanted family information. Some nontrivial percentage of men who believe they are a child's father may be the father in every way *except* being the genetic father. If genetic testing is done on the child and both (putative) parents, this "false paternity" is easily detected. If genetic testing is done on two putative siblings, they may be shown to be half-siblings. Who, if anyone, should be told of these genetic family relationships continue to be a largely opaque area in genetic testing.

One beneficial aspect of (appropriate) genetic tests is that they only have to do it once. Except for tumors, a person's genome does not change substantially over a lifetime. The meaning of that genomic sequence may change, however, as new information is discovered about connections between various genome sequences (or collections of sequences) and disease risks. Ideally, the clinician will be aware of possible changes in the meaning of a patient's tested genetic variants. This is particularly important, of course, if the tested individual has a "variant of unknown significance," as because its significance may change from unknown to known. But even variants that are classified as pathogenic or clinically insignificant may change their meaning with new knowledge, particularly knowledge of other modifying sequence segments.

At least one lawsuit has already been launched over an updating question. In 2008, a toddler died as a result of seizures [26]. Eight years later, his mother sued the genetic testing firm. It had initially said that child had a "variant of unknown significance" but later reclassified it as disease-causing. The suit alleges that knowledge of the effect of this variation would have improved the boy's treatment, quality of life, and survival—and further that the firm should have both reclassified the variation earlier, before the boy died, and informed the family. The case is currently tied up in procedural fights, but other fights about updating results, and informing patients of the updates, will undoubtedly follow.

Finally, clinical genetic testing puts pressure on the boundary between medical practice and medical research. Especially if a patient's variant is of unknown or unclear significance, either medical information on the patient at the time of testing or subsequent follow-up information may be of great research value. And yet patients may not want their cases used in research, even if their data are "anonymized." Finding the line between using these potentially invaluable data for research and respecting a patient's right not to be used in research is a continuing tension.

Steps Toward -Broad Genetic and Genomic Testing

The division between today's targeted genetic tests and tomorrow's broadband genomic tests is not entirely accurate. For one thing, "targeted" specific tests for specific DNA variants are likely to continue into the future, at least until highly accurate WGS becomes a nearly universal part of good medical care. More importantly, some "broadband" genetic or genomic tests already exist. Three important examples are multiplex tandem mass spectrometry as a part of neonatal genetic screening, SNP chips, and array-based comparative genomic hybridization (aCGH). For varying reasons, however, none of them has been deeply revolutionary.

The adoption of tandem mass spectrometry for neonatal genetic screening in the mid-2000s was a major change in neonatal screening [3, 12]. Neonatal genetic screening has its roots in the early 1970s with the rise of state-mandated tests for all infants for a handful of serious genetic diseases. Tandem mass spectrometry technology changed this model. This method of looking at proteins (not directly at DNA) allowed simultaneous screening for scores of genetic diseases. The cost of the machine was substantial; the costs of

adding extra screening tests were minimal. The result was the rapid expansion of mandatory screening in every American state to include 30–50 different conditions, including many for which the value of early intervention is unclear.

Tandem mass spectrometry allows multiplex testing and hence is a "broadband" form of genetic testing, but in more meaningful ways, it remains targeted. Although it tests for scores of diseases, each disease is quite rare. The vast majority of screened babies will have no positive results, and almost no babies will be so terribly unlucky as to have positive results for two different diseases, each of which is found in only one baby in, say, 10,000.

SNP chips provide another example approaching broadband testing technology. These arrays, pioneered in the mid-1990s for research uses by Affymetrix and other firms, allow the rapid categorization of a DNA sample by a set of chosen SNPs, locations in the genome where some people have one nucleotide, while others have a different nucleotide. The original SNP chips could simultaneously detect which nucleotide was found in thousands of different locations in a DNA sample for a few thousand dollars, whereas today's chips can look at a million SNPs for around a hundred dollars.

SNP chips are also a broadband technology—of a sort. They can deliver millions of pieces of genomic data for a small price. The problem is that the data they provide has very little clinical value. Although it is possible to have medically important SNPs, these are unusual. SNPs are almost never medically significant in themselves; their medical significance, if any, comes from the fact that they are usually found along with nearby alleles that are medically significant.

The result is that the medical risk attributable to having an A instead of a G in a particular SNP will usually be quite small in absolute terms, even if it is statistically significant. Statistical significance can make it useful as a research tool because it indicates that something nearby (and hence inherited along with that SNP allele) may be causally related to a disease, but the small risk makes it almost useless for direct medical application. Thus, before FDA's action limiting 23andMe's uses of its SNP chip, results came with the interpretation of the customer's "genetic risk" of over 250 diseases. For almost all of those diseases, the absolute change in risk is too small to be useful. For example, one of the strongest disease associations claimed by 23andMe was that a particular set of SNPs will lead to a person having a fourfold higher than normal risk of being diagnosed with Crohn's disease. The underlying risk of Crohn's disease in the general population, however, is 0.7%. A fourfold increase takes that risk all the way to 2.8%. Few people will change their lives because their risks of that disease are 1 in 35 instead of 1 in 143. (The now FDA-approved 23andMe tests have more substantial effects, though sometimes at the cost of looking at only a few of the many possible risky mutations in a gene.)

aCGH is another "partially broadband" technology. It is a useful tool for revealing whether, for any given spot, a sample has too little, too much, or just the right number of copies of DNA. This is mainly important for recognizing copy number variations, ranging from whole chromosomes (thus, seeing, for example, whether a sample has three copies of chromosome 21 or only one copy of the X chromosome) to insertions or deletions of DNA that are several thousand base pairs long. This technology, therefore, is also somewhat broadband: one test will reveal any regions of the sampled genome that have other than the usual two copies (or, in the case of the X and Y chromosomes in males, other than the normal one copy). It has thus become a widely used tool for the testing for aneuploidies, such as trisomy 13, 18, and 21, as well as for aneuploidies of the sex chromosomes—X0, XXY, XXX, XYY, and others. Apart from those major aneuploidies, however, the importance and meaning of copy number variations remain generally unclear. Some have been associated, at least on a research basis, with various conditions, but the number of clinically meaningful associations between copy number variations at the less than whole chromosome level and particular diseases remains, as yet, small.

WGS (and its less comprehensive relative, whole-exome sequencing or WES) seems to provide the best of all these tests or holds the promise of doing so. Like SNP chips and aCGH, it looks across the whole genome (or, in the case of WES, all the parts of the genome that directly code for protein), but, unlike them, it will provide powerful information about many sites. These characteristics raise all the problems of targeted genetic testing and more.

Broadband Genomic Testing: Ethical and Legal Issues

We have already entered the era of clinical WES and WGS. Tens of thousands of people have now had whole exomes or genomes sequenced from their bodies' tissues—some for curiosity, some in search of a diagnosis for a mysterious childhood syndrome, and some to have tumors sequenced in the hopes of finding a better treatment against their individual cancer variant. Exome sequencing of tumor DNA to detect somatic changes has been particularly popular with one firm, Foundation Medicine, with over $50 million in revenues for over 67,000 clinical tests in 2017 [27]—even though the technique has not been shown, except anecdotally, to be effective. (Although in many of the cases so far, particularly those involving cancer, only the exome has been sequenced, the remainder of this chapter will refer to WGS to include both exome and genome, in the assumption that cheap WGS will eventually drive out WES.)

WGS is not yet in widespread clinical use, but its use will grow rapidly as its price goes down and its accuracy

improves. Soon, people whose doctors think they need a test for a particular genetic risk will be offered WGS instead of testing for just the appropriately targeted genes. Before much longer, we will see neonatal genetic screening for 30–50 diseases replaced by WGS with its power to predict thousands of diseases. Eventually almost all people with access to good health care will probably have their whole-genome sequences in their electronic medical records.

The wide use of WGS, effectively for screening purposes, holds the promise of substantial health benefits, if done wisely. It also holds the certainty of substantial ethical, legal, and practical challenges during its implementation. Those challenges, like the issues confronting traditional medical genetic testing, can be seen in five categories: accuracy, informed consent, return of results, DTC provision, and "other concerns." In each case, the move to WGS or WES complicates the issues, sometimes massively.

Deciding to Test: Medical Appropriateness and Informed Consent

The initial problems for clinical WGS are deciding when to use such a test and how to obtain informed consent.

With a traditional single trait or single gene, one generally asks whether the test is appropriate for the patient. That will still be the case in some uses of WGS. Looking for a genomic cause for a mysterious syndrome or looking for points of attack in a tumor's genome will usually need to be done through WGS because the specific variants of interest could be anywhere in the genome. The use of WGS purely for screening purposes, however, either as neonatal genetic testing or as a routine part of medical care, raises different questions. Screening is not the same as looking for the cause or nature of a known or suspected condition. Decisions to screen require a different, and generally broader, calculation of the individual and societal costs and benefits rather than purely an emphasis on an individual patient. Someone—whether governments, professional organizations, consensus conferences, or others—will need to decide whether and to what extent population-wide screening uses of WGS will be appropriate. That decision will not normally rest with an individual's physician or perhaps even with the individual.

Between the extremes of necessary WGS and screening WGS lies opportunistic use of WGS. A patient may present good medical reasons for getting genetic testing targeted at a particular gene or trait, but if the price of WGS is near, or below, the price of the targeted testing, it may be tempting—to the doctor, the patient, and whoever is paying for the test—to order WGS. It is conceivable that the WGS could be examined *only* for the targeted genes or traits, in which case the other information would be discarded and in that case it would, in effect, just be another method of doing targeted testing. But, more likely, the WGS information would be used both to answer questions about the trait of interest and to screen this particular patient. In that case, the physician should only recommend, or order, WGS if she is confident that, on average, the information it brings will have some net benefit to the patient. If WGS has not yet been established to have positive (or, at least, nonnegative) expected value as a general screening tool, the physician should resist the temptation—and possibly the urgings of the patient and the payor—to order a medically useless WGS test instead of the targeted test.

WGS makes informed consent much more difficult and, in some ways, frankly impossible. With traditional testing the patient can learn about the advantages and disadvantages of being tested for particular genes or traits and make an informed decision whether to accept that testing. With WGS the patient is being tested, at least potentially, for *every* genetically influenced trait and every stretch of genome. No patient can learn about each one of the thousands of genetic traits before deciding to accept or reject testing on each specific trait. There are too many traits and too little time.

Instead, patients will need to learn about the kinds of results that WGS could provide them and to decide what kinds of results they want. Are they interested in hearing about genomic variations that are only of reproductive significance? Do they want to learn only about risks above a certain cutoff, for example, those that are more than double the average risk and that have an absolute risk of over, say, 10%? Do they want to learn about risks for which, at least at this point, there are no useful medical interventions? That conversation could, at least in theory, take place after the WGS was performed and before results were returned, but facing these questions might help the patient decide whether to undertake WGS at all. It is thus better done as part of the informed consent process.

In addition to information about the possible results, patients will also need to be told some background facts about WGS. These include not only what it is but what the realistic chances are that the WGS will produce false positives and false negatives that could affect them. They also need to consider the possible effects on their genetic privacy, as well as the possible implications for their parents, siblings, children, and other genetic relatives. And they need to know that the interpretation of their genome may change, so that WGS will not be a "do it once and forget it" procedure.

Test Accuracy

It will be neither ethically nor legally appropriate to make widespread use of grossly inaccurate sequencing results in clinical decisions, and, at least in the near future, accuracy of WGS will be a major issue, both for its analytic and its clinical validity.

The first question will be: "how well will WGS detect the actual sequence of the genome?" Laboratories, researchers, and clinicians know how well the current testing protocols identify genetic variations, but each sequencing machine and each protocol under which the WGS is done will affect this basic sequence accuracy. How accurate are the sequencers? We do not really know. Some companies have reported accuracy levels; Complete Genomics, for example, reported as early as 2008 that its sequence calls were 99.999% correct, making 1 error in 100,000 calls. (This would, in a 3.4 billion base pair haploid genome, still mean 34,000 errors.) In 2011, a team headed by Michael Snyder compared two sequences generated by two different firms' sequencers, those of Illumina and of Complete Genomics [28]. They reported that of 3.7 million single nucleotides where the tested individual was known to vary from the reference human genome, two firms agreed in their calls of only 88% of them. One of the firms appeared to be more accurate on the divergent calls than the other, but in both cases, a large fraction of their discordant SNP calls was wrong. Still worse, the team looked at the calls made on short insertions (about 50 base pairs or fewer) and short deletions (about 200 base pairs and fewer). The concordance rate on these "indels" between the two platforms was only about 25%.

In addition, we know that there are other aspects these sequencing techniques do poorly at this time. These include determining the length of repeating sequences, some of which are involved in serious human diseases, such as Huntington disease. They also do not necessarily "phase" the results by revealing on which chromosome different variants are found. If a genome shows two different deleterious mutations in a gene known to be involved in an autosomal recessive disease, the person would be unaffected, but a carrier if both of the mutations are on the same chromosome should be affected if they are on different chromosomes.

WGS is nowhere near being ready to be used by itself across the genome for clinical purposes, although some centers may offer it. Any clinical use currently must require a protocol that confirms the most important findings using independent methods. The Snyder team suggests a number of strategies, from sequencing samples using multiple platforms to doing WES to validate findings in the more important parts of the genome and to using more established Sanger sequencing or array capture technologies to validate particularly important findings [28]. The total accuracy of the resulting WGS will depend not just on the raw accuracy of the sequencing machine but of the accuracy of the whole protocol, including the informatics pipeline and interpretation.

The clinical side of accuracy is even more daunting. At least there is some "gold standard" of reality to the sequence (analytic validity), but the medical implications of that sequence often will be much less clear. Some genetic variants, including presumably those most commonly found in humans, can be safely considered nonpathogenic; others, with long track records in medical genetics, such as the 185delAG mutation in *BRCA1* or more than 39 CAG repeats in the *HTT* gene, are equally well known to be serious. But WGS will turn up hundreds of thousands of VUSs, sequence variants that are not known to be either clearly safe or dangerous. How will they be called?

Apart from testing and interpretation in academic laboratories, one possibility is that different firms will spring up to provide, as a service, genome sequence interpretation, either with or independently from the actual sequencing work. (At least one such firm, Personalis, already exists, though it limits itself now to research uses only.) In other cases, the firms that do the sequencing may also do the interpretation. If multiple firms provide interpretative services, will their answers be consistent—and will their answers be "right"?

The DTC SNP business has already provided a worrisome example of this issue. In 2009 a team headed by Craig Venter sent samples from the same five individuals to both 23andMe and Navigenics and compared the results [13]. Both firms did well (though not perfectly) in calling the underlying SNPs; they agreed more than 99.7% of the time on the 500,000–one million SNPs. But when Venter's group looked at the interpretation of the results with respect to 13 diseases, they found that the two firms disagreed on the relative risk about a third of the time. For four diseases, the firms agreed entirely; for another seven, they agreed half the time or less. In 2010 the US Government Accountability Office did a similar study, with similar results [29].

Now imagine the difficulties of different companies interpreting whole-genome sequences, each with thousands of different genetic risks. If each company makes its own calls based on its own proprietary, nonpublic decision-making algorithms, it will be impossible to compare the bases for the calls. More importantly, it will make it very difficult for anyone—a doctor providing a second opinion in this case or researchers interested in that specific disease—ever to assess who is right.

A similar problem already exists with respect to the "traditional" genetic testing of *BRCA1* and *BRCA2*, where Myriad Genetics has refused since 2004 to share its database of sequence variants and patient outcomes, making informed second opinions nearly impossible. Although the Supreme Court's decision in *Association for Molecular Pathology v. Myriad Genetics, Inc.* [30] has removed Myriad's patent monopoly on testing for *BRCA1* and *BRCA2* mutations, their database of mutations, the fruit of that monopoly, remains their property. (There is an initiative called "Free the Data!" that is trying to create a similar, open database, in part by asking physicians and patients who used Myriad to share their, de-identified, test results.)

This interpretive step is both crucial and difficult. It will become even more difficult without broad and open access to information needed to assess the likely accuracy of any one interpretation as shown by patient outcomes. At the research level, the "Global Alliance," announced in early June 2013, hopes to avoid some of these problems by providing broad access, with research subject consent, to genomic and health data from sequencing experiments [31, 32]. Something similar is likely to prove necessary for clinical uses.

Communicating Results

In clinical WGS, communicating results will be no different in kind than in traditional genetic testing. The testing laboratory (in this case, perhaps through a separate WGS analysis firm) will return results to the physician who ordered them, who will in turn need to explain the relevant ones to the patient. In degree, however, the difference in quantity with WGS will effectively transform the activity.

Traditional testing provides information usually on just one disease or risk, while WGS will provide information—positive, negative, or indeterminate—on any risk, trait, or disease that is associated with genomic sequences. Even today, that amounts to thousands of diseases and susceptibility risks, as well as hundreds of pharmacogenomic traits, hundreds of SNP chip associations (as WGS necessarily reveals SNPs as well), and scores of physical or behavioral traits not related to disease. Each of those numbers will only grow as our understanding of the relationships between specific genomic sequences, or combinations of sequences, and phenotypes grows. This embarrassment of riches leads to two questions: what information should be returned and how should it be returned?

The first question has already begun to be debated, at least around its edges. In 2013 the American College of Medical Genetics and Genomics (ACMG) issued guidelines recommending that when a patient is offered WGS as part of the investigation of a particular risk or disease, the laboratory must return to the physician, and the physician should discuss with the patient, any highly penetrant, serious risks for which there is good medical intervention, whether or not they have anything to do with the reason for seeking genetic testing [7]. The ACMG guidelines initially listed 56 such conditions. In late 2016 it added four more genes, but subtracted one, for a current total of 59 [33] as of the time of writing. Thus, if someone was tested for long QT syndrome risk, but the WGS showed high risks for Lynch syndrome, that second risk would have to be disclosed. The European Society for Human Genetics (ESHG), by contrast, has recommended that only expressly sought-after risks be disclosed [34]. And at least one article has been published decrying the ACMG position for removing from patients the right to remain ignorant of their risks [35], coupled with another article supporting the ACMG position [36] and an ACMG "clarification" [37].

Note that the debate, thus far, is merely about what results physicians *should* return when they are doing the WGS on their patient for a specific indication or indications and the patients have not requested any additional information. It does not deal with screening uses of WGS at all or, even when it is being used for targeted purposes, with what physicians should do when patients request broader information or how physicians, before offering WGS, should determine what their patients want. As discussed above, the categories of results that patients want to receive back from WGS should be a major topic of conversation before ordering such a test.

The ESHG position seems both morally and, at least in the USA, legally questionable. A physician has a fiduciary obligation to put his or her patient's interests first. To *not* to tell the patient about known or readily discoverable information concerning a highly penetrant and serious genetic risk, for which there is a good medical intervention, seems a breach of the physician's obligation to the patient [38]. Physicians who, when examining lungs in a CAT scan, see (or read in a radiologist's report) something highly suspicious in the liver will not ignore this unlooked for finding. At least some American case law finds that they can be liable if they do. A patient perhaps should be able to retain a "right to ignorance," but only after being informed of the possibility of such findings and expressly requesting not to be informed. If a patient made that decision after adequate information, and the physician documented the decision, the physician's legal risks should be minimal.

But what about the cases, which ultimately should become the majority, where the testing is being done for screening purposes or where the patient says "tell me everything." In both cases, the patient should be told at least whatever scientifically valid information seems likely to be significant. (The patient who wants to be told "everything" could also be given a flash drive with his whole sequence.) What might be significant to one patient might not be important to another, so the physician and patient will need to have had some discussion about what the patient wants, preferably before the test was ordered, but if not then, before results are returned.

In the absence of that discussion, the physician should return at least the kinds of highly penetrant, serious disease risks for which good medical intervention exists, as listed by the ACMG Standards and Guidelines, but that must be viewed as a minimum. Those criteria were created for cases in which the testing was being performed for one specific indication but other things turned up; in screening, the idea is to look generally for problems.

What information should be returned, absent an express discussion, should probably include disease risks that are substantial (in absolute risk) and significantly higher than the average person's (in relative risk). Whether they should go beyond those that are medically actionable is a harder question and one that really needs to be answered directly by the patient.

The second question is of a different nature—it is not a normative question of "what" should be done, but a practical question of "how." In the screening context, especially with a patient who wants to be told "everything" (even if qualified to everything significant), the numbers could quickly become unwieldy.

In 2009 Stanford bioengineer Stephen Quake had his genome sequenced on a sequencing machine he had invented. In 2010, 32 authors published a medical analysis of this genome [39]. They concluded about 100 findings should be shared with him. Even 3 min of discussion of each of 100 issues amounts to 5 h of counseling. Few counselors could provide good information on 100 random genetic issues; fewer could talk for 5 h, and even fewer patients could listen for 5 h. And, no one would pay for that counseling [10].

Information, including genomic information, is only useful if it is properly understood. Patients who do not understand the significance of their results, as noted above, might overreact or underreact to them in ways that could be risky. Yet how many patients will understand, on their own, the meaning of more than 100 genetic risks? And how many of their personal care physicians will know themselves, or be able to convey if they do know, the meaning of more than a handful of those risks? Using genetic counselors sounds good, but there are no more than 3500 genetic counselors in North America today, and they are all busy. Thus, the interpretation by appropriately trained pathologists and their direct involvement in multidisciplinary teams that work together to arrive at an interpretation that takes into account the patient's clinical context and wishes, and that is versed in the delivery of such interpretations, may become increasingly important.

The challenge, perhaps biggest for WGS, will be coming up with economically feasible ways to convey complex, important probabilistic information to hundreds of millions of people, many with very limited knowledge of genetics. The movie and television industries, computer game companies, advertising agencies, and others need to be engaged in order to produce systems that allow individuals to find out, probably online, what their WGS means. Face-to-face counseling with trained professionals should, however, remain part of communicating results; face-to-face, real-time communication (perhaps online through some HIPAA-compliant version of Skype or a similar venue) gives a better chance to answer questions and to spot confusion or misunderstandings.

DTC Marketing

The current lack of regulation of DTC for traditional and SNP chip testing also applies to WGS. Several firms seem to be currently offering DTC WGS: a firm originally named DNA DTC but now called Gene by Gene is offering WGS for just under $3000 (and WES for about $1100) [40]. Interestingly, it is offering only the naked sequence, without any interpretation, calling it a nonmedical service—customers will have to figure out what it means, medically or otherwise, on their own.

Any concerns about DTC for individual tests or SNP chips are magnified for WGS, because of the breadth of information—and misinformation—the whole sequence can convey. Both the decision to order WGS without any professional advice and the return of information from WGS without any professional advice seem reckless. The customers may not know what they are ordering or what they have received.

Other Issues

In discussing traditional testing methods, this chapter looked at six concerns: confidentiality, discrimination, testing children, family relationships, updating test results, and the relationship of clinical genetic testing to research. WGS makes each of those more complicated.

With WGS, breaching confidentiality will become more tempting. With traditional genetic testing, a breach of confidentiality leads only to information about one (or a few) variant(s) or trait(s). WGS information, because of its breadth, is more useful to someone who wants to find out something about the patient, whether a hacker or the police. Police might, for example, seek clinical WGS records for a suspect to compare with crime scene DNA; electronic medical record databases could, in effect, supplement the forensic databases of convicts and arrestees. Some of the confidentiality breaches may be unauthorized and hence illegal, but others, such as those following court orders, will be fully legal.

WGS would also magnify, in some respects, discrimination concerns because one can learn of many different risks in a person. On the other hand, WGS will show that everyone has *some* genetic weaknesses. In the long run, this seems likely to lead to less discrimination, as the perpetrators realize they could themselves become victims and because society sees more tangibly the shared nature of these risks.

If WGS is done on children, either for neonatal screening or as part of a test for a particular disease, it will strain the current consensus on not testing children for conditions for which nothing need be done in childhood. WGS will provide information on both childhood-onset and adult-onset diseases; if the latter information is not revealed immediately

to the parents, then what should be done with it—should it be put in an envelope marked "do not open until your 18th birthday"? Also, WGS for a child might reveal genetic risks for the child's parents or siblings that could be important right away. The ACMG Standards and Guidelines prompted more controversy by recommending the immediate return to parents of positive results in children's sequences for any of its 56, now 59, listed genes [35]. This controversy will have to be resolved before widespread use of WGS in children.

Comparing WGS makes spotting family relationships very easy. With the number of variants revealed by WGS, even fairly distant relationships between people, such as second or third cousin, should be ascertainable with a high degree of likelihood based on the percentages of variants shared. This increased utility for finding family relationships is another factor that may drive people to try to breach confidentiality—such as frustrated genealogists or state officials trying to find a genetic father to sue for child support.

Updating may be the issue that complicates WGS the most. With tests for a single gene, updating the result may or may not ever be necessary. A clear and definitive "safe" or "high-risk" variant seems unlikely to change; a VUS, however, would need to be monitored and updated, either from the physician's personal knowledge of the area or from some periodic reanalysis. With WGS, not only will each patient have tens of thousands of VUSs, but the overall interpretation of the human genome will change every day with new discoveries on particular variants. The individual's genome will not change, but the interpretation of any individual's genome is likely to change fairly frequently. Not only will some VUSs be resolved, but also many "known" variants will have their effects refined, with better knowledge of risk, severity, age of onset, or the effects (positive or negative) of particular environmental factors or of variants in other genes.

In effect, the interpretation of a patient's genome will be a test that will have to be repeated regularly, like a blood pressure test or a Pap smear. Once optimal accuracy of WGS has been achieved, no office visit or actual test may be required, but the existing genome sequence will need to be run through the analysis protocol every few years. The field should agree on just how often such reinterpretation is appropriate. (Of course, if the cost of WGS becomes low enough, it may be cheaper to resequence the patient every few years than to store the data. This would, however, require the costs of sequencing to continue to plummet at the same time the declining costs of computer memory stalled.)

WGS will increase the pressure to blur the line between research and clinical data. Looking at whole-genome sequences along with a host of clinical, phenotypic information in the electronic medical records of millions of patients will likely be the most useful way to make progress in understanding the effects of different sequence variants, alone and in combination. Yet those patients may not have agreed to have their records used in research. And given the wealth of potentially identifying details contained both in the phenotypes in medical records and in the genotypes, almost any patient could, with some effort, be reidentified from his or her records. In the long run, it would be good for everyone for this information to be as widely available for research as possible. But to do so without the consent of the patients would not only be unethical and, most likely, illegal, but it would risk a political backlash against medical research and genetics generally. Resolving this dilemma will require careful effort.

Conclusions

We have had clinical genetic testing for over 45 years. It has raised some complex ethical and legal questions, but those questions have largely been answered, if not perfectly, at least satisfactorily. New technologies, and particularly WGS, will raise those ethical and legal questions to new and higher levels. The old answers may be useful as guides, but they cannot provide satisfactory solutions by themselves. This change in degree really is a change in kind. Implementing clinical WGS effectively will require serious improvements in sequencing technologies but even more serious interdisciplinary efforts to deal with the ethical, legal, and, ultimately, very practical problems it raises. The time for such efforts is quite urgent now.

References

1. Greely HT. Human genomics research: new challenges for research ethics. Perspect Biol Med. 2001;44(2):221–9.
2. Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. Annu Rev Genomics Hum Genet. 2007;8:343–64.
3. Greely HT. Testing infant destinies. Nature. 2012;492:192.
4. Am College Med Genet, Points to consider in the clinical application of genomic sequencing. 2012. http://www.acmg.net/StaticContent/PPG/Clinical_Application_of_Genomic_Sequencing.pdf. Accessed 5 Jul 2013.
5. Biesecker LG. Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. Genet Med. 2012;14:393–8.
6. Green RC, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.
7. Green RC, et al. Clinical genome sequencing. In: Ginsburg G, Willard H, editors. Genomic and personalized medicine. 2nd ed. San Diego: Elsevier; 2013. p. 102–22.
8. Manolio TA, et al. Implementing genomic medicine in the clinic: the future is here. Genet Med. 2013;15(4):258–67.
9. Mayer AN, et al. A timely arrival for genomic medicine. Genet Med. 2011;13:195–6.
10. Ormond KE, et al. Challenges in the clinical application of whole-genome sequencing. Lancet. 2010;375:1749–51.
11. Schrijver I, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the association for molecular pathology. J Mol Diagn. 2012;15:525–40.

12. Timmermans S, Buchbinder M. Saving babies?: the consequences of newborn screening. Chicago: University of Chicago Press; 2012.
13. Ng PC, et al. An agenda for personalized medicine. Nature. 2009;461:724–6.
14. Gibbs JN. LDTs: the Saga continues, Update Magazine. 2017. https://www.fdli.org/2017/04/ldts-saga-continues/.
15. Genome Web. Lab group submits citizen petition against FDA regulation of lab developed tests. 2013. http://www.genomeweb.com/print/1237821?hq_e=el&hq_m=1588054&hq_l=1&hq_v=162223e076. Accessed 24 Aug 2013.
16. Gottlieb S. Remarks by Dr. Gottlieb at the American rClinical Laboratory Association Annual Meeting. 2018. https://www.fda.gov/NewsEvents/Speeches/ucm599551.htm.
17. Am College Med Genet. ACMG statement on direct-to-consumer genetic testing. 2008. https://www.acmg.net/docs/ACMG_Statement_on_DTC_4.07.08.pdf..
18. Zettler PJ, et al. 23andMe, the Food and Drug Administration, and the future of genetic testing. JAMA. 2014;174(4):493–4. https://doi.org/10.1001/jamainternmed.2013.14706.
19. Pollack A. 23AndMe will resume giving users health data, NY Times; 2015.
20. FDA. FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions. 2017. https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm.
21. Weintraub A. FDA clears 23AndMe's DTC breast Cancer gene test -- but 'Buyer beware', Forbes. 2018. https://www.forbes.com/sites/arleneweintraub/2018/03/06/fda-clears-23andmes-dtc-breast-cancer-gene-test-but-buyer-beware/#230563a63c1e.
22. Nakashima E. From DNA for family, a tool to make arrests, Wash Post. 2008. At http://www.washingtonpost.com/wp-dyn/content/article/2008/04/20/AR2008042002388.html.
23. Greely HT. “Genotype discrimination”: the complex case for some legislative protection. Univ Pa Law Rev. 2001;149:1483–505.
24. Ross LF, et al. Technical report: ethical and policy issues in genetic testing and screening of children. Genet Med. 2013;15:234–45.
25. Kaiser J. NIH studies explore promise of sequencing babies’ genomes, Science. 2013. http://www.sciencemag.org/news/2013/09/nih-studies-explore-promise-sequencing-babies-genomes.
26. Ray T. Mother's negligence suit against Quest's Athena could broadly impact genetic testing labs, Genome Web. 2016. https://www.genomeweb.com/molecular-diagnostics/mothers-negligence-suit-against-quests-athena-could-broadly-impact-genetic.
27. Genome Web, Foundation medicine preliminary Q4 revenues up 70 percent, Genome Web. 2018. https://www.genomeweb.com/molecular-diagnostics/foundation-medicine-preliminary-q4-revenues-70-percent.
28. Lam HY, et al. Performance comparison of whole genome sequencing platforms. Nat Biotech. 2012;30:78–82.
29. Government Accountability Office. Direct-to-consumer genetic tests: misleading test results are further complicated by deceptive marketing and other questionable practices. Washington, DC: GAO; 2010.
30. Association for Molecular Pathology v. Myriad Genetics, Inc. 2013. 569 U.S 576, 133 S.Ct. 1958.
31. Global Alliance. Creating a global alliance to enable responsible sharing of genomic and clinical data. 2013. http://www.broadinstitute.org/news/globalalliance. Accessed 2 Sept 2013.
32. Hayden EC. Geneticists push for global data sharing. Nature. 2013;498:16–7.
33. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics, Genetics in Medicine. Am College Med Genet. 2017;19:249–55.
34. van El CG, et al. Whole-genome sequencing in health care: recommendations of the European society of human genetics. Euro J Hum Gen. 2013;21:580–4.
35. Wolf SM, et al. Patient autonomy and incidental findings in clinical genomics. Science. 2013;340:1049–50.
36. McGuire AL, et al. Ethics and genomic incidental findings. Science. 2013;340:1047–8.
37. Am College Med Genet. Incidental findings in clinical genomics: a clarification. 2013. www.acmg.net/docs/Incidental_Findings_in_Clinical_Genomics_A_Clarification.pdf. Accessed 5 Jul 2013.
38. Clayton EW, et al. Managing incidental genomic findings: legal obligations of clinicians. Genet Med. 2013;15:624–9.
39. Ashley EA, et al. Clinical assessment incorporating a personal genome. Lancet. 2010;375:1525–35.
40. Gene by Gene. https://www.genebygene.com/pages/dnadtc.

Part III

Clinical Implementation of Diagnostic Genomics

Transitioning Discoveries from Cancer Genomics Research Laboratories into Pathology Practice

13

Tamara Jamaspishvili and Jeremy A. Squire

Introduction

Cancers of an identical primary site can be heterogeneous in molecular pathogenesis, clinical course, and treatment responsiveness, which reflects the existence of multiple cancer subtypes [1]. The differentiation of these subtypes is often based on biomarkers that distinguish important cancer features such as the aggressiveness of the disease (prognostic biomarkers) or the response to treatment (predictive biomarkers). The latter have fueled an increasing interest in biomarkers, given the potential they hold for individualized or personalized medicine. This new field focuses on differences between people and the potential for these differences to influence medical outcomes. With individualized or "precision" medicine, a person's cancer may be subtyped based on an explicit biomarker that is present or absent or that may have increased or decreased expression levels. This may result in a greater likelihood of receiving treatment that is appropriate and effective for a specific tumor in a particular cancer patient. Individualized medicine contrasts markedly with the traditional "empiric method," which uses a standardized treatment for the whole patient population with an established presentation of disease symptoms, based on long-standing generic descriptions of the average patient (Fig. 13.1).

Nowadays, tumor biomarkers, together with new genomic and proteomic technologies, provide powerful tools for the early identification of cancer patients and recurrent disease and for defining therapeutic responsiveness. In spite of the rapid developments in biotechnology and genomics, the pace of acceptance of new markers in clinical practice is surprisingly low. The slow uptake is due to the substantial reasons presented below and elsewhere [1–3]. In this chapter we (1) summarize the importance of personalized medicine and describe some of the biomarkers and genetic tests which are being used in pathology practice now, (2) describe the translational research cycle and draw attention to some of the challenges faced in delivering practice-changing discoveries, (3) discuss the impact of genomic biomarkers on the design of new clinical trials, and (4) briefly review the guidelines and recommendations for moving successful biomarkers into clinical practice.

Cancer-Associated Biomarker Categories

Personalized, i.e., patient-oriented, research refers to a continuum from initial studies in humans to comparative effectiveness and outcome research and the integration of this research into the health-care system and clinical practice. The goal of patient-oriented research is to optimize the translation of innovative diagnostic and therapeutic approaches to the point of care, as well as to help researchers meet the challenge of contributing to high-quality, cost-effective health care [4]. It involves ensuring that the right patient receives the right clinical intervention at the right time, ultimately leading to better health outcomes [5, 6]. In order to make patient-oriented care effective, there is a great need to discover more promising, reliable cancer-specific biomarkers and translate them successfully into clinical use.

In general, biomarkers are biological measurements that are used to aid clinical practice. The National Cancer Institute defines a biomarker as a "biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease" [7]. A biomarker may be used to see how well the body responds to a treatment for a disease or condition [8]. The Biomarkers

T. Jamaspishvili
Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada

Division of Cancer Biology and Genetics, Queen's University, Kingston, ON, Canada

J. A. Squire (✉)
Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada

Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_13

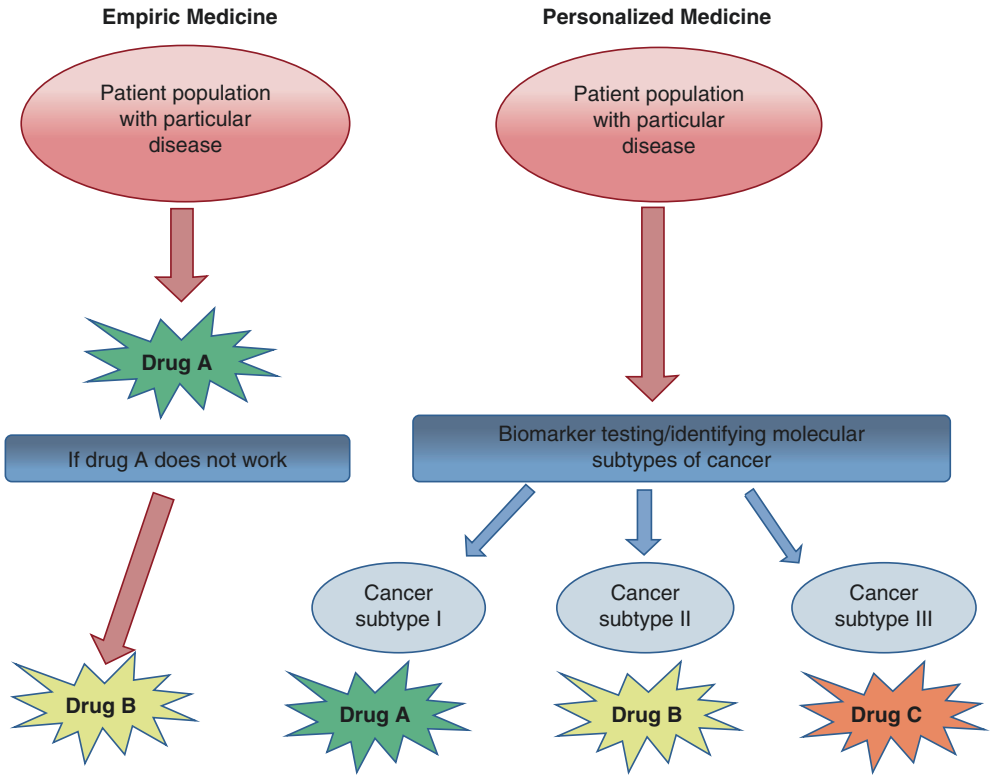


Fig. 13.1 Empiric treatment versus patient-oriented treatment. Individualized medicine is contrasted with the traditional "empiric method," which uses a standardized treatment for all patients with a certain disease

Consortium (managed by the Foundation of National Institutes of Health) states that "biomarkers are characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention" [9].

There are five different categories of cancer biomarker measurements that can be assayed either once at baseline (diagnostic, prognostic, and predictive) or repeatedly (disease screening, disease monitoring, and molecular imaging) during the course of the disease. A marker may belong to a single or to multiple biomarker categories.

A diagnostic biomarker is an indicator measurement that will aid in the detection of malignant disease in an individual. PSA (prostate-specific antigen) is the best-known cancer biomarker for early detection of prostate cancer. Serum PSA has been widely used for almost 25 years in screening for prostate cancer and has brought about a dramatic increase in early detection of the disease. Unfortunately, the low specificity of elevated serum PSA as a cancer biomarker results in a significant number of men who do not actually have prostate cancer undergoing unnecessary needle core biopsies [10, 11]. To address these concerns, the US Preventive Services Task Force (USPSTF) reconsidered the potential harms and relative benefits of using PSA as a screening biomarker. It was found that there was insufficient evidence to recommend routine use of PSA as a screening test at any age (see section "The Biomarker Development Process"). The PCA3 (prostate cancer antigen 3) RNA biomarker test has been introduced as a simple additional urine assay to address the significant diagnostic dilemma in new cases of prostate cancer [12, 13]. The specificity of this test in prostate cancer is 74% compared to only 21–51% (depending on grade) for serum PSA, which at least increases the potential for this type of assay in predicting the likelihood of a positive needle core biopsy [14–16]. Using a cutoff of 4.0 ng/mL, the PSA blood test has a sensitivity of 67.5–80% compared to 52% sensitivity for the PCA3 urine test. PSA is used for both as a diagnostic and a prognostic test after the USPSTF recommended against its routine use as a general screening biomarker, except in high-risk patients with a family history. Nowadays, PSA is more appropriately used as part of the diagnostic work-up on a new patient rather than as a primary screening test, though it can be still used for both purposes.

Screening biomarkers are an important subclass of biomarkers that must have high sensitivity and a good negative predictive value (specificity is less critical) in a clinical setting. These biomarkers are designed to robustly differentiate patients with disease from those without a disease. A perfect screening biomarker should have 100% sensitivity and 100% specificity, but at present none of the available biomarkers achieve these ideal performance standards. Another good example of a currently used screening biomarker is the widespread testing for HPV (human papillomavirus) DNA as part of cervical cancer screening programs. The HPV molecular test is more sensitive with a high negative predictive value than either conventional cytology (PAP smear) or

liquid-based cytology methods. An example is the cobas® HPV (Roche Molecular Systems, Inc.) DNA test, which has been used as an adjunct to conventional screening methods in the USA and in some European countries since 2011. In the ATHENA screening trial, this test was able to quantify the risk of precancer and cervical cancer in HPV 16+ and/or HPV 18+ women who either had atypical squamous cells of undetermined significance (ASC-US) or they had normal cytology [17]. In 2014, the FDA announced approval of the HPV DNA test as a primary screening method for cervical cancer for women over the age 24 [18].

Prognostic biomarkers are often defined as measurements made at diagnosis that provide information about patient prognosis. Prognostic biomarkers may predict disease recurrence (disease-free survival) and/or cancer-related death (cancer-specific survival) or overall survival for an individual patient in the absence of treatment or in the presence of standard primary treatment. Thus, prognostic markers typically give information about patient outcomes and tumor aggressiveness. For example, estrogen receptor (ER)-positive breast cancer patients have longer survival in the absence of systematic therapy than those patients who are ER negative [19]. CA125, which is present in a subset of ovarian cancers, is not used for detection of early cancers because the serum levels are elevated in only 50% of patients with stage I disease [20, 21]. This biomarker is usually used to evaluate response to chemotherapy, relapse, and disease progression in ovarian cancer patients. Gupta and Lis performed comprehensive evaluation of the existing literature on the prognostic role of CA125 and suggested that postoperative levels of serum CA125 are also a strong prognostic factor for estimating overall survival and progression-free survival in ovarian cancer [22].

Disease-monitoring biomarkers are assays that are performed repeatedly over time. A change in disease status during treatment will be reflected by a concomitant change in the biomarker status. Examples of biomarkers used for such monitoring are as follows: PSA in prostate cancer, CA125 in ovarian cancer, CEA in colorectal cancer, CA19–9 in pancreatic cancer, and CA15–3 or CA27.29 in breast cancer.

Predictive biomarkers are used to predict response or resistance to a specific cancer therapy, i.e., they are used to identify the patients who are likely or unlikely to benefit from a specific treatment. For example, in addition to its role as a prognosticator, tumor ER positivity is considered to be a predictive biomarker in breast cancer because such patients are far more likely to benefit from antiestrogen therapy such as tamoxifen. On the other hand, ER negativity is a predictive biomarker for benefit from conventional cytotoxic chemotherapy. Human epidermal growth factor receptor 2 (Her2/neu) amplification is a predictive marker for benefit from trastuzumab (Herceptin®), doxorubicin, and taxanes [23, 24]. In some situations, predictive biomarkers can be used to identify patients who may not benefit from a particular drug. For example, advanced colorectal cancer patients whose tumors have KRAS mutations are typically poor candidates for treatment with epidermal growth factor receptor (EGFR) antibodies [25, 26].

Cancer Genomics: From Research to Pathology Practice

The successful completion of the Human Genome Project stimulated a shift in emphasis from studying genes and proteins as individual biomarkers to current objectives to better understand their interactions in pathways of therapeutic importance. Thus, genomics, proteomics, transcriptomics, and metabolomics are now providing excellent opportunities for researchers to learn more about complex diseases like cancer by studying the overall response of cells to a mutation or to changes in the disease microenvironment. It is important to note that technologies that are used for biomarker discovery are often not exactly the same technologies that will be routinely used in a clinical laboratory. However, it is clear that discoveries made using genomic and proteomic technologies, coupled with advances derived from applied bioinformatics, are showing great promise for simpler and more cost-effective analysis of clinical samples.

Genomic Technologies Used for Biomarker Discovery

Gene Expression Arrays

Gene expression analysis has been one of the first high-throughput molecular profiling technologies with widespread adoption for biomarker discovery. Microarrays enable simultaneous analysis of tens of thousands of genes and thus the rapid identification of new potential biomarkers. Gene expression analysis measures the activity of cellular RNA (mRNA) in a tissue or bodily fluid at a given point in time, and it may provide information about the current status of a disease or the likelihood of future disease. RNA levels are dynamic and change as a result of pathology or environmental signals [27]. Certain patterns of gene activity may be used to diagnose a disease or to predict how an individual will respond to treatment over time. Methods used for gene expression analysis are diverse, ranging from real-time reverse transcription polymerase chain reaction (RT-PCR) to microarray screening technologies, which have been widely used in research, and are now beginning to be applied in clinical settings.

The most significant genomic biomarkers that have emerged in recent years include *BCR-ABL1* for CML

(chronic myeloid leukemia) diagnosis and monitoring of treatment responses [28], Her2/neu for diagnosis and prognosis of the breast cancer subtype which benefits from monoclonal antibody (trastuzumab [Herceptin®]) treatment [29], and detection of *EGFR* (epidermal growth factor receptor) and *KRAS* mutations for predictive purposes in lung [30] and metastatic colon cancer [31]. Discoveries from molecular profiling of RNA and DNA continue to generate many new candidate biomarkers that have potential similar to these successful genomic biomarkers.

The use of DNA expression microarrays has provided one of the most powerful tools to discover subsets of clinically important genes in human cancer [32]. Such expression arrays have been used to obtain major insights into progression, prognosis, and response to therapy on the basis of gene expression profiles (see the section on gene expression tests, below). Typically microarrays have been used to discover subsets of genes whose expression levels can be used to provide a distinct molecular subclassification of disease state. Once such a distinguishing genetic signature with likely clinical relevance has been discovered, custom-made arrays or other molecular biology methods are used to develop preclinical or clinical testing.

Genome-Wide Association Studies (GWAS)

GWAS is a comprehensive approach that identifies and correlates single-nucleotide polymorphisms (SNPs) to complex diseases such as cancers and is predominantly carried out with SNP microarrays specifically designed to interrogate millions of different polymorphisms in the human genome. GWAS is also very helpful as a biomarker discovery tool [32]. Results obtained from GWAS are typically cross-referenced with data from the HapMap Project or the 1000 Genomes Project in a process called imputation that aims to substitute values for missing data [33]. The advantage of GWAS is that it is unbiased and less likely to miss important genes or pathways than methods that use selected genes. Analysis of the large complex datasets generated by GWAS poses several challenges: (1) it requires large sample numbers and advanced bioinformatics to determine statistical significance; (2) there often remains a high likelihood of false-positive associations; and (3) with such marked biostatistical complexity, small differences may be missed due to stringent biostatistical corrections. With the introduction of high-throughput next-generation sequencing (NGS) into clinical medicine, diagnostic genomics is becoming an integral part of advanced molecular oncology. The USA recently launched the Precision Medicine Initiative in 2015 that includes a million patients as part of a multimillion dollar longitudinal cohort study to understand the hurdles and pitfalls of NGS-based applications and to accelerate the progress of personalized medicine [34, 35].

Next-Generation Sequencing (NGS)

The comprehensive screening power of NGS promises to help mine the remaining "unannotated regions" of the genome for novel sequence-based biomarkers that are below the resolution levels for detection by conventional microarray analysis [36]. In NGS all sequence information from a patient sample is aligned to a full-length reference genome to match all sequencing reads to their exact genomic locations [37]. Counting the number of sequencing reads that align to a given genomic location is analogous to microarray intensities for a probe with a specific sequence, and this metric can provide an estimate of relative expression levels. With slight modification in the NGS experimental design, DNA copy number, expression levels, and differential methylation can be determined. Sequencing technologies can further identify variation between samples by identifying genomic locations, whereas reads that do not perfectly match the reference genome may indicate individual genetic variation such as SNPs, loss of heterozygosity (LOH), as well as copy number variation (CNV) [38, 39].

While sequencing costs continue to decrease over time, costs associated with downstream data analysis are expected to grow by ~50% between 2010 and 2020 [40]. There are two types of NGS technology: (a) targeted sequencing of genes or so-called gene panel sequencing and (b) whole-exome (WES) or whole-genome sequencing (WGS) both for clinical management and for discovery of new disease-associated genes.

Gene panel sequencing can detect base-pair substitutions (gene mutations, SNPs), short insertions and deletions (indels), duplications or deletions of large chromosomal regions, and gene copy number changes. The advantage of targeted NGS is that the method works well with relatively low amount of DNA present in FFPE samples and provides high depth of coverage (up to 1000×), which makes it ideal for using in clinical laboratories. Such NGS panels have been designed for diagnostic, prognostic, and predictive purposes to detect and monitor regions of interests and specific gene sets. Although gene panel sequencing can detect CNVs, the method is not sufficiently sensitive for detection of low copy number changes or for evaluation of complex gene rearrangements [41, 42]. While whole-exome sequencing (WES) provides DNA sequence data of just the genome coding regions, whole-genome sequencing (WGS) provides full sequence data of all genome coding exons as well as all the intervening noncoding regions. Whole-genome sequencing looks at the genome more broadly allowing for a more accurate detection of genome rearrangements and is the most sensitive approach to characterize copy number changes that are often not evident with other sequencing approaches such as targeted sequencing. The disadvantage of this broader sequencing is the high cost of analysis and inability to capture intratumoral heterogeneity at sufficient depth.

In addition, data analysis and interpretation is the biggest drawback [43]. Although whole-exome and whole-genome sequencing are more comprehensive approaches compared to targeted sequencing, whole-exome sequencing covers only 1% of the genome that is translated into protein, and therefore, a large number of noncoding regions are ignored from analysis. A number of recent studies have demonstrated that mutations in noncoding regions may have direct tumorigenic effects, and therefore, future diagnostic genomics will need to move toward more complete 100% genome sequencing [44]. Current clinically available sequencing-based tests are discussed in section "Gene Expression and Sequencing-Based Tests".

Role of Bioinformatics and Genomic Datasets in the Public Domain

In order to facilitate the biomarker discovery process, it was recognized that there was a need for freely accessible datasets containing comprehensive information associated with DNA and with RNA expression. Most journals now require that investigators make such genomic data publically available in a standardized format for open access in silico analysis. All data must be MIAME (minimum information about a microarray experiment)-compliant. In other words, MIAME comprises the minimum requirements that should always be included with published microarray datasets, as suggested by the Functional Genomics Data Society (http://www.fged.org). The most popular genomic datasets are GEO, ONCOMINE, and ArrayExpress Archive, described below.

GEO (the Gene Expression Omnibus) is the biggest public repository that was designed to utilize features of the most commonly used molecular profiling methods today. These include data generated from microarray analyses as well as sequence technologies and include gene expression profiling, noncoding RNA profiling, chromatin immunoprecipitation (ChIP) profiling, genome methylation profiling, SNP genomic variation profiling, array comparative genomic hybridization (aCGH), serial analysis of gene expression (SAGE), and protein arrays (http://www.ncbi.nlm.nih.gov/geo/).

ONCOMINE is a cancer microarray database and Web-based data mining platform aimed at facilitating discovery from genome-wide expression analyses [45]. Using the ONCOMINE platform, researchers can easily compare gene expression profiles between cancer and normal samples; compare gene expression between different molecular, pathological, and clinical cancer subtypes; and investigate expression of genes in pathways and networks associated with cancer. It is possible to identify pathways, processes, chromosomal regions, and regulatory motifs activated in cancer and also search for genes that distinguish and predict cancer types and subtypes (http://www.oncomine.org).

ArrayExpress Archive/Gene Expression Atlas is a European database that contains functional genomic experiments including gene expression data. Here, researchers can query and download data collected according to MIAME and MINSEQE (minimum information about a high-throughput nucleotide sequencing experiment) standards. It is also an atlas that can be queried for individual gene expression under different biological conditions across experiments (http://www.ebi.ac.uk/arrayexpress).

Integration Approaches to In Silico Datasets

For in silico analysis, information is extracted from publicly available genomic datasets and then analyzed by the researcher using a computer to look for various patterns associated with particular diseases. In silico analysis can be applied, for example, to determine the location of mutations in a certain tumor suppressor gene, to look for copy number changes for particular genes, and to compare gene/protein expression patterns between cancerous and normal samples. Commercial bioinformatics software (such as Nexus™, BioDiscovery, Inc., California, USA, or Partek®, Partek Inc., Saint Louis, USA) enables users to manage, integrate, visualize, and analyze data generated from high-throughput gene expression analysis, aCGH, SNP arrays, and NGS datasets.

The advantages of in silico methods are that they are rapid and avoid the need for expensive experiments to evaluate a biomarker's clinical value. Moreover, bioinformatics permits the investigator to search for a biomarker in one dataset and attempt to validate it in another. However, the utility of in silico analysis depends on the quality of the clinical data collected, as well as the coverage and accuracy of the annotations used to report the genomic data. It can also be difficult to compare results across several datasets because of the differences in genomic methods. For these reasons, in silico analysis in biomarker discovery is often considered an initial step that must be followed by rigorous experimental validation prior to preclinical investigation.

Clinically Applicable Gene-Based Assays

A very important aspect of marker development is to translate it to the clinic, once its usefulness has been established. A potential marker can be tested in different sources, including tumor tissues and body fluids such as serum and urine. The methods used should be of rapid execution, reliable, and ideally not very expensive. As our understanding of complex diseases grows, additional biomarkers are being identified and developed into new and improved diagnostic tools that can analyze multiple biomarkers simultaneously. Often, such

biomarker assays establish a complex molecular profile of the disease and provide an estimate of the likelihood of a response to a given treatment. They combine the values of multiple variables to yield a single patient-specific result. Such multigene assays commonly use PCR tests or gene expression microarrays, the results of which are integrated into an algorithm to organize and prioritize individual markers, thereby producing a readily accessible result [46]. The common examples of this modality are discussed below and some are already FDA cleared or approved.

Gene Expression and Sequencing-Based Tests

In spite of the fact that microarray technologies are costly, gene expression tests are increasingly being implemented in modern clinical practice as an aid to conventional diagnostic, prognostic, and predictive decision tools used in cancer management. Some of the most recently used examples are discussed below.

ColoPrint® (Agendia, Amsterdam, the Netherlands) is a microarray-based gene expression profile used to predict the risk of distant recurrence of stage II and III colon cancer. ColoPrint® combines a multigene panel, which includes seven colon cancer-related genes and five reference genes, with a proprietary algorithm for determining risk of recurrence (http://www.agendia.com). ColoPrint uses the same technology, methods, and quality control as FDA-cleared assays (i.e., MammaPrint®), though it is not approved by the FDA. Similarly, Genomic Health, Inc. provided the Oncotype DX® colon cancer test for stage II colon cancer patients by evaluating expression levels of 12 genes. The results of the test are reported as a quantitative Recurrence Score® result, which is a score between 0 and 100 that correlates with the likelihood of a person's chances of having the cancer return [47]. At present this test it is not FDA approved. The assay is only performed by the developers in their Clinical Laboratory Improvement Amendments (CLIA) commercial laboratory. Genomic Health also provides MMR (mismatch repair) testing by immunohistochemistry on colon tumor samples, which, in combination with Oncotype DX®, may help the clinician in making treatment decisions (http://www.oncotypedx.com). Stage II colon cancer patients with MMR-deficient (MMR-D) tumors have a much lower risk of recurrence compared to patients with MMR-proficient (MMR-P) tumors [48].

BluePrint® is an 80-gene expression signature which classifies breast cancer into basal-type, luminal-type, and ERBB2-type cancers. The BluePrint® molecular subtyping profile, combined with the patient's MammaPrint® (see below) test results, provides a greater level of clinical information to assist in therapeutic decision-making (http://www.agendia.com). BluePrint® does not require FDA clearance because it is considered a class I, low-risk device under FDA regulations.

MyPRS™/MyPRS Plus™ (my prognostic risk signature) is a tool for guiding treatment in patients with multiple myeloma. It analyzes all of the nearly 25,000 genes in a patient's genome to determine the gene expression profile (GEP) that is associated with a particular patient's condition. The GEP is made up of the 70 most relevant genes (GEP70) which aid in the prediction of the patient's outcome (http://www.signalgenetics.com). Both MyPRS™ and MyPRS Plus were developed by Myeloma Health, LLC, who determined performance characteristics in a CLIA-certified laboratory. The FDA has indicated that these tests do not require either clearance or approval at present.

MammaPrint® (Agendia, Amsterdam, the Netherlands) is based on microarray technology using 70 cancer-related and about 1800 non-cancer-related genes (http://www.agendia.com). The test stratifies patients into two distinct groups: low risk or high risk for distant recurrence, with no intermediate-risk patients. With low-risk patients, hormonal therapy (e.g., tamoxifen) might be sufficient, avoiding the necessity of aggressive treatment such as chemotherapy. The test was cleared by the FDA as a class II device in 2007. However, the FDA did not evaluate treatment outcomes as a result of use of this "prognostic" device. In addition, the EWG (the Evaluation of Genomic Applications in Practice and Prevention [EGAPP] working group) found that "data were adequate to support an association between the MammaPrint Index and 5 or 10 year metastasis rates, but the relative efficacy of testing in ER-positive and ER-negative women is not clear." Also, study subjects were European, and how characteristics of other demographic populations might affect test performance is not known [49]. The MINDACT (Microarray In Node-Negative Disease May Avoid Chemotherapy Trial) is designed to compare the effectiveness of MammaPrint test results versus clinical evaluation in predicting 15-year disease-free survival and overall survival (EORTC (European Organization for Research and Treatment of Cancer), MINDACT 2008). This trial will compare clinical response to endocrine therapy alone versus endocrine therapy combined with chemotherapy regimens (anthracycline-based, docetaxel-capecitabine, letrozole).

The Oncotype DX® breast cancer test (Genomic Health, Inc., Redwood City, CA) uses RT-PCR to study gene expression profiles in formalin-fixed, paraffin-embedded (FFPE) breast cancer tissues. Oncotype DX analyzes expression of 21 genes, 16 cancer related and 5 normative [50]. The test is intended for stage I or II, lymph node-negative, and ER-positive breast cancer patients, who will be treated with tamoxifen. Results are reported as a Recurrence Score™ (RS; scale of 0–100). Patients are divided into low-, intermediate-, and high-risk categories. Oncotype DX® claims to provide information beyond conventional risk assessment tools, including how likely the woman is to benefit from chemotherapy in addition to tamoxifen therapy (http://www.genomichealth.com). The TAILORx (Trial Assigning Individualized Options for Treatment) trial was

designed to determine the benefit of chemotherapy for women with intermediate risk. The trial has shown that gene expression test could identify women with a low risk of recurrence who could be spared chemotherapy [51]. The test is not FDA cleared but is available at the Genomic Health, Inc. CLIA-certified laboratory.

The most extensively studied tests among those listed above are Oncotype DX® breast cancer and MammaPrint®. In many countries these new tests are being offered for clinical use, but there remains a need for more comprehensive long-term studies to assess whether test outcomes lead to clear beneficial effects for patients and are cost-effective.

There are also a number of sequence-based gene panel tests that have been developed recently that provide precise information on mutations of clinical importance. These include clinical tests of germ line DNA for the risk of hereditary disorders and tests of tumor DNA for therapeutic decision-making in cancer [52].

The hotspot panel is a collection of frequently mutated hotspots that are either therapeutically actionable or with diagnostic/prognostic significance. There are two types of hotspot cancer panels currently commercially available to guide for treatment: one for the choice of therapy and the other for the amount of medication.

The AmpliSeq™ Cancer Panel v1, developed by the Life Technologies, covers 739 clinically relevant hotspot mutations from 46 cancer genes, including well-established tumor suppressor genes and oncogenes. The similar panel from ThermoFisher (Ion AmpliSeq™ Cancer Panel v2) has become very popular as a clinically validated test that is compatible with FFPE samples, and it has been adopted by many academic institutes and private laboratories in North America [52]. PGxOne™ developed by Admera Health is a hotspot panel (http://www.admerahealth.com/pgxone/), which screens for 152 frequently mutated sites from 13 well-established pharmacogenomics genes that affect drug absorption, metabolism, or activity. The data from the panel provide information for physicians to prescribe appropriate doses for effective treatment based on the presence of specific actionable mutations. Several institutions offer similar panels as lab-developed procedures performed in CLIA-certified laboratories.

The disease-focused panels are designed to detect germ line mutations to screen for the risk of inherited diseases or to diagnose suspected genetic diseases in carriers. The hereditary cancer panels are widely used tests since approximately 5–10% of all cancers are considered to be hereditary. More than 100 cancer susceptibility syndromes have been reported, including hereditary breast and ovarian cancer syndrome, Lynch syndrome, Cowden syndrome, and Li-Fraumeni syndrome. Today around 227 tests are available for hereditary cancer screening in clinical laboratories.

Comprehensive panels include all genes associated with all diseases. Illumina's TruSight One is an example of such a comprehensive panel. This panel includes more than 60 well-established subpanels and covers 4813 genes having known association with clinical phenotypes. Such panels minimize test development and validation efforts and enables physicians to request testing for specific disease(s) if clinically indicated, without any additional efforts.

Whole-genome sequencing is the most comprehensive tool for future clinical application. It can provide full coverage of all protein-coding regions like WES as well as intronic and other noncoding regions associated with inherited diseases. With the recent release of Illumina HiSeq X Ten, a human genome can be sequenced at 30x coverage under $1000 for the wet lab portions of the analysis.

Protein Chips

Similar to using DNA chips for identification of gene expression profiles in particular tumors, the advent of "protein chips," which enables the analysis of thousands of proteins expressed by a single tumor sample at the same time, has helped researchers to better understand the molecular basis of disease, including disease susceptibility, diagnosis, progression, and potential points of therapeutic interference. The basic format of most protein chips is similar to that of DNA chips, such as the use of glass or plastic printed with an array of molecules (e.g., antibodies) that can capture proteins. Ideally, a protein chip would be able to predict a cancer state by a simple serum or urine test. This technology is likely to see considerable additional development and application in the coming years [53].

Fluorescence In Situ Hybridization (FISH)

Quantification of multiple mRNA levels in tumors is expensive, technically demanding, and not readily available in a routine clinical setting. FISH provides an alternative way to diagnose and identify predictive or prognostically important genetic alterations. The method is simple, fast, and reliable and therefore has been widely accepted for clinical use in human cancer. It is used to assess various genetic alterations (amplifications, deletions, translocations). FISH can detect genomic anomalies over a much greater dynamic size range than other techniques. In the past decade, the technique has been developed to include multicolor FISH assays so it is now possible to assess complex genomic alterations [54]. Recent improvements have been made to FISH in the form of chromogenic in situ hybridization (CISH) and silver-enhanced in situ hybridization (SISH). These techniques use peroxidase enzyme-labeled probes whose signals do not decay over time and allow the specimen to be viewed using bright-field microscopy. CISH and SISH have been used to assess Her2/neu gene status [55].

Assessment of Her2/neu amplifications in breast cancer, to assess prognosis and to predict treatment outcome, is the most common example of FISH use in clinical settings [56]. Other examples include the recently developed commercialized test eXagenBC. The latter promises to provide a tailored prognosis in node-positive and node-negative breast cancer patients and is based on assessment of DNA copy numbers of three genomic regions (around *CYP24*, *PDCD6IP*, and *BIRC5*) for ER-positive and progesterone (PR)-positive tumors and three different genes (*NR1D1*, *SMARCE1*, and *BIRC5*) for ER-negative and PR-negative tumors in both node-negative and node-positive patients. The eXagenBC test uses a prognostic index (PI) from an algorithm to integrate the information from the three genes and predict recurrence rates. This test may provide greater accuracy compared to other criteria for recurrence risk assessment and therefore has been suggested for routine clinical use [57].

Additional promising prognosticators are fusion genes such as *TMPRSS2-ERG* translocations and *PTEN* deletions in prostate cancer which show great promise for identification of aggressive prostate cancers. *PTEN* deletions have been associated with earlier biochemical relapse following radical prostatectomy. Prostate cancers showing homozygous *PTEN* deletions, termed "*PTEN* null," have been strongly associated with metastasis and androgen-independent progression, i.e., castration-resistant prostate cancers (CRPC) [58–60]. One important new FISH biomarker is the echinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase (*EML4-ALK*) fusion gene, present in a small subset of non-small-cell lung cancers (NSCLC). Such tumors are particularly sensitive to ALK inhibitors such as crizotinib which has been approved by the FDA in 2011 for the treatment of locally advanced or metastatic non-small-cell lung cancers that are ALK-positive [61, 62]. The FDA also approved the Vysis ALK Break Apart FISH Probe Kit (Abbott Molecular, Inc.) that is a diagnostic test designed to detect rearrangements of the *ALK* gene in NSCLC [63].

Polymerase Chain Reaction (PCR)

Clinical diagnostic applications of real-time PCR or real-time quantitative PCR (qPCR) have been widely implemented by hospital-based clinical laboratories [64]. In translational research, qPCR is simple and one of the fastest, most reliable and cheapest molecular techniques for the validation of a newly discovered biomarker. A qPCR assay can be used to identify gene amplifications, deletions, fusions, overexpression, and mutations down to single base changes, and therefore, these very sensitive and specific molecular tests are among the most widely used methods to translate recent discoveries in cancer research into clinical practice.

Examples of clinically applicable qPCR assays in cancer diagnostics and prognostics include the detection of *BCR-ABL1* transcripts in patients with chronic myeloid leukemia (CML) who are then subjected to tyrosine kinase inhibitor (imatinib [Gleevec®]) treatment as a first-line therapy and to quantification of minimal residual disease (MRD) by qPCR [65]. Recently highly sophisticated methods have been developed using DNA-based and RNA-based PCR assays for the detection of *BCR-ABL1* transcripts that were previously not detectable by conventional PCR methods [66, 67]. Thyroid cancer is another example where qPCR assays play an important role: in this case they have a diagnostic and predictive role. Real-time PCR can be used to diagnose papillary thyroid carcinomas (PTCs) harboring a point mutation in *BRAF* or *RAS*, or a *RET-PTC* rearrangement (>70%), and they can help diagnose follicular thyroid carcinomas (FTCs) that harbor either *RAS* mutations or *PAX8/PPARγ* rearrangements [68]. *RAS* mutations may also be found in benign thyroid lesions. In addition, sporadic and hereditary medullary thyroid carcinomas (MTCs) are both associated with point mutations in the *RET* gene. Thus, molecular testing is now an important component of thyroid cancer diagnosis and management [68, 69].

Assays that simultaneously amplify (or detect) two or more target fragments (or detect sequence changes within target fragments) are termed duplex and multiplex real-time PCRs, respectively. It is noteworthy that the multiplexing of biomarkers has many advantages over single biomarker measurements, especially when trying to identify the best diagnostic or prognostic models for various human cancers (prostate cancer, as an example, is discussed below) [70]. One commercially available real-time PCR assay (HemaVision, DNA Technology, Aarhus, Denmark) is widely used in clinical laboratories to simultaneously detect 28 fusion genes and more than 80 breakpoints and splice variants in patients with acute myelogenous leukemia (AML) and acute lymphoid leukemia (ALL) ([71]; http://www.biocompare.com).

Classical cytogenetic methods (e.g., conventional karyotyping) continue to provide well-established diagnostic findings to clinicians. However, the detection of certain genetic abnormalities (translocations or fusion genes) that often have been missed by conventional cytogenetics is now feasible with high reliability using newer molecular techniques that have advantages over traditional methods. These may include a shorter turnaround time, automated analyses, and a lack of the prior requirement of dividing cells [72].

Impact of Genetic Biomarkers on Drug Development and Clinical Trial Designs

Genetic biomarkers now have tremendous impact in every phase of drug development, from drug discovery to preclinical evaluations through each phase of clinical trials and into

routine clinical use [73]. In the early phases of drug development, biomarkers are used to evaluate the activity of small molecule therapeutics in animal models, to investigate mechanisms of action and to provide essential preclinical data needed for the various later stages of clinical trials. If the preclinical phase of drug development is successful, then it is followed by an application to the FDA as an investigational new drug (IND). The purpose of an IND is "to ensure that subjects will not face undue risk of harm" in a clinical investigation that involves the use of a drug. The IND is the mechanism by which the investigator, or pharmaceutical sponsor, provides the requisite information to obtain authorization to administer an investigational agent to human subjects [74]. By doing so, the compound can be tested for dose response, efficacy, and toxicity. After an IND is approved, the next steps are clinical phases 1, 2, and 3. Phase 1 trials determine safety and dosage and identify side effects (patient number: 20–80); phase 2 trials are used to obtain an initial assessment of efficacy and to further explore safety of the drug or treatment in a larger number of patients (100–300); and in phase 3 trials, the treatment is given to large groups of patients (>1000) to confirm effectiveness, monitor side effects, compare efficacy to established treatments, and collect information that will allow it to be used safely.

In clinical trials which are designed to validate and assess the usefulness of a prognostic or predictive biomarker, the major issues are to obtain sufficient statistical evidence of treatment benefit in patients who are positive for the predictive or prognostic biomarker and then to examine the biological relationships associated with the biomarker's expression and the molecular pathways targeted by the therapeutic agent. Often, such studies utilize a retrospective analysis of a biomarker in available tissues from patients with known response who have been treated similarly [75]. Before initiating studies to confirm the clinical utility of a novel biomarker, it is necessary to conduct validation trials in which several criteria must be met. First, specific testable hypotheses must be proposed based on scientific evidence of the predictive properties of the putative biomarker relative to the existing (standard) treatment. In addition, any prognostic benefit is assessed as well. A novel biomarker is considered promising for clinical utility when it demonstrates the following features in the validation study: (1) the marker is independently associated with clinical outcome; (2) its biological effects are specific for the cancer of interest as opposed to normal tissues, other disease states, or other cancers; (3) the marker's prevalence in the target population is high; and (4) the methods of marker measurement are feasible and reproducible.

In the next phase of the evaluation of clinical utility of the predictive or prognostic biomarker, two major issues have to be considered: the selection of an appropriate patient population and the choice of the most appropriate end point. For example, when evaluating predictive markers of therapeutic efficacy in the adjuvant setting, the primary end point usually is overall, disease-free, or recurrence-free survival. Possible primary end points for metastatic disease trials would include response rate, time to progression, survival, or risks of toxicity [75].

With respect to clinical trial designs for new drugs or treatment options and companion biomarkers, randomized controlled trials (RCT) are the most popular, because they limit the potential for bias by randomly assigning one arm to an intervention and the other arm to nonintervention (or placebo). This minimizes the chance that the incidence of confounding (particularly unknown confounding) variables will differ between the two groups. Currently, some phase 2 and most phase 3 drug trials are randomized, double-blind, and placebo-controlled. Traditional RCT designs are not always well suited for drugs with molecular targets and associated biomarkers. Newer clinical trial designs have incorporated the recent discoveries of molecular oncology [76]. These trial designs are much more efficient because study arms are enriched based on mutational profiles associated with a specific actionable drug response. For example, the standard randomized approach in a clinical trial for trastuzumab would not be very effective without the use of an enrichment design, because the drug has little effect on Her2/neu-negative patients. Because almost 75% of patients are Her2/neu negative, a standard design would require a large sample size to detect the treatment effect of trastuzumab on Her2/neu-positive patients. An enrichment clinical trial design is used to evaluate a treatment or a drug in which the effect can be readily demonstrated on a specific subset of the study population. Often such a subset is identified by a biomarker test that is used to select those patients who are likely to respond well to the treatment. Efficiency of the study thus depends on the prevalence of test-positive patients and on the relative effectiveness of the new treatment in test-negative patients [76]. In the enrichment designs, the number of randomized patients is often substantially smaller than for a standard design.

Another new type of clinical trial is the "basket" phase 2 design, which is based on the idea that the presence of a molecular marker will predict response to targeted therapies, independently of tumor histology. Basket trials can be nonrandomized or randomized and can include a single drug or multiple individual drugs [76]. The MATCH (Molecular Analysis for Therapy Choice) clinical trial, launched by the National Cancer Institute, opened with 400 clinical sites and 10 drugs is an example of a large multidrug basket design [77]. In this trial, more than 3000 patients with advanced metastatic cancer of many histologic types have been genomically tested with a common platform and triaged to a nonrandomized substudy with an actionable drug.

The "umbrella" trial design is a similarly innovative approach that takes patients with the same type of cancer and assigns them to different arms of a study based on their

mutations and the availability of a targeted therapy. The BATTLE I (Biomarker-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination) phase 2 trial for patients with non-small-cell lung cancer is an example of a phase 2 umbrella trial [78]. In this trial, patients' samples were assayed for four candidate biomarkers based on genomic or transcriptomic alterations. The patients were then randomly assigned to receive one of the four drug regimens. The analysis of this trial was the same as for the randomized basket designs, but in the umbrella design, the conclusion about whether targeting was useful was limited to patients with the single selected primary site of disease.

The Translational Research Continuum

Despite the rapid pace of biomarker discovery in recent years, there are still very few validated genetic biomarkers of proven and robust clinical utility [79]. This poor performance reflects that the clinical development of new biomarkers is just as difficult as the development and approval of a new drug. Here we will outline the bench to bedside pipeline and discuss how best to facilitate the successful development of biomarkers and molecular targeted treatments, respectively. Throughout the cancer research process, many challenges are faced during the transition of a new discovery from the "research bench" through the phases of laboratory and clinical validations. Unfortunately the majority of "exciting discoveries" never succeed in overcoming the rigorous evaluations and are not accepted as part of routine clinical practice or used for laboratory testing by pathologists (Fig. 13.2).

Challenges in Preclinical and Clinical Research

A major factor contributing to the lack of use of genetic biomarkers in clinical trials is the poor quality of published preclinical data. This has been the focus of a recent commentary by Begley and Ellis [3]. IND trials rely heavily on the literature and on having a comprehensive understanding of the agent's target, its associated biomarker, and the various downstream consequences of the drug. Very often, however, the biological hypothesis around a new agent and its companion biomarker is uncertain or questionable. The lack of reproducibility of preclinical "research assays" when applied to patient samples may prevent the application of novel biomarkers in a clinical setting. Some of the issues that are considered to be associated with poor uptake of research biomarkers by trialists and clinical laboratories are summarized in Table 13.1.

The Biomarker Development Process

The biomarker development process requires multiple collaborative mechanisms, knowledge networks, and consortia to facilitate biomarker fruition in clinical practice. The critical limitation in biomarker development is the lack of a proper structure in the biomarker discovery process as is present in the process of testing a new drug. After proving, among other things, the clinical validity and clinical utility of a newly discovered biomarker (see below), a biomarker is not considered "validated" and cannot be recommended for use in clinical practice until independent research groups at

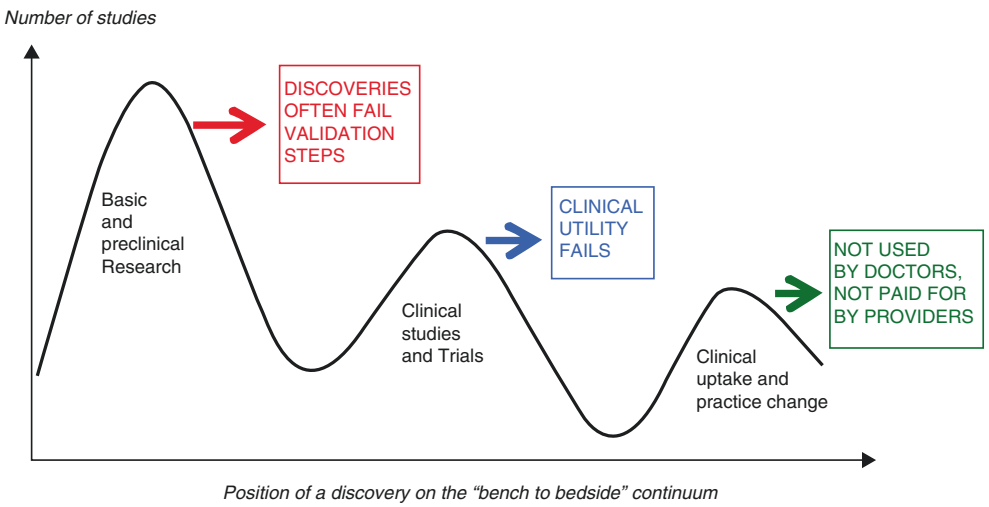


Fig. 13.2 The translational research continuum. This graph schematically depicts the three major obstacles that impede an exciting research discovery (*leftmost* peak) moving though the validation phase from preclinical research into clinical trials (*middle* peak) and onto clinical or laboratory practice (small peak on *right*). The graph illustrates the continuing gap between basic biomedical research and clinical research and knowledge. This gap limits the capacity to translate the results of provocative discoveries generated by basic biomedical laboratory research to the bedside, as well as to successfully engage and educate health-care providers in the benefits of the discoveries

Table 13.1 Challenges in preclinical and clinical research

Challenges in preclinical research:
Research staff does not use SOPs (standard operating procedures) or operate following GLP (good laboratory practice) standards.
Biased comparison groups in the study (case versus controls).
Statistically underpowered study size, inappropriate statistical analyses, including over-fitting of data.
Challenges in clinical research:
Independent groups are unable to generate concordant results due to the lack of coordination between biomarker research laboratories/ lack of standardized protocols across laboratories.
Lack of "good-quality samples." so-called convenience samples (from local bio-repository) may be too homogeneous to provide evidence for clinical relevance of biomarker to the whole population of the patients.
Clinical heterogeneity often leads to wrong conclusions.
New testing technologies lack appreciation of interlaboratory performance, standardization, quality control, and cost-effectiveness and cannot be used widely by clinical laboratories (e.g., mass spectrometric protein profiling).
Lack of pre-analytical studies.
Lack of funding for translational research.

multiple sites have demonstrated concordant results in separate trials. The challenge is firstly to determine which data are required to perform these studies and, secondly, to obtain, share, and pool these data together and to provide adequate support to analyze the pooled datasets. A solution would be to apply uniform standards, which should facilitate effective translation of newly discovered biomarkers to the clinical setting. Therefore, numerous collaborative mechanisms, knowledge networks, and consortia have emerged in order to facilitate biomarker discovery and enhance the delivery process to the clinic. Examples of such mechanisms such as the Early Detection Research Network (EDRN) and The Biomarkers Consortium (TBC) demonstrate the value of a national coordinated approach [80, 81].

Guidelines (known as the Standards for Reporting of Diagnostic Accuracy, or STARD statement) have been developed for diagnostic studies and were inspired by CRGs (Cochrane Review Groups) in 1999. For prognostic studies, guidelines known as REMARK criteria were developed by NCI-EORTC (National Cancer Institute-European Organisation for Research and Treatment of Cancer) [82–84]. The STARD initiative aims to improve the reporting quality and diagnostic accuracy of publications describing new biomarkers. The statement consists of a checklist of 25 items, and the decision to include items in the checklist was based on evidence linking these items to either bias, variability in results, or limitations of the applicability of results to other settings [82]. The checklist can be used to verify that all essential elements are included in the report of a research study.

REMARK (REporting recommendations for tumor MARKer prognostic studies) guidelines were developed primarily for studies of prognostic markers, especially those evaluating a single tumor marker while possibly adjusting for other known prognostic factors. The guidelines suggest relevant information that should be provided about the study design, preplanned hypotheses, patient and specimen characteristics, assay methods, and statistical analysis methods [83].

While some biomarkers have already been approved by the FDA, the use of others has been recommended in clinical guidelines by various cancer societies [5]. A recent example of this is a test for epidermal growth factor receptor (*EGFR*) mutation in patients with advanced NSCLC, which determines whether or not first-line EGFR tyrosine kinase inhibitor therapy is indicated [5, 85]. The introduction of biomarkers into routine clinical practice is considered in the framework tumor marker utility grading system (TMUGS) which was designed to evaluate the clinical utility of tumor markers and to propose a hierarchy of "levels of evidence" that might be used to determine if available data support the use of a marker or not [86]. TMUGS provides guidelines to determine the clinical utility of known and future tumor markers, as well as guidance on biomarker assay design, interpretation, and use in clinical practice. This evidence scale has been widely cited and used for deciding whether to recommend the use of a tumor marker in clinical practice and for design and conduct of tumor marker studies [87, 88]. This evidence scale has recently been revised to distinguish data generated from prospective clinical trials, in which the marker is the primary objective of the study, from those in which archived specimens are used [1, 75, 89]. Starting in 2000, the Office of Public Health Genomics (OPHG) at the Centers for Disease Control and Prevention (CDC) established the analytic framework ACCE Model Project based on four main criteria for evaluating a genetic tests:

1. *Analytic validity* is a component of clinical validity (see below) describing how accurately and reliably the test measures the genotype of interest. Analytic validity assesses technical test performance and includes analytic sensitivity (detection rate), analytic specificity (false-positive rate), reliability (repeatability of test results), and assay robustness (resistance to small changes in pre-analytic or analytic variables).
2. *Clinical validity* describes the accuracy with which a test predicts a particular clinical outcome and clearly separates two subgroups of patients with different outcomes within a large population. When a test is used diagnostically, clinical validity measures the association of the test with the disorder [90], and when used predictively, it measures the probability that a positive test will result in the appearance of the disorder within a stated time period.
3. *Clinical utility* is a balance of benefits and harms when the test is used to influence patient management, i.e., the evidence that the use of the marker improves outcomes

compared to not using it. Evaluation of clinical utility factors and the available information about the effectiveness of the interventions for people who test positive and the consequences for individuals with false-positive or false-negative results.

4. *Ethical, legal, and social implications* (ELSI) refer to other implications which may arise in the context of using the test and cut across clinical validity and clinical utility criteria. In 2004, a new initiative, termed EGAPP™ (evaluation of genomic applications in practice and prevention) was created by OPHG at the CDC "to better organize and support a rigorous, evidence-based process for evaluating genetic tests and other genomic applications that are in transition from research to clinical and public health practice in the U.S." [49, 91].

The US Preventive Services Task Force (USPSTF) is an independent panel of non-federal experts in prevention and evidence-based medicine and is composed of primary care providers. The USPSTF strives "to make accurate, up-to-date, and relevant recommendations about preventive services in primary care. It conducts scientific evidence reviews of a broad range of clinical preventive health care services (such as screening, counseling, and preventive medications) and develops recommendations for primary care clinicians and health systems" (http://www.uspreventiveservicestaskforce.org). These recommendations are published in the form of "Recommendation Statements." Also, the USPSTF stratifies the evidence by quality about the effectiveness of treatments or screening by three different levels (Table 13.2). For example, in 2002, USPSTF deemed the evidence to be insufficient to recommend routine use of PSA as a screening test among men younger than age 75. The recommendation, however, does not include the use of PSA test for surveillance after diagnosis or treatment of prostate cancer. The USPSTF reviewed the available evidence again in 2011 and in a draft report concluded that population benefit from PSA screening was inconclusive, recommending against PSA-based prostate cancer screening at any age [92, 93]. The USPSTF makes evidence-based recommendations about clinical preventive services such as screenings, counseling services, or preventive medications. Currently the majority of USPSTF recommendations are not in favor of widespread use of cancer screening using biomarkers. However, as more DNA-based biomarkers are developed, it seems likely that the benefits of screening may outweigh the risks for some of the diseases where early intervention can prevent disease progression (http://www.uspreventiveservicestaskforce.org/uspstopics.htm#AZ).

Table 13.2 Stratification of evidence by quality [94]

Level 1:	Evidence obtained from at least one properly designed randomized controlled trial
Level 2-a:	Evidence obtained from well-designed controlled trials without randomization
Level 2-b:	Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group
Level 2-c:	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence
Level 3:	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

Conclusions

Various consortia, grading systems, and collaborative initiatives discussed in this chapter are basically founded and developed in North America and are part of the goal to provide evidence-based medicine, which seeks to assess the strength of the evidence of risks and benefits of treatments, diagnostic tests, and biomarkers. Similar systems exist in Europe though they are not discussed here. The development and application of high-throughput sequencing have led to the precision medicine initiative in cancer. At the same time, radical changes in clinical trial design, combined with accelerated biomarker development, suggest there will be greatly improved response rates for patients and reduced cancer mortalities for many more tumor types. Networking infrastructures throughout the world developed to date have a goal of sharing and pooling analyzed data to complete the biomarker discovery → development → validation continuum. Increased collaboration between such consortia will continue to accelerate biomarker development and the use of genomics in clinical oncology. Global harmonization of guidelines in the years ahead will likely underpin the success of biomarker translation from bench to bedside.

References

1. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst. 2009;101:1446–52.
2. Brooks JD. Translational genomics: the challenge of developing cancer biomarkers. Genome Res. 2012;22:183–7.
3. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. Nature. 2012;483:531–3.
4. Strategy for Patient-Oriented Research-CIHR. http://www.cihr-irsc.gc.ca/e/41204.html. Accessed 10 Feb 2013.
5. Ziegler A, Koch A, Krockenberger K, et al. Personalized medicine using DNA biomarkers: a review. Hum Genet. 2012;131:1627–38.
6. Henney AM. The promise and challenge of personalized medicine: aging populations, complex diseases, and unmet medical need. Croat Med J. 2012;53:207–10.

7. Madu CO, Lu Y. Novel diagnostic biomarkers for prostate cancer. J Cancer. 2010;1:150–77.
8. NCI dictionary of cancer terms-National Cancer Institute. http://www.cancer.gov/dictionary?cdrid=45618. Accessed 10 Feb 2013.
9. Foundation for the National Institutes of Health. http://www.fnih.org/work/key-initiatives/biomarkers-consortium. Accessed 10 Feb 2013.
10. Rao AR, Motiwala HG, Karim OM. The discovery of prostate-specific antigen. BJU Int. 2008;101:5–10. https://doi.org/10.1111/j.1464-410X.2007.07138.x.
11. Barry MJ. PSA testing for early diagnosis of prostate cancer. N Engl J Med. 2001;344:1373–7.
12. Haese A, de la Taille A, van Poppel H, et al. Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. Eur Urol. 2008;54:1081–8.
13. Kirby RS, Fitzpatrick JM, Irani J. Prostate cancer diagnosis in the new millennium: strengths and weaknesses of prostate specific antigen and the discovery and clinical evaluation of prostate cancer gene 3 (PCA3). BJU Int. 2009;103:441–5.
14. Day JR, Jost A, Reynolds MA. PCA3: from basic molecular science to the clinical lab. Cancer Lett. 2011;301:1–6.
15. Laxman B, Morris DS, Yu J, et al. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. Cancer Res. 2008;68:645–9.
16. Jamaspishvili T, Kral M, Khomeriki I, et al. Quadriplex model enhances urine-based detection of prostate cancer. Prostate Cancer Prostatic Dis. 2011;14:354–60.
17. cobas® 4800 HPV Test [package insert, CE]. Branchburg, NJ: Roche Molecular Systems, Inc; 2012.
18. Practice Bulletin No. 157. Obstet Gynecol. 2016;127:e1–e20.
19. Kiba T, Inamoto T, Nishimura T, et al. The reversal of recurrence hazard rate between ER positive and negative breast cancer patients with axillary lymph node dissection (pathological stage I-III) 3 years after surgery. BMC Cancer. 2008;8:323.
20. Nossov V, Amneus M, Su F, et al. The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125? Am J Obstet Gynecol. 2008;199:215–23.
21. Nustad K, Bast RC Jr, Brien TJ, et al. Specificity and affinity of 26 monoclonal antibodies against the CA 125 antigen: first report from the ISOBM TD-1 workshop. Int Soc Oncodev Biol Med. 1996;17:196–219.
22. Gupta D, Lis CG. Role of CA125 in predicting ovarian cancer survival – a review of the epidemiological literature. J Ovarian Res. 2009;2:13.
23. Gennari A, Sormani MP, Pronzato P, et al. HER2 status and efficacy of adjuvant anthracyclines in early breast cancer: a pooled analysis of randomized clinical trials. J Natl Cancer Inst. 2008;100:14–20.
24. Hayes DF, Thor AD, Dressler LG, et al. HER2 and response to paclitaxel in node-positive breast cancer. N Engl J Med. 2007;357:1496–506.
25. Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26:1626–34.
26. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. Per Med. 2010;7:33–47.
27. UnitedHealthcare®. Medical Policy 2012; PDF. Accessed 18 Feb 2013.
28. Tohami T, Nagler A, Amariglio N. Laboratory tools for diagnosis and monitoring response in patients with chronic myeloid leukemia. Isr Med Assoc J. 2012;14:501–7.
29. Leong AS, Zhuang Z. The changing role of pathology in breast cancer diagnosis and treatment. Pathobiology. 2011;78:99–114.
30. Kim ST, Sung JS, Jo UH, et al. Can mutations of EGFR and KRAS in serum be predictive and prognostic markers in patients with advanced non-small cell lung cancer (NSCLC)? Med Oncol. 2013;30:328.
31. Laszlo L. Predictive and prognostic factors in the complex treatment of patients with colorectal cancer. Magy Onkol. 2010;54:383–94.
32. Trevino V, Falciani F, Barrera-Saldana HA. DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med. 2007;13:527–41.
33. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9.
34. Ashley E. Towards precision medicine. Nat Rev Genet. 2016;17:507–22.
35. Collins D, Sundar R, Lim J, et al. Towards precision medicine in the clinic: from biomarker discovery to novel therapeutics. Trends Pharmacol Sci. 2017;38:25–40.
36. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010;11:31–46.
37. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol. 2009;27:455–7.
38. Fertig EJ, Slebos R, Chung CH. Application of genomic and proteomic technologies in biomarker discovery. Am Soc Clin Oncol Educ Book. 2012;32:377–82.
39. Schrijver I, Aziz N, Farkas DH, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. J Mol Diagn. 2012;14:525–40.
40. Sboner A, Mu X, Greenbaum D, et al. The real cost of sequencing: higher than you think! Genome Biol. 2011;12:125.
41. Frampton GM, Fichtenholtz A, Otto GA. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31:1023–31.
42. Gray P, Dunlop C, Elliott A. Not all next generation sequencing diagnostics are created equal: understanding the nuances of solid tumor assay design for somatic mutation detection. Cancers (Basel). 2015;7:1313–32.
43. Khotskaya YB, Mills GB, Mills Shaw KR. Next-generation sequencing and result interpretation in clinical oncology: challenges of personalized Cancer therapy. Annu Rev Med. 2017;68:113–25.
44. Shen T, Pajaro-Van de Stadt SH, Yeat NC, et al. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Front Genet. 2015;6:215.
45. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004;6:1–6.
46. College of American Pathologists In Vitro Diagnostic Multivariate Index Assays (IVDMIA). http://www.cap.org/apps/cap.portal?_nfpb=true&cntvwrPtlt_actionOverride=%2Fportlets%2FcontentViewer%2Fshow&cntvwrPtlt%7BactionForm.contentReference%7D=committees%2Ftechnology%2Fivdmia.html&_pageLabel=cntvwr. Accessed 01 Jul 2017.
47. Oncotype DX official website. http://www.oncotypedx.com/en-US/Colon/HealthcareProfessionals/RecurrenceRisk/ScoreResult. Accessed 01 Jul 2017.
48. Venook AP, Niedzwiecki D, Lopatin M, et al. Validation of a 12-gene colon cancer recurrence score (RS) in patients (pts) with stage II colon cancer (CC) from CALGB 9581. J Clin Oncol. 2011;29(15):3518.
49. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP working group: can tumor gene expression profiling improve outcomes in patients with breast cancer? Genet Med. 2009;11:66–73.
50. Oncotype IQ Official website. http://www.oncotypeiq.com/en-US. Accessed 01 Jul 2017.
51. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast Cancer. N Engl J Med. 2015;373:2005–14.
52. Dong L, Wang W, Li A, et al. Clinical next generation sequencing for precision medicine in cancer. Curr Genomics. 2015;16:253–63.

53. Massie CE, Mills IG. Mapping protein-DNA interactions using ChIP-sequencing. Methods Mol Biol. 2012;809:157–73.
54. Bayani J, Squire JA. Application and interpretation of FISH in biomarker studies. Cancer Lett. 2007;249:97–109.
55. Penault-Llorca F, Bilous M, Dowsett M, et al. Emerging technologies for assessing Her2 amplification. Am J Clin Pathol. 2009;132:539–48.
56. Mass RD, Press MF, Anderson S, et al. Evaluation of clinical outcomes according to Her2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. Clin Breast Cancer. 2005;6:240–6.
57. Davis LM, Harris C, Tang L, et al. Amplification patterns of three genomic regions predict distant recurrence in breast carcinoma. J Mol Diagn. 2007;9:327–36.
58. Yoshimoto M, Cunha IW, Coudry RA, et al. FISH analysis of 107 prostate cancers shows that PTEN genomic deletion is associated with poor clinical outcome. Br J Cancer. 2007;97:678–85.
59. Lotan TL, Wei W, Ludkovski O, et al. Analytic validation of a clinical-grade PTEN immunohistochemistry assay in prostate cancer by comparison with PTEN FISH. Mod Pathol. 2016;29:904–14.
60. Troyer DA, Jamaspishvili T, Wei W, et al. A multicenter study shows PTEN deletion is strongly associated with seminal vesicle involvement and extracapsular extension in localized prostate cancer. Prostate. 2015;75:1206–15.
61. Sholl LM, Weremowicz S, Gray SW, et al. Combined use of ALK immunohistochemistry and FISH for optimal detection of ALK-rearranged lung adenocarcinomas. J Thorac Oncol. 2013;8:322–8.
62. Lindeman NI, Cagle PT, Beasley MB, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol. 2013;8:823–59.
63. Abbot Molecular official website. https://www.molecular.abbott/int/en/products/oncology/vysis-alk-break-apart-fish-probe-kit Accessed 01 Jul 2017.
64. Murphy J, Bustin SA. Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard? Expert Rev Mol Diagn. 2009;9:187–97.
65. Bartley PA, Ross DM, Latham S. Sensitive detection and quantification of minimal residual disease in chronic myeloid leukaemia using nested quantitative PCR for BCR-ABL DNA. Int J Lab Hematol. 2010;32:222–8.
66. Ross DM, Branford S, Seymour JF, et al. Patients with chronic myeloid leukemia who maintain a complete molecular response after stopping imatinib treatment have evidence of persistent leukemia by DNA PCR. Leukemia. 2010;24:1719–24.
67. Goh HG, Lin M, Fukushima T, et al. Sensitive quantitation of minimal residual disease in chronic myeloid leukemia using nanofluidic digital polymerase chain reaction assay. Leuk Lymphoma. 2011;52:896–904.
68. Nikiforov YE. Molecular diagnostics of thyroid tumors. Arch Pathol Lab Med. 2011;135:569–77.
69. Cantara S, Cappezzone M, Marchisotta S, et al. Impact of proto-oncogene mutation defect in cytological specimens from thyroid nodules improves the diagnostic accuracy of cytology. J Clin Endocrinol Metab. 2010;95:1365–9.
70. Fu Q, Zhu J, Van Eyk JE. Comparison of multiplex immunoassay platforms. Clin Chem. 2010;56:314–8.
71. Choi HJ, Kim HR, Shin MG. The author response: diagnostic standardization of leukemia fusion gene detection system using multiplex reverse transcriptase-polymerase chain reaction in Korea. J Korean Med Sci. 2011;26:1401.
72. King RL, Naghashpour M, Watt CD, et al. A comparative analysis of molecular genetic and conventional cytogenetic detection of diagnostically important translocations in more than 400 cases of acute leukemia, highlighting the frequency of false-negative conventional cytogenetics. Am J Clin Pathol. 2011;135:921–8.
73. Dancey JE, Bedard PL, Onetto N, et al. The genetic basis for cancer treatment decisions. Cell. 2012;148:409–20.
74. Holbein ME. Understanding FDA regulatory requirements for investigational new drug applications for sponsor-investigators. J Investig Med. 2009;57:688–94.
75. Sargent DJ, Conley BA, Allegra C, et al. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol. 2005;23:2020–7.
76. Simon R. Genomic Alteration-Driven Clinical Trial Designs in Oncology. Ann Intern Med. 2016;165:270–8.
77. Conley BA, Doroshow JH. Molecular analysis for therapy choice: NCI MATCH. Semin Oncol. 2014;41:297–9.
78. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung Cancer. Cancer Discov. 2011;1:44–53.
79. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. N Engl J Med. 2011;364:340–50.
80. Srivastava SS, Kramer BS. Early detection cancer research network. Lab Investig. 2000;80:1147–8.
81. Zerhouni EA, Sanders CA, von Eschenbach AC. The biomarkers consortium: public and private sectors working in partnership to improve the public health. Oncologist. 2007;12:250–2.
82. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem. 2003;49:1–6.
83. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2006;97:1180–4.
84. Hayes D, Ethier S, Lippman M. New guidelines for reporting tumor marker studies in breast cancer research and treatment: REMARK. Breast Cancer Res Treat. 2006;100:237–8.
85. Keedy VL, Temin S, Somerfield MR, et al. American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) Mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. J Clin Oncol. 2011;29:2121–7.
86. Hayes DF, Bast RC, Desch CE. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst. 1996;88:1456–66.
87. Locker GY, Hamilton S, Harris J, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol. 2006;24:5313–27.
88. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007;25:5287–310.
89. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. J Natl Cancer Inst. 2010;102:152–60.
90. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol. 2002;156:311–8.
91. Teutsch SM, Bradley LA, Palomaki GE, et al. The evaluation of genomic applications in practice and prevention (EGAPP) initiative: methods of the EGAPP Working Group. Genet Med. 2009;11:3–14.
92. Chou R, Croswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2011;155:762–71.
93. Prensner JR, Rubin MA, Wei JT, et al. Chinnaiyan. Beyond PSA: the next generation of prostate cancer biomarkers. Sci Transl Med. 2012;4:127rv3.
94. U.S. Preventive Services Task Force. Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. DIANE Publishing; 1989, pp. 21–25.

Bioinformatics Tools in Clinical Genomics

14

David K. Crockett, Karl V. Voelkerding, Alan F. Brown, and Rachel L. Stewart

Abbreviations

BAM	Binary Alignment Mapping file format
BWA	Burrows-Wheeler aligner
BWT	Burrows-Wheeler transform
GATK	Genome Analysis Tool Kit
HGMD	Human Gene Mutation Database
IGV	Integrative Genomics Viewer
NGS	next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
TVC	Torrent Variant Caller
VCF	Variant Call File format
VUS	variant of uncertain significance
WGS	whole-genome sequencing

Every laboratory adopting NGS must undergo several learning curves, first regarding the implementation of new chemistries and instrumentation that continues to evolve, the second being the establishment of a bioinformatics pipeline to digest large volumes of analytic data, and, lastly, the acquisition of experiential knowledge to properly interpret and prioritize genomic results. Implementation and validation of a robust bioinformatics pipeline have proven to be a significant bottleneck for most laboratories, and the field has seen an outgrowth of specialized annotation companies that can aid laboratories in data analysis. In this chapter, basic concepts and principles of bioinformatics required for the analysis of NGS data are presented. We discuss the spectrum of NGS data generation, processing and alignment, and variant calling and interpretation. The Illumina and Ion Torrent sequencing technologies are emphasized due to their current market dominance in the NGS landscape. We include bioinformatics considerations and approaches for clinical diagnostic applications. A subsection is devoted to computational approaches for the identification of candidate genes from exome and whole-genome sequencing studies.

D. K. Crockett · K. V. Voelkerding (✉) · A. F. Brown
ARUP Laboratories, Department of Pathology,
University of Utah School of Medicine,
Salt Lake City, UT, USA
e-mail: voelkek@aruplab.com

R. L. Stewart
ARUP Laboratories, Department of Pathology,
University of Utah School of Medicine,
Salt Lake City, UT, USA

Department of Pathology and Laboratory Medicine,
University of Kentucky College of Medicine,
Lexington, KY, USA

Introduction

Whether one works at the bench or at the bedside, we are constantly interfacing with an avalanche of large data sets. Recent advances in high-throughput analysis platforms such as next-generation sequencing (NGS) beg the question of how these data will be managed and utilized by the field of clinical pathology. Historically, large research projects such as the Human Genome Project did not realize their full utility without years of subsequent bioinformatics analysis and data interpretation. Because genomic information is increasingly being used in the practice of medicine, bioinformatics is becoming an essential component in medical research and the clinical diagnostic laboratory. As the cost to sequence DNA drops and implementation widens, our clinical knowledge of significant variants will expand, and this trend is likely to continue [1]. It is important to emphasize that without expert computational analysis, the sequencing results themselves are, in essence, just a very large file of seemingly random As, Ts, Cs, and Gs. Thus, there is a general need to better understand the field of bioinformatics and how it is impacting clinical pathology. Ultimately, increased automation and deep hierarchical learning will be necessary to improve efficiency and better understand subtle interactions between genomic components [2–4].

It is interesting to note that the actual term *bioinformatics* appeared well before the current "genomics revolution."

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_14

In 1978, a Dutch theoretical biologist (Paulien Hogeweg) first coined this term in reference to the study of information processes in biotic systems [5]. One common definition of bioinformatics is "Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data" [6]. This includes any computational methods to acquire, store, organize, archive, analyze, or visualize such data. To others, bioinformatics is simply a grammatical contraction of "biological informatics" and may call to mind the computer science disciplines of information science or information technology, underscoring the large amounts of data to be analyzed and managed [7]. It is also quite notable that the majority of this computational work is not in the Microsoft Windows® environment. A more typical setting is command line parsing, scripting, and making queries on Unix/Linux hardware using programming languages and tools such as Perl, Python, Java, R, and SQL, among others.

The unprecedented volumes of both qualitative and quantitative NGS data have driven a renaissance in bioinformatics research and development resulting in the proliferation of a diversity of open-source and commercial software algorithms to support the computational processing, analysis, and interpretation of NGS results [8]. These efforts have facilitated a broad dissemination of NGS into every facet of biomedical research and a growing list of clinical diagnostic applications from targeted multigene panels to whole-genome sequencing (WGS).

Every laboratory adopting NGS undergoes a learning curve with respect to analyzing NGS data (see Note 14.1). This has proven to be a significant bottleneck due to a rapid expansion of the field, the specialized nature of bioinformatics knowledge, and a lack of trained personnel [9]. The goal of this chapter is to introduce basic concepts and principles of bioinformatics required for the analysis of NGS data. The spectrum of NGS data generation, processing and alignment, and variant calling and interpretation is discussed. The Illumina and Ion Torrent sequencing technologies and associated data analyses are emphasized due to their current dominant roles in the NGS landscape. A subsection is devoted to computational approaches for the identification of candidate genes from exome and WGS studies [10]. The chapter concludes with a discussion of in silico predictors and test reporting strategies.

Note 14.1
Pathologists can strive to become more aware of this rapidly moving field by familiarizing themselves with bioinformatics resources such as websites like *SeqAnswers* [11], *BioStar* [12], and *Bio-IT World* [13], software and server tutorials such as *Galaxy* [14] or *GATK* [6], continuing education or certificates such as *NGS short courses* [15] and *AMIA 10x10* [16] offerings.

Methods

Next-Generation Sequencing Bioinformatics

The bioinformatics processing of NGS data can be operationally divided into three major steps:

1. Generation of a sequence read file containing linear nucleotide sequence (e.g., ACTGGCA) accomplished using instrument-specific software (FASTQ file).
2. Mapping and aligning sequence reads to a reference sequence and identifying differences (variants) between sequence reads and reference (SAM/BAM file).
3. Annotation of variants (variant call format (VCF) file to report generation).

Steps 2 and 3 use either open-source or commercial software algorithms while referencing publicly accessible, online databases. Each of these steps is later presented in greater detail. We first discuss bioinformatics workflow considerations pertinent to the analysis of Illumina and Ion Torrent sequence data.

Illumina: Primary Bioinformatics Overview

NGS raw sequence base data derived from Illumina sequencers are comprised of four-color (ATCG) fluorescent images optically recorded after each successive sequencing cycle. The images captured reflect single-nucleotide base incorporation into individual sequencing clusters, with each cluster representing a single clonal amplicon as seen in Fig. 14.1. Illumina utilizes a quality control measure termed "chastity" quality filtering for acceptance or rejection of an individual cluster that is applied after the first 25 cycles of a sequencing run. Specifically, during the first 25 cycles, the highest fluorescent intensity base incorporated into a cluster is recorded, and its intensity is compared to the next highest fluorescent base recorded for the cluster. This information is used to calculate the chastity filter ratio which is derived by taking the fluorescence of the highest fluorescent intensity base and dividing it by the fluorescence of the same highest fluorescent intensity base plus the fluorescence of the next highest fluorescent intensity base. A ratio of 0.6 or greater is considered a "passing" ratio. A cluster "failure" is defined when two or more base incorporation events have chastity ratio values less than 0.6 in the first 25 cycles of a sequencing run. This quality filter removes overlapping and low intensity clusters which often represent high cluster densities causing cluster overlap and mixed sequencing signals. The dominant error mode for Illumina sequencing is the category of single-nucleotide substitutions.

An additional factor contributing to the overall error rate is the phenomenon termed "phasing." Each Illumina cluster is

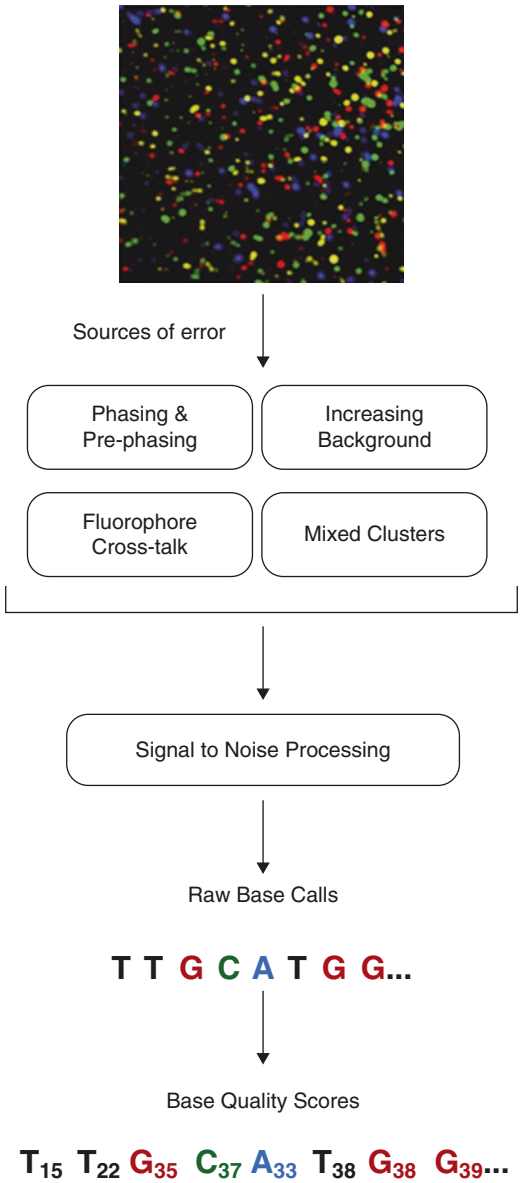


Fig. 14.1 Processing of signal-to-noise and generation of base calls for Illumina sequence data. A flow cell image generated by the Illumina Genome Analyzer is shown at the top. This represents an overlay of the four unique reversible dye terminators (fluorophores) in red, yellow, blue, and green. Respective fluorophores are incorporated into individual clonal clusters during a sequencing cycle. Process steps for generating raw base calls and sources of error are indicated. Base calls and associated quality scores from actual sequence are shown at the bottom. A adenine, C cytosine, G guanine, T thymine [10]. (*Reprinted from* Coonrod et al. [10] *with permission from Archives of Pathology & Laboratory Medicine. Copyright 2013 College of American Pathologists*)

comprised of ~1000 clonal amplicon strands, and each strand is sequenced in parallel, providing a sum total signal for the individual cluster. As each base is incorporated individually within a growing DNA strand, base incorporation can become out of phase within a clonal cluster if one base is skipped (phasing) or if multiple bases are incorporated in a single cycle (pre-phasing), resulting in non-uniform fluorescence within a clonal cluster [17]. Additionally, background fluorescence on the flow cell increases during the analytical run, which results in a decreased signal-to-noise ratio (see Note 14.2).

Note 14.2 As noted above, errors can occur by virtue of mixed signals in a cluster composed of more than one unique, overlapping clonal amplicon, especially if they have not been eliminated by the chastity filter. In addition, overlap in the emission spectra of each of the four fluorophores can make it difficult to determine which base was incorporated (fluorophores cross talk) when clonal clusters are physically close to each other. As sequence reads become longer, particularly over 100 base pairs in length, they become prone to "sticky" fluorophores which attach at a gradually increasing error rate.

Initially, real-time image analysis and base calling are performed through the Illumina instrument's Sequencing Control Software Real-Time Analysis (SCS/RTA). The Firecrest module locates fluorescent clusters in the captured image files (.TIF), assigns signal intensities, estimates "noise," and provides X and Y coordinates for each cluster. During this process, a maximum signal intensity is established for the run, as well as a threshold intensity to suppress background fluorescence. The output from image analysis provides the input for base calling. The Bustard application transforms Firecrest signal intensities into base calls following normalization for signal cross talk and cluster phasing and pre-phasing. Then, Bustard assigns the base with the highest signal-to-noise intensity ratio as the "true" base call with an attached base quality (Q) score [$Q = -10 \times \log10(e)$]. The Q score is logarithmically related to error probability (e) and is conceptually analogous to the Phred quality score used in Sanger sequencing [18, 19]. For example, bases with Q20 and Q30 quality scores have a 1:100 and 1:1000 probability of being called incorrectly, respectively. The Q score is calculated for each base along the sequence read and is used as a standard quality control metric for downstream data analysis.

For Illumina sequencing, a key metric is the number of bases with quality scores equal to or greater than 30. As an example, in a representative whole-genome sequencing data set comprised of 100 base length pair-end reads, approximately 88% of bases had Q scores of 30 or greater. After assignment of the Q scores, the Illumina sequence files are converted into a text-based file format termed FASTQ. The FASTQ file contains sequence reads that have passed filter metrics, and each read is associated with an identifier that indicates its location on the flow cell (e.g., lane and tile). The linear sequence is displayed, and each base is assigned a base quality score designated using ASCII coding. The FASTQ file format, also employed by Ion Torrent and other NGS technologies, has become the de facto information exchange currency in NGS.

Illumina: Secondary Data Analysis (Mapping, Alignment, and Variant Calling)

At this point, the primary data analysis (image analysis, base calling) is complete, and the sequencing file enters a secondary Linux-based application server. For MiSeq instruments, this process takes place using the onboard PC; however, NextSeq and HiSeq models can utilize Illumina's cloud-based BaseSpace Application Hub. Overall, secondary analysis is customizable but can include generation of the FASTQ file, de-multiplexing of indexed samples, alignment to the reference genome, and initiation of variant analysis and counting. While Illumina has made recent revisions to improve compatibility with existing open-source software, many laboratories choose to load the FASTQ file into open-source academic software programs beginning with alignment due to improved efficiency or sensitivity (BWA-MEM, NovoAlign; see below).

Sequence reads in FASTQ files are used for two main computational purposes: assembly and alignment. While the majority of diagnostic applications employ alignment to a reference sequence, assemblies are performed when no reference genome exists for the sequenced DNA (e.g., uncharacterized or novel bacteria and viruses). Algorithms used for assembly seek and join overlapping sections of sequence reads to generate longer length "contigs." The length of contigs can be increased by using longer and paired-end sequence reads to yield an overlapping genomic scaffold onto which subsequent alignments can be performed.

Mapping and aligning are the processes of determining the best match between the sequencing reads and the reference sequence. Reference genome files (FASTA) for many species, including several versions of the human genome, are available for download through several annotated databases including UCSC Genome Bioinformatics, Ensembl, and NCBI [20–22]. Due to the large number of sequence reads in NGS data sets, NGS alignment algorithms seek to decrease computational time. The two major approaches are a sophisticated data compression method termed the Burrows-Wheeler transform (BWT) and a method based on a hash table. The open-source Burrows-Wheeler aligner (BWA) algorithm [23] that has become one of the standards for sequence alignment uses the BWT method [24, 25]. An example of a popular commercial aligner that utilizes a hash table method is NovoAlign [26]. Prior to initial alignment, the reference sequence and sequence reads are converted into a population of shorter length sequences with each sequence given an identifier for computational tracking. The use of shorter length sequences (also known as "seeds") allows mapping and alignment to proceed more rapidly [27]. When using either BWT or hash table-based aligners, parameters for initial mapping and alignment need to be established, including the number of nucleotide mismatches permitted across a given read or seed length and whether gaps in alignment are allowed to accommodate insertions and deletions (indels). The operator may elect default settings or set more or less restrictive parameters. Often indels are significantly underestimated or potentially missed due to penalties from short-read alignment algorithms; longer amplicon read lengths, pair-end sequencing, and adjustment of aligner settings can improve indel call accuracy.

The output of most alignment algorithms is a file format termed SAM (sequence alignment map) which contains read position information and orientation in relationship to the reference sequence and a confidence value for the alignment. A reduced size, binary version of SAM is the BAM format. BAM files are typically compressed and allow for improved processing efficiency. Using initial alignment criteria, which are typically more permissive than secondary algorithms, the output is a data set known to contain inaccuracies.

After initial alignment, SAM/BAM files are used as inputs into secondary algorithms to refine and increase alignment accuracy prior to identifying variants (differences between the sequencing reads and the reference sequence). Two popular open-source software programs for refining alignments and calling variants are the Genome Analysis Tool Kit (GATK) [28] and SAMtools [29]. As displayed in Fig. 14.2, three major refinement steps are used in GATK: (1) removal of reads with the same start and end points, referred to as

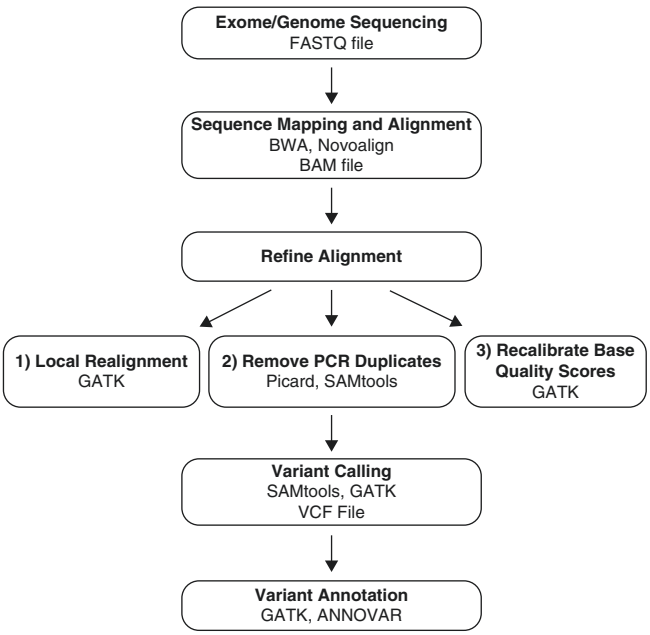


Fig. 14.2 A representative bioinformatics workflow for analysis of Illumina sequence data. Steps required to generate a final annotated variant list from raw sequencing data are also indicated. Where applicable, open-source programs are listed along with the file type generated. BWA Burrows-Wheeler aligner, GATK Genome Analysis Tool Kit, PCR polymerase chain reaction, VCF Variant Call Format. (*Reprinted from* Coonrod et al. [10] *with permission from Archives of Pathology & Laboratory Medicine. Copyright 2013 College of American Pathologists*)

PCR duplicates; (2) local realignment to improve accuracy of indel calls; and (3) recalibration of base quality scores. PCR duplicates originate by sequencing identical fragments generated by PCR during library preparation. Nucleotide errors introduced during PCR can be propagated and appear in variant files. To mitigate this, PCR duplicate removal is performed leaving only a single read with overall highest base qualities. The impact of local realignment on indel accuracy in detecting a 3 bp deletion in the FOXP3 gene is illustrated in Fig. 14.3 with aligned reads visualized in the open-source Integrative Genomics Viewer (IGV) [30, 31]. Read length and base quality affect the size of indels that can be accurately identified. Recalibrating base quality scores is done to adjust Phred-like quality scores generated by the Illumina platform, which have been shown to deviate from the true error rate. After the initial and refined alignments, variants are tabulated in a Variant Call File format (VCF) that contains several parameters including the chromosomal position of the variant, the reference base, and the alternative base(s) (e.g., single-nucleotide variant or SNV, indel). Variants can be visualized using the Integrative Genomics Viewer (IGV; Broad Institute) by importing BAM files.

Illumina: Coverage and Variant Calling

For many applications, NGS libraries are comprised of randomly overlapping fragments, exceptions being certain

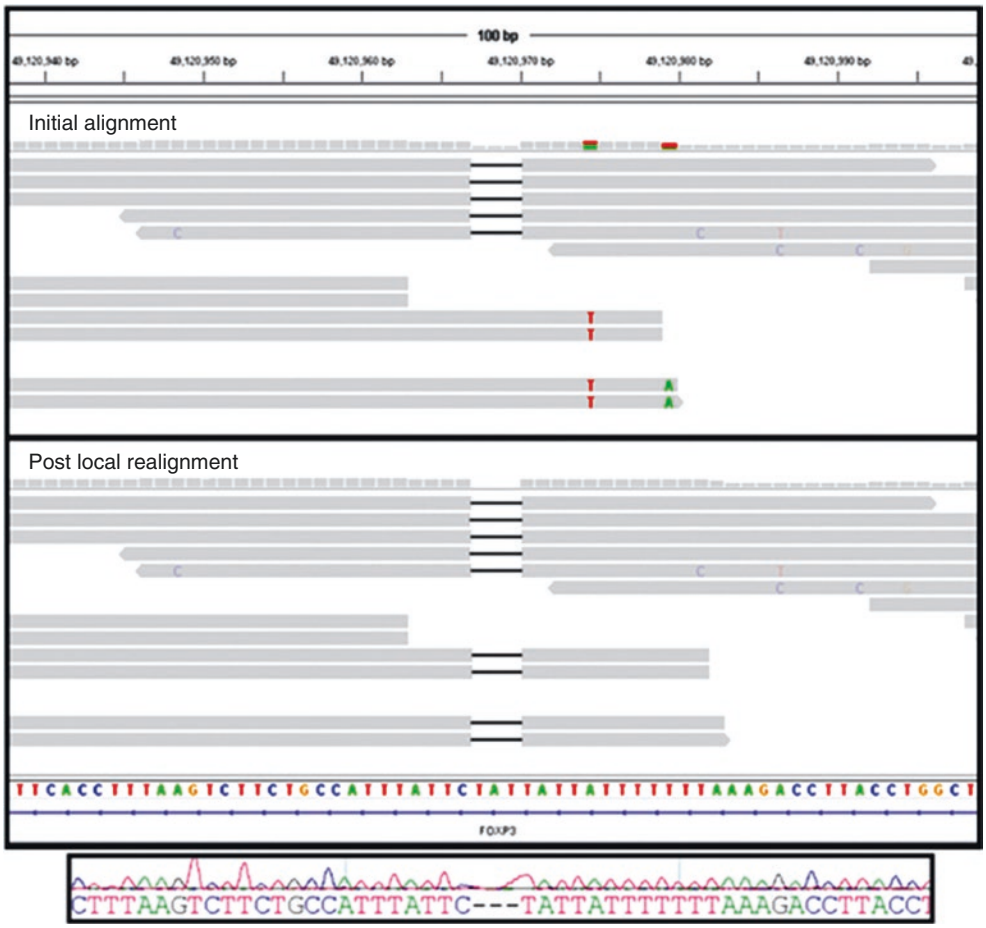


Fig. 14.3 Generating refined alignments may improve local mapping around insertions and/or deletions. The local realignment of an indel visualized in the Integrative Genomics Viewer enhances results from initial mapping with an apparent 3 bp deletion in the FOXP3 gene on the X chromosome of a male. Initially, five reads contain the deletion (depicted by black bars within the read), and four reads do not contain the deletion. Importantly, the initial four read alignments that do not contain the deletion are flanked by perceived single-nucleotide variants (SNVs) (a T variant in red and an A variant in green). The initial alignment suggests heterozygosity for the deletion on the X chromosome, but this is unlikely, given that the sequence reads are derived from a male. The lower panel (post-local realignment) shows all reads in agreement to contain the 3 bp deletion. In addition, the flanking, false-positive SNVs are no longer present. The Sanger sequencing trace shown below confirms the deletion and zygosity of the g.49120967_49120971delTAT deletion. Importantly, larger indels (>15 bp) may be underestimated or missed by current variant callers. Confirmation or further assessment by Sanger sequencing or fragment analysis (somatic *FLT3*-ITDs) is recommended. (*Reprinted from* Coonrod et al. [10] *with permission from Archives of Pathology & Laboratory Medicine. Copyright 2013 College of American Pathologists*)

targeted enrichment approaches that employ PCR. As such, multiple reads align to the reference in a staggered or non-staggered fashion, respectively. This multiplicity can be quantified by enumerating the number of reads aligned to the reference sequence at a given location and is termed the "read coverage depth." Bidirectional sequencing yields forward and reverse strand sequences, and under ideal conditions, their percentages would be approximately equal. The percentage of reads containing a variant is referred to as the "allelic read percentage" or the "variant allele fraction" (VAF). Figure 14.4 depicts a heterozygous single-nucleotide variant (SNV) and the overall concepts of coverage and allelic read percentage. Germline variant calling accuracy is greater when there is a consensus among aligned reads consistent with either heterozygosity or homozygosity. Ideally, a sample with a heterozygous SNV would have approximately a 50/50 ratio of reads containing variant and reference nucleotides, and forward and reverse strand reads would be equally represented in both variant- and reference-containing reads. A homozygous variant would ideally be present in 100% of aligned sequences with equal representation of forward and reverse strands. In practice, a wider range of allelic read percentages and forward and reverse strand percentages (manifesting as "read strand bias") is typically observed.

Read strand bias can arise from several sources including differential PCR amplification of library fragments, sequencing errors in difficult to sequence regions, and misalignment

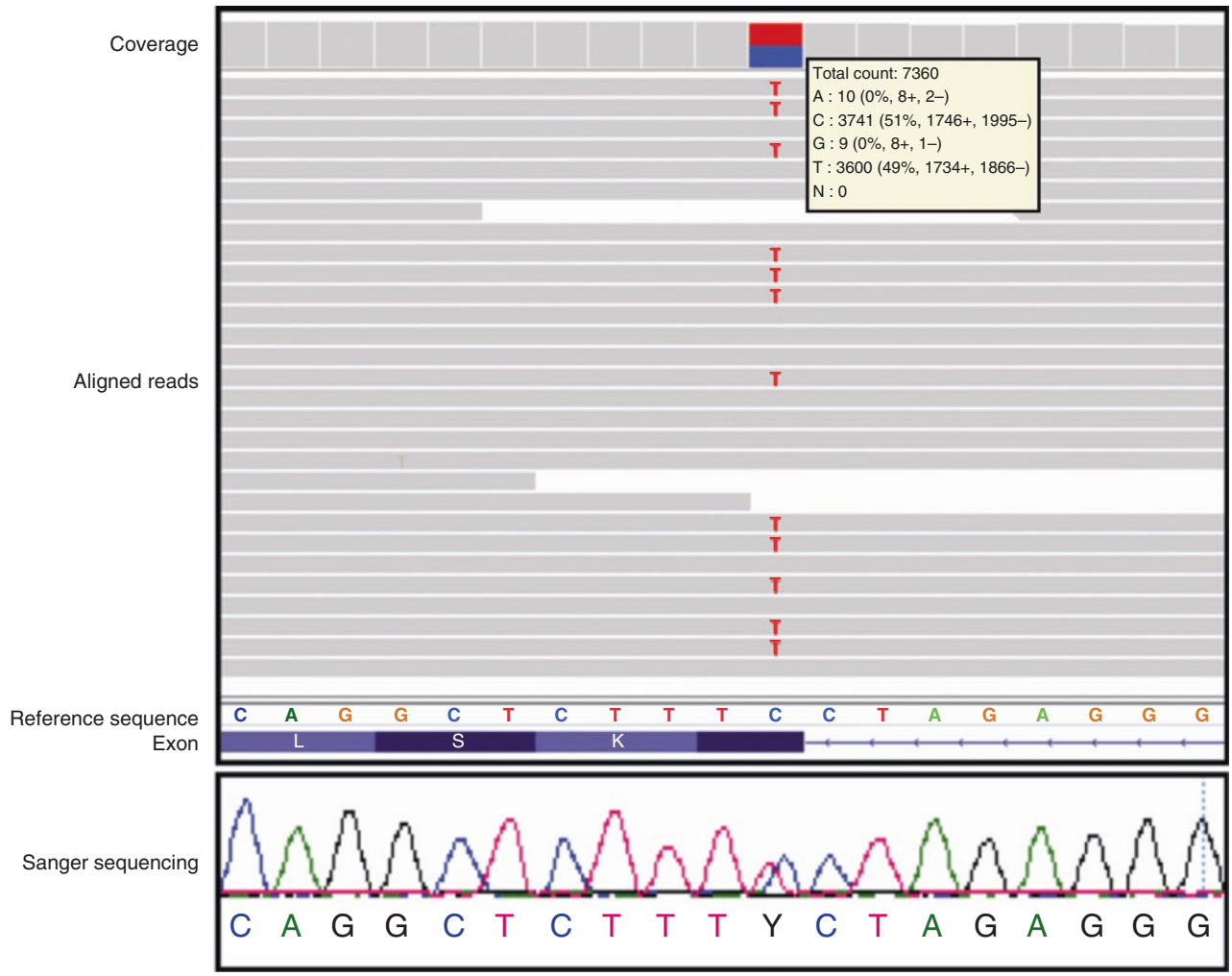


Fig. 14.4 Illumina sequence reads and alignment as seen in a popular genome browser (Integrative Genomics Viewer, IGV). Gray boxes across the top row represent read depth (coverage). Mapped and aligned reads are shaded gray, with variance from the reference highlighted by a unique color. In this example, a cytosine to thymine (C > T) variant change is present, with variant Ts highlighted in red in the aligned reads. The heterozygous change is also indicated above (in Coverage), and, in this case, because the variant is heterozygous, the box is split into two colors, red for the variant nucleotide (T) and blue for the reference nucleotide (C). The reference nucleotide sequence (reference sequence) is shown below the aligned reads along with the location of the exon and amino acid translation of the sequence (exon). This region was also Sanger sequenced to confirm the heterozygous (C > T) variant. The Sanger trace from this sequence is shown at bottom. (*Reprinted from* Coonrod et al. [10] *with permission from Archives of Pathology & Laboratory Medicine. Copyright 2013 College of American Pathologists*)

of related sequences (e.g., pseudogenes or highly homologous genes). The minimum number of reads required for variant calling needs to be empirically determined for the application. For germline variant detection, a coverage of approximately 30-fold has been found to offer a balance between sensitivity, specificity, and sequencing costs in the research setting, whereas clinical laboratories typically aim for higher coverage depths (>50-fold) in an effort to increase variant detection and improve variant call confidence [32–34]. For somatic variant detection, a coverage greater than 500-fold is sought by most clinical laboratories performing cancer-related targeted gene panels [35] with many sites using robust extraction and retrieval methods that can achieve mean coverage of >3000-fold. Coverage depths across sequenced regions are variable due to factors such as differential ligation of adapters to fragments during library preparation and differential amplification during clonal expansion, thus necessitating that sufficient sequencing is performed to meet clinically required coverage.

Finally, whereas the above discussed 50/50 and 100% variant ratios are relevant to identifying heterozygous and homozygous germline variants, respectively, they are less resolute when applied to heterogeneous scenarios such as somatic variants in cancer samples which may be comprised of a mixture of tumoral subclone and normal germline cell populations. In a histologically "pure" tumor sample, a VAF near 50% would suggest a heterozygous variant in tumor cells that likely occurred early in clonal evolution, whereas a VAF greater than 60% could suggest regional loss of heterozygosity or homozygosity in the tumor. To improve sensitivity in identifying low-level somatic variants (<5% of reads), increased read coverage into the >500-fold range is routinely being used for clinical testing [36, 37].

Illumina: Variant Annotation

Annotation is accomplished by introducing a VCF file into software that ascribes additional features to variants. Examples of open-source software that contain annotation functions are ANNOVAR [38, 39], GATK [28], and SnpEff [40], among others. Annotation outputs include many features including chromosomal location of base change from reference, whether the variant is in a gene and its respective location (e.g., exon, intron, splice site), the consequence of the change to a codon (e.g., synonymous versus nonsynonymous, missense versus frameshift), and zygosity (e.g., homozygous or heterozygous). Often incorporated into annotation software programs are algorithms that predict the functional impact of variants on proteins such as Sorting Intolerant from Tolerant (SIFT) [41–43], Polyphen2 [44, 45], and MutationTaster [46–48]. Notably, for clinical interpretation, the initial predictions generated by these software programs must be reviewed by experienced sequence analysts and laboratory directors. Additional information regarding in silico prediction algorithms is discussed later in this chapter.

Clinical laboratories may choose to maintain a composite system of sequencing and bioinformatics annotation in-house or to implement a distributive model where only sequencing is performed in-house, while bioinformatics are performed by an external reference laboratory or service provider. Alternatively, a laboratory may outsource only their report generation (variant classification and clinical trial matching). Typically, external laboratories or service providers offer customizable, cloud-based interfacing to allow the primary laboratory to directly access proprietary software or upload sequencing files for off-site analysis. While this option is an additional expense to the laboratory and typically includes an annual retainer fee and percentage of reimbursement rate per test, it may represent a preferable option for laboratories that do not have sufficient internal bioinformatics capabilities.

Ion Torrent: Bioinformatics Overview

To generate DNA sequence information, Ion Torrent™ technology relies on the relatively simple biochemical components of DNA polymerase and natural nucleotides. The nucleotide incorporation detection process monitors hydrogen ion release as known nucleotides are incorporated into growing DNA strands. Hydrogen ion signals from individual reaction wells are detected by a proprietary ion sensing technology that utilizes field effect transistors scaled in a massively parallel configuration using semiconductor technology. Analogous to Illumina technology, signal-to-noise ratios are algorithmically converted into nucleotide base calls with associated quality scores. The linear sequence file output is converted into the FASTQ format which can then be put into a variety of open-source and commercial software for subsequent mapping, alignment, and variant calling. In practice, most groups use software developed by Ion Torrent which has been optimized for Ion Torrent sequence read data.

Ion Torrent: Data Flow and Sequence Generation

Each ion sequencing chip (e.g., 314, 316, 318 Ion PGM Chip Kits) contains a high-density array of micro-machined wells that are placed over an ion-sensitive layer and a proprietary ion sensor. During each nucleotide flow over the chip, when a nucleotide is incorporated into a growing strand of DNA, a hydrogen ion is released. The ion release changes the pH of the solution and is detected by the chip's ion sensor. The raw

pH value from each well is converted into a voltage and captured as a digital representation of that voltage. If the nucleotide that flows over the chip is not complementary, no incorporation occurs, and thus no change in pH or voltage is recorded. In this way, analysis of these data can reveal the base incorporated during the nucleotide flow. This process transforms the chemical information to digital information in a conceptually simple and direct manner.

The Ion Sequencer outputs raw sequencing data in the form of Data Acquisition (DAT) files. These DAT files reflect the conversion of raw pH values in each well into a digital representation of the change in voltage. The raw DAT data files are then transferred to the Torrent Server for analysis pipeline processing. On the Torrent Server, the raw signal measurements are converted into incorporation measurements and then into base calls for each read. Figure 14.5 shows the steps in the Torrent pipeline from the point of view of the data files that are generated. As shown, each data file output by one step is put into the next step in the pipeline.

The sequence generation step, also called base calling, determines the actual sequence of individual nucleotide bases in each sample. The Torrent algorithm, BaseCaller, runs automatically during the Torrent Suite pipeline. It is important to note that the Torrent Suite base calling algorithm is optimized for Ion Torrent data. Although the base calling module uses fairly stringent filters that are designed to increase the accuracy of results, filters can be adjusted if a given sequencing application requires maximizing the number of reads. After sequence data are generated, the Torrent Server automatically performs sequence alignment on those data.

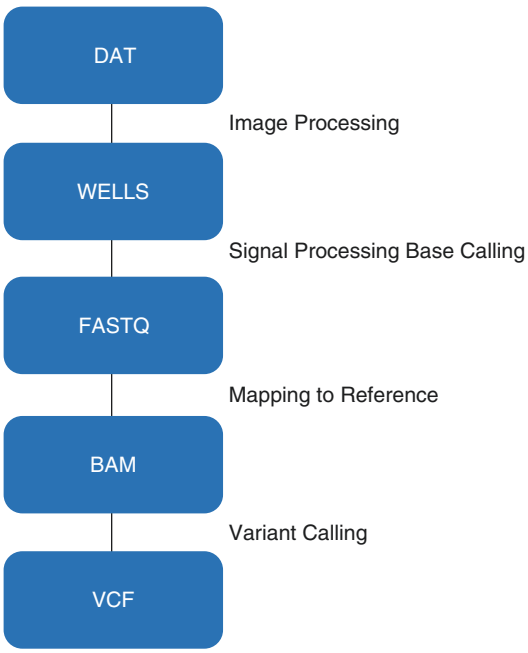


Fig. 14.5 The bioinformatics pipeline from Torrent Suite is as follows: DAT (Data Acquisition) contains raw voltage measurements from the chip. The WELLS file contains the nucleotide incorporation signals for the flow for each well. FASTQ contains the nucleotide calls (sequence calls) and associated quality values. BAM (Binary Alignment Map) contains mapped reads with their alignment to the reference genome. VCF (Variant Call Format) contains variants called on the input DNA sample. Each variant call details how a given DNA position (sequence) found in the sample differs from the reference genome, for instance, by an insertion of bases, deletion of bases, or change in a base

Ion Torrent: Sequence Alignment

During sequence alignment (also known as mapping), base calls generated by the Torrent Suite analysis process are aligned to a reference genome in the BAM file format. Several alignment metrics are also produced at this time. The Torrent Suite utilizes the Torrent Mapping Alignment Program (TMAP). This is a sequence alignment software program that is developed specifically to meet Ion Torrent data mapping challenges (see Note 14.3).

Note 14.3 Ion Torrent™ data requires special consideration during the alignment process for several reasons, including (1) reads generated by Ion Sequencers are variable in length and are expected to increase over time as chemistries improve and (2) the principal error mode associated with Ion data relates to miscalling homopolymer lengths and results in insertion or deletion errors during alignment and post-processing.

TMAP incorporates three common alignment algorithms (BWA-short, BWA-long, and Sequence Search and Alignment by Hashing Algorithm, SSAHA). The main indexing structure in TMAP uses a compressed suffix array (FM-index) based on the Burrows-Wheeler transform (block-sorting compression). The initial alignment approach uses all three algorithms to quickly produce a list of candidate mapping locations. These candidate locations are then refined using the more accurate Smith-Waterman algorithm. Resulting alignments are aggregated to identify the optimized mapping location. User-defined parameters then determine if all alignments, a subset of alignments, or a random best alignment is reported. TMAP employs a two-stage mapping strategy to maintain sensitivity and specificity while significantly reducing runtime. In this two-stage mapping, reads that do not align during the first pass are given to the second stage with a new set of algorithms and/or parameters. The output of sequence alignment is a BAM file containing mapped reads. Each BAM file can be analyzed to obtain various metrics, including quality estimates and read length estimates. Various chip loading and alignment metrics can be viewed in the Torrent Browser's Reports summary pages.

TMAP has key advantages over other alignment tools. To deal with varied length reads and error profiles that are specific to Ion Torrent™ data, the re-implemented versions of the three popular alignment algorithms have been optimized. In this way, TMAP results are expected to perform significantly better when compared against the original algorithms alone. Since TMAP amalgamates the candidate locations for all three methods and identifies the best alignment, final accuracy and specificity can benefit from the advantages of each individual algorithm. The overall performance is comparable to other alignment software in terms of CPU burden and RAM utilization. It is notable that the combined performance of multiple algorithms is equal to or better than the total performance of running each algorithm separately. However, some technical issues to avoid when using TMAP are described in Note 14.4.

Note 14.4 TMAP is recommended for alignment, as its algorithms are tuned to handle Ion Torrent data in an optimal manner. Some common issues to avoid when using TMAP are:

1. TMAP is not an assembler. TMAP is an alignment tool that gives the location where a particular read from the sequencing instrument aligns to the reference genome, but TMAP does not create a consensus file for the reference.
2. TMAP currently does not support RNA-Seq data.
3. Alignment programs must be optimized for the sequencing platform. The user can optionally use a different alignment program (outside of the Torrent Browser and the Torrent Suite Software). To align Torrent data with other alignment programs (besides TMAP), the user must ensure that the program's filters are set for Torrent data, not for data from other platforms. Incorrect results may occur if the data are not aligned with the correct parameter settings.
4. TMAP performs best when a restrictive error tolerance (such as maximum threshold of five mismatches) is not specified.
5. Typically, quality scores in repetitive sequences are lower than in nonrepetitive sequences.
6. It is not recommended to install other utilities named tmap on one's Torrent Server. For example, the European Molecular Biology Open Software Suite (EMBOSS) also includes a utility named tmap. If one installs EMBOSS on the Torrent Server, it will be likely to see name conflicts with the two different tmap programs.
7. Running TMAP using a partial or incomplete reference sequence may cause reads originating from homologous regions to be incorrectly mapped to target regions, which in turn may cause a downstream variant calling application to produce false-positive variant calls.

In terms of realignment, the Torrent Browser supports redoing sequence alignment through two different interfaces. The first is by way of a plugin, where the alignment plugin runs the TMAP alignment module and optionally supports aligning against a different reference genome. The second is by reanalysis, where the Run Report reanalyze feature supports rerunning TMAP and also supports changing TMAP parameters for the new alignment.

Ion Torrent: Variant Calling

To perform variant detection, sequencing reads are first mapped to a reference genome to generate a read pileup. This read pileup is compared to the reference sequence, and SNPs and indel variants are identified. Ion's variant calling algorithms make calls based on the consensus accuracy, independently of the variants identified in raw reads.

Several software strategies exist for calling variants using Ion Torrent sequence data. These approaches were built to support different sequencing applications, such as targeted sequencing using Ion AmpliSeq™ or Ion TargetSeq™ selection technologies, as well as traditional genome or exome sequencing experiments. These software options are listed here with each software approach and their respective workflows described in more detail below.

1. Torrent Variant Caller plugin. This is a SNP and indel calling analysis module that is part of the Torrent Suite Software and is accessible through the Torrent Browser. It is designed to be initiated automatically after sequencing data have been generated and bases called. This plugin can also be initiated manually to process previously generated data sets.

2. Ion Reporter™ Software. This is a cloud-based software service that provides both variant calling and annotations. It incorporates log and traceability features that are essential to performing routine sequencing assays.

3. Third-party software. A list of commercial partners that have provided software tools and customized services for detecting and visualizing variants is available on the Thermo Fisher webpage.

The Torrent Variant Caller (TVC) plugin accepts the aligned reads (.bam) generated by Torrent Suite Software as input. The plugin produces an output file (.vcf) of an annotated list of SNPs and indel variants called in each sample. By simply configuring the run plan before running the chip on the Ion PGM™ Sequencer, a user can set the TVC plugin to run automatically upon completion of the primary analysis of the chip. The TVC plugin can also be run (or rerun) manually at any time after the primary analysis has completed (see Note 14.5).

Note 14.5 The TVC plugin offers several advantages. It has been optimized for Ion Torrent data, it is an included

component of the Torrent Suite Software (automatically updated with each new release), and it is supported by Ion Torrent. To initiate an analysis with the TVC plugin, users set up three key configurations:

1. Workflow – Users can select from a set of preconfigured workflows, according to the expected variant frequency in the sample (germline, all variants greater than 20%, or somatic, at least one variant less than 20%) and the library type.
2. Scope of analysis – Users can provide a reference sequence for alignment that is larger than the regions sequenced (e.g., the whole human genome) and then restrict the variant calling analysis to a specified region by uploading a target regions file in BED format. The region file reduces the run time of the TVC plugin.
3. Scope of reporting – The user can also require the scope of the variant report to include a specified region, even if a variant has not been specifically identified in that region, by uploading a hot spot region file in BED format.

Also note that, typically, germline variant frequency refers to a relatively pure population, whereas somatic variant frequency is found in a mixed population.

When completed, TVC outputs five primary report tables. The most inclusive is the Variant Calls table. This table provides details about each variant called, including chromosome, position, and sequence coverage. For each variant called, the genomic position listed in the "Position" column is hyperlinked to open the Integrative Genomics Viewer (IGV) and displays all reads pertaining to that variant. The TVC plugin's results page also contains a File Links section, which lists the output files generated by the Torrent Variant Caller plugin. This allows any output files to be conveniently available for loading into IGV or other third-party tools for further visualization or analysis.

Release of Torrent Suite Software v4.4 (from February 2017) offered several improvements, including faster signal processing, additional customization including community-developed plugins, and improved variant calling. Another feature was improved well characterization ("bead finding") for more accurate background model signal processing and better phase parameter estimation. Additional filter options designed to increase the accuracy of the results were also incorporated. Importantly, users may wish to reanalyze and optimize data based on their sequencing application. For example, data quality may weigh most importantly for detection of rare gene variants. On the other hand, maximizing read depth for counting applications such as somatic variant detection and gene expression may be most important. By simple software interface, filtering can also be returned to the less stringent filtering of earlier versions, or additional, more stringent filters can be added to provide the most accurate output [49].

Ion Reporter™ Software is a suite of bioinformatics tools meant to streamline and simplify analysis, reporting, and archiving of sequencing data. The initial design is for researchers performing repeated analysis of sequencing assays. Ion Reporter™ Software also integrates comprehensive public annotations to reduce the bioinformatics work needed to understand the impact of detected variants. Although the annotation and interpretation scope is much wider than that of the TVC plugin, the underlying variant calling algorithms are the same as those of the TVC plugin. Some advantages of using Ion Reporter™ Software are that it automatically adds annotations to variants, gives preconfigured workflows (to compare pairs and trios of samples), and provides an audit trail and version control. It also scales to efficiently utilize the computing power of cloud resources including the Torrent Suite Storage System.

Currently, the Ion Reporter™ Software can perform mapping, variant calling, and annotation starting with the input of an unaligned BAM file. Alternatively, it can perform only variant annotation starting with a variant file in VCF format produced by the TVC plugin. To manually or automatically move sequencing data from the Torrent Server to Ion Reporter™ Software, users launch the Ion Reporter™ Uploader plugin. If preconfigured prior to starting a sequencing run, the Ion Reporter™ Uploader plugin uploads the unaligned read BAM file to Ion Reporter and automatically performs mapping, variant calling, and annotation. Ion Reporter™ Software data upload, analysis, and storage are packaged for each chip and purchased as an addition to the other consumables.

In terms of third-party software, Ion Torrent has partnered with commercial software providers including DNASTAR (SeqMan) and SoftGenetics (NextGENe), Partek, and Avadis NGS. The intent is to provide alternative solutions for end-to-end workflows focusing on variant calling. Each of these software products provides a comprehensive solution, which allows users to identify variants, annotate those variants, and perform multi-sample comparisons. In addition, these products allow users to integrate additional genomics data into a single viewer. Workflows specific to Ion Torrent must be selected within third-party software solutions for optimal results (see Note 14.6).

Note 14.6 Issues to note include avoiding the use of the TVC plugin or Ion Reporter™ Software on results from primary analysis methods (base calling and aligning) not optimized for Ion Torrent data. Running the TVC plugin or Ion Reporter™ Software over a long reference genome without defining a target region with a BED file causes run times to be significantly longer than with a BED file. One should avoid deriving conclusions about variants based only on raw accuracy or consensus accuracy. The TVC plugin uses consensus accuracy for candidate calls and also uses the raw reads to model errors and true variants. Therefore, the variant

calls made by the TVC plugin cannot be verified solely by viewing the reads in an alignment browser, because the browser does not illustrate the power of the variant caller algorithm across multiple reads at the same position.

All of the initial data analysis steps for NGS reads as described above must be in place prior to the additional filtering, annotation, and interpretation steps leading to gene discovery or clinical diagnostics.

Pipeline Validation and Quality Control Metrics

The "pipeline" by which laboratories process sequencing results is highly customizable. While general themes exist, each custom pipeline is different and expresses an individual set of strengths and weaknesses. For laboratories performing the entire genomic testing process (sample handling, sequencing, and bioinformatics), it is vital to validate each major processing step to assure that the total test performance meets design requirements [50]. General validation guidelines have been difficult to fully implement given the often proprietary restrictions of software supplied by commercial NGS companies and widespread use of custom-made software in both academic and reference settings. Examples of quality control metrics for bioinformatics pipelines include establishing thresholds for sequencing run quality in regard to percentage read alignment, average read quality and length, mean coverage, and reproducibility of variant frequency in characterized controls [51]. All laboratories performing NGS testing should independently validate their bioinformatics pipeline under the supervision of a qualified medical professional specializing in NGS testing and interpretation [52]. Validation should attempt to examine the individual components and the composite systems in a clinically appropriate manner using the sample types and variants of intended use [50]. Potential weaknesses (indel allelic frequency, GC-rich areas, pseudogenes, off-target annealing) should be specifically interrogated to minimize inaccuracies. Furthermore, an increasing number of variant callers directly reference online databases to provide an expandable list of columns displaying variant-specific observational and predictive information. While this feature enhances efficiency, the accuracy under different variant types should be verified manually. Lastly, all activities to validate and ensure quality control of the bioinformatics pipelines should be handled as a documented laboratory procedure following accreditation standards.

Data Integrity, Security, and Storage

As gene panels become larger, read coverage deeper, and exome and genome sequencing more widely available, data storage has become a management issue, with concerns regarding cost, integrity, reanalysis, and security. Many laboratories have experienced increasing value for off-site data storage with cloud-based retrieval. Security and privacy remain paramount to data management, and all storage solutions must abide by current HIPAA security standards. Due to their longevity and singularity, genomic results are especially prone to compromise by unexpected disaster, security breaches, and insufficient research privacy. The prospective costs of storing and securely maintaining genomic results should be clearly delineated in budgetary planning with interval reassessment given the continuously evolving parameters of data storage. Data security and loss prevention (integrity) should be managed by implementing security measures that prevent unauthorized access or manipulation, fully control all output devices within the LIS, and optimize a backup schedule with redundancy. Further data transmission integrity and formatting standards including encryption should be verified during the pipeline validation. In-depth resources including several recent review articles and the College of American Pathologists' (CAP, Northfield, IL, USA) Molecular Pathology Checklist (Section: Bioinformatics Pipeline for NGS) are available for additional information [51, 53–55].

Exome and Genome Sequencing for Causal and Candidate Gene Discovery

The use of exome and genome sequencing for causal and candidate gene identification requires additional bioinformatics strategies beyond variant calling. Genomes contain approximately 3–3.5 million noncoding and coding variants, whereas exomes contain 15,000–20,000 variants in coding regions. Whether starting with genome or exome data sets, the primary approach to identifying causal or candidate genes is to:

1. Focus on variants that are in coding or splice site regions (IGV).
2. Exclude higher population frequency variants (>1%, healthy homozygotes) by assuming non-pathogenicity (ExAC, gnomAD, dbSNP, observed institutional frequency).
3. Determine if remaining variants have known or predicted deleterious impact on gene function (HGMD BioBase, Alamut, ClinVar, COSMIC, OMIM, PolyPhen-2, SIFT, MutationTaster Literature Search).
4. Rank variants on the basis of deleterious impact and presence within genes of biological relevance to patient phenotype and co-segregation with that phenotype if a family study is being performed (priority or tier-based classification).

Efficiently accomplishing these steps requires the establishment of an integrated process that draws upon several curated databases [e.g., 1000 Genomes, dbSNP, Online Mendelian Inheritance in Man (OMIM), Human Gene Mutation Database (HGMD BioBase)] and variant impact prediction tools [e.g., Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping v2 (PolyPhen-2), Genomic Evolutionary Rate Profiling (GERP), etc.].

The above approach is referred to as a logic tree or heuristic filtering method, and in practice, most clinical laboratories performing causal and candidate gene identification have established in-house custom workflows to accomplish these steps in analysis. An emerging trend is the development of several commercial software programs to accomplish heuristic filtering, though one limitation of heuristic methods is that they do not provide any measure of statistical uncertainty for a given variant or candidate gene. In this context, new causal and candidate gene discovery prediction algorithms are being developed such as the Variant Annotation, Analysis, and Selection Tool (VAAST 2.0) that compares allele frequencies between cases, controls, and background data sets in conjunction with modeling variant severity by amino acid substitution analysis to provide a list of variants, each associated with a VAAST ranking score and a p-value [56, 57]. The p-value is a measure of the probability that a variant is statistically significant in a case, as compared to the control data set.

Another recently reported approach to predicting causative variants describes a statistical method using a weighted sum approach, which takes into account "background" variation in genes to avoid having large or highly variable genes in the population rank high on the candidate list, can accommodate related or unrelated data sets, can incorporate linkage or functional data, and uses a computational approach to generate a measure of statistical certainty (p-value) for individual genes [58]. One could combine both heuristic and probabilistic approaches as shown in Fig. 14.6, and the

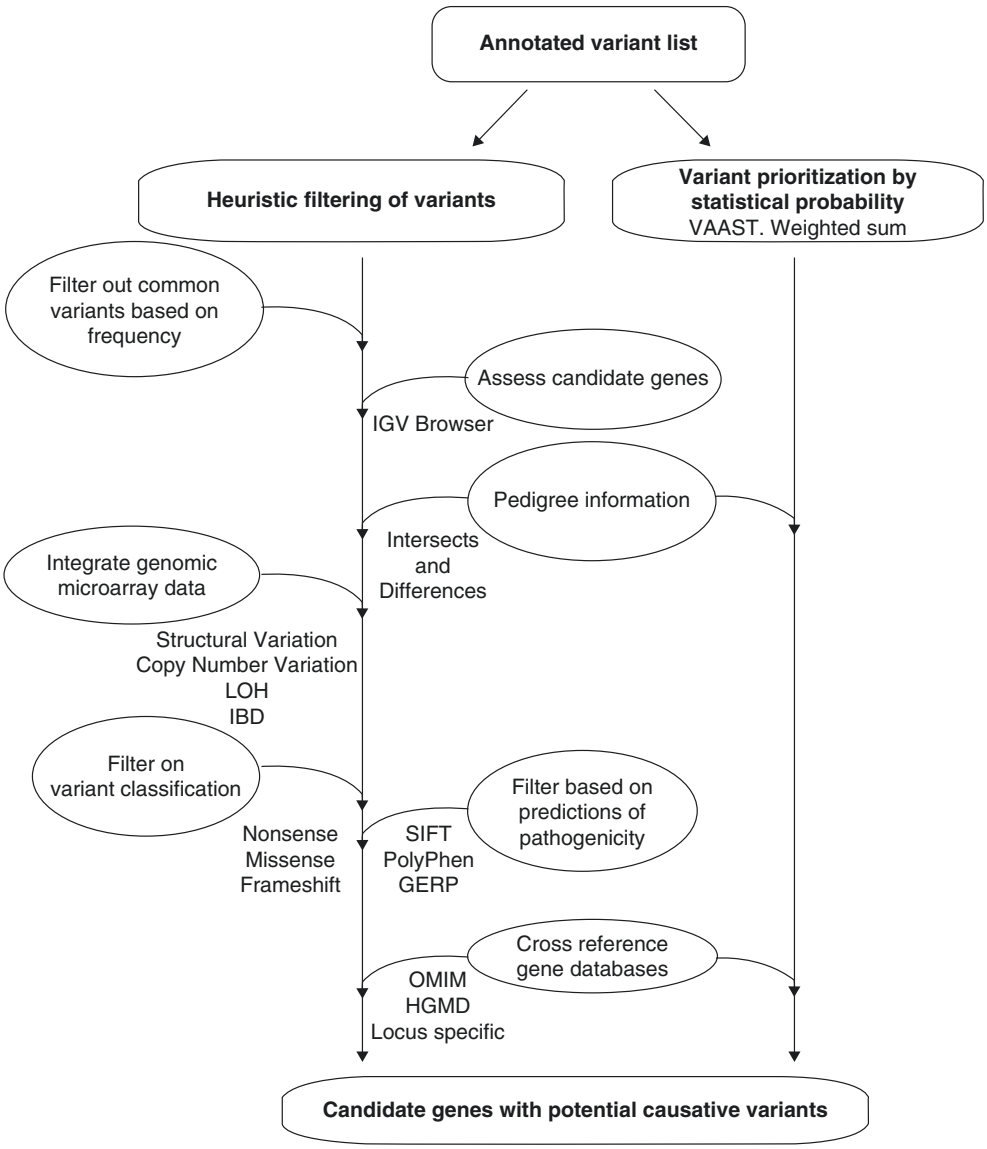


Fig. 14.6 Diagram of approaches for candidate gene discovery from exome and genome sequencing data. Annotated variant lists can be analyzed with heuristic filtering approaches, statistical probability approaches, or a combination of both, to generate candidate gene lists. Multiple process steps are involved in heuristic filtering as depicted. Steps that incorporate pedigree information and cross-referencing of gene databases are critical components of both heuristic and statistical probability approaches. GERP Genomic Evolutionary Rate Profiling, HGMD Human Gene Mutation Database, IBD identity by descent, IGV Integrative Genomics Viewer, LOH loss of heterozygosity, OMIM Online Mendelian Inheritance in Man, PolyPhen Polymorphism Phenotyping, SIFT Sorting Intolerant from Tolerant, VAAST Variant Annotation, Analysis, and Selection Tool. (*Reprinted from* Coonrod et al. [10] *with permission from Archives of Pathology & Laboratory Medicine. Copyright 2013 College of American Pathologists. Reprinted in a modified form with permission from Coonrod* et al.)

outputs of each approach are then cross-compared to generate causal and candidate gene lists. One commercial software that incorporates both heuristic and probabilistic (i.e., VAAST) approaches has been developed by Omicia [59]. In terms of final variant interpretation for clinical use, other notable commercial efforts include approaches such as Ingenuity Variant Analysis [60], SV Bio Genome to Mutation (G2M) [61], Agilent Cartagenia Bench Lab [62], PierianDx Clinical Genomicist Workspace [63], and Golden Helix VarSeq [64].

Finally, heuristic or probabilistic approaches can also be complemented by including information from high-density microarrays. For example, array data in family studies can be mined to identify linkage regions and regions of identity by descent that co-segregate with the phenotype of affected individuals in a pedigree. Focusing on these regions may reduce and prioritize specific genomic regions for evaluation. Increasingly, laboratories are utilizing exome sequencing in the diagnostic evaluation of undiagnosed germline disorders. By running comparison exomes (parents, siblings, affected relatives) in parallel with the affected individual often referred to as "trios," sequence analysts are able to drastically narrow the pool of potential causative genes and approximate potential inheritance risk (de novo, recessive, X-linked) for future offspring [65–67]. Recently, recommendation regarding the development and validation of clinical exome sequencing for germline variants has been published to aid laboratories in ensuring robust design and quality control [68].

In Silico Predictors

Medical genetics involves diagnosis, management and risk determination, and genotype-to-phenotype correlation of gene variants relating to disease [69, 70]. In monogenic diseases, gene mutations are typically curated as either pathogenic or benign. However, in some cases, variants may be classified as "unknown" or of "uncertain" significance because they have not been clearly associated with a clinical phenotype. The expense of time and labor to validate disease association of a given variant of uncertain significance (VUS) may be cost prohibitive [71, 72]. To help bridge this genotype-phenotype gap, prediction algorithms are used to narrow the uncertain "gray area" between pathogenic and benign sequence variants [73–76]. While in silico prediction algorithms can provide helpful supportive information, they should not be used as the sole determinant of variant pathogenicity. Furthermore, the clinical significance of intergenic and intronic variants may be difficult to assess.

There are several established methods for predicting mutation severity, many of which have been available online for years. Prediction tools such as PolyPhen-2 [45, 77] and SIFT [41] are primarily based on multiple alignment of different species assessing the degree of evolutionary conservation for specific bases and amino acids. More recently introduced, MutPred [75] calculates the probability of deleterious mutations by disrupted molecular mechanism. Additionally, PMut [78] is a neural net, based and trained on human mutations. However, prediction algorithms are not always in agreement with curated data or each other and are, as yet, primarily research tools [79–81]. A brief description of a representative sample of these online prediction tools may serve to improve our understanding.

SIFT was first published in 2003 by Ng and Henikoff from work done at the Fred Hutchinson Cancer Research Center in Seattle [41]. The algorithm predicts whether an amino acid substitution will affect the function of a protein based on both sequence homology to various orthologs and physical properties of amino acids. SIFT is a multistep procedure that (1) searches for and chooses similar (>90% identical) sequences, (2) makes an alignment of these sequences, and (3) calculates scores based on the amino acids appearing at each position in the alignment. It was initially developed and trained on nsSNP data sets from LacI, lysozyme, and HIV protease [82]. This algorithm works especially well when adequate numbers of sequence homologs are available for multiple alignment.

PolyPhen-2 (Polymorphism Phenotyping v2) is an EMBL-based tool described in 2002 by Ramensky et al. and updated in 2013 by Adzhubei et al. [45, 77]. It was developed to predict the possible impact of an amino acid substitution on the structure and function of a human protein using physical and comparative considerations. It was originally developed from a set of disease-causing mutations in human proteins with known structures extracted from the SWISS-PROT database and correlated to the OMIM database [83]. Because the algorithm relies on predicted structural disruption, it works especially well when protein structure is known and less reliably when a solved protein structure is not available.

MutationTaster2 was updated in 2014 and utilizes a Bayes classifier and database review to generate predictions for protein-coding, intronic, and splice site variants [48]. Alterations reported in a homozygous state in at least four individuals from either 1000 Genomes Project or HapMap are disregarded as nonpathogenic, whereas known pathogenic variants from ClinVar are automatically predicted to be disease-causing.

PMut was updated in 2017 by the Molecular Modeling Unit at the Institut de Recerca Biomédica, Parc Científic de Barcelona, Spain [78, 84]. It is based on a two-layer neural network and was trained using human mutation data. It allows for either prediction of single-point amino acid mutations or scanning of mutational hot spots and is focused on predicting Mendelian pathological mutations. Results are

obtained using a computational technique known as alanine scanning, which scores the predicted pathogenicity index (including accessibility and predicted secondary structure changes) associated with a mutation to alanine for all residues. A graphical interface for Protein Data Bank (PDB) structures, when available, and a database containing hot spot profiles for all nonredundant PDB structures are also accessible from the PMut server.

MutPred2 is a recently updated prediction algorithm that [75, 85] builds on the established SIFT method but offers improved classification accuracy based upon protein sequence and models changes of structural features and functional sites between wild-type and mutant sequences with output of probabilities of gain or loss of structure and function. It was trained on a set of disease-associated SNPs from cancer and the OMIM disease archive. This predicted disruption of molecular function again works especially well for well-studied proteins, for which homolog and solved structure is available.

The Rare Exome Variant Ensemble Learner (REVEL) was first described in 2016 and attempts to abrogate the typically limited statistical and functional information associated with rare missense variants by using multiple in silico prediction algorithms in combination to implement an ensemble method for predicting pathogenicity [86]. REVEL is specifically designed for rare variants that lack database references or functional study information. This ensemble method generates random forests drawn from 13 individual in silico prediction algorithms to predict pathogenicity in missense variants that are then scored from 0 to 1. Comparison has shown significantly better discrimination and higher area under the curve for REVEL-based predictions of rare, neutral SNPs compared to other ensemble and individual prediction tools.

Splice site predictors have become increasingly available for in silico prediction and variant classification. While documentation continues to expand, splice site mutations appear to comprise a minor but significant proportion of pathogenic variants and have the potential to provide variable functional impacts based on the number and function of constitutive and alternative splice variants in an individual gene. A number of computational methods providing predictive impact scores have been developed to identify splice sites, determine intron-exon boundaries (gene finding), and predict exonic splicing enhancer activity [87, 88]. Many of these programs represent open-source platforms freely available to sequence analysts (GeneSplicer, NNSPLICE, HumanSpliceFinder, NetGene2, MaxEntScan, EX-SKIP, RESCUE-ESE, ESEFinder) [89–98]. Additionally, subscription services like Alamut Visual (v2.9; Interactive Biosoftware) provide a set of splice site predictors allowing for consensus evaluation between multiple predictors [99]. While each prediction tool may have an individual set of strengths and weaknesses, users are likely to experience a higher degree of confidence when using multiple splice site predictors simultaneously [100]. Importantly, these computational models should not be used as a stand-alone tool for variant classification but rather as a supportive element in a manifold argument.

Variant Classification

Gene variants are currently being identified at a tremendous pace using retrospective observational data, numerous in silico prediction models, and better delineation of workflow and reporting. Recent endeavors such as the NCBI Genetic Testing Registry, MutaDATABASE, 1000 Genomes, and Human Variome Project draw attention to this growing interest in gene variant annotation and clinical interpretation in human disease [101–104]. Furthermore, accurate prediction of phenotypic severity for novel mutations and uncertain gene variants as relating to disease function is of great importance to medicine and biology. Informatics tools for predicting disease severity of uncertain gene variants may assist in the improvement of genetically informed patient care.

A standardized framework for evaluating potential disease association with novel and uncertain variants has been proposed that provides definitions and prioritization schemes for reporting [105]. Most laboratories proceed through heuristic variant classification by interrogating a number of databases and computational predictors to derive a composite probabilistic conclusion. Population (ExAC, gnoMAD, 1000 Genomes, dbSNP) and disease (ClinVar, OMIM, HGMD) databases cite specific variants by respective frequency. Computational in silico predictive algorithms provide pathogenicity scores based on evolutionary conservation, protein structural alteration, comparative impact, and Markov models. While initially developed for coding and splice site variants, newer algorithms are beginning to assess noncoding sequence variants [106, 107].

Generally, databases and prediction models find agreement (pathogenic, benign) unless the variant has not been described or has ambiguous predictive impact. In cases of disagreement or ambiguity, caution is advised, and the uncertainty should be addressed in the variant classification (e.g., VUS). In silico prediction results should not be used as the sole evidence for variant classification. Examples of in silico prediction results for a clear pathogenic variant with strong supporting evidence as well as a variant with unknown pathogenicity and conflicting prediction results are displayed in Figs. 14.7 and 14.8, respectively. Historically, many variants initially classified as VUS due to limited specific information have been subsequently reclassified as benign following years of retrospective monitoring and expanded population databases. Advantages of gene-specific algorithms

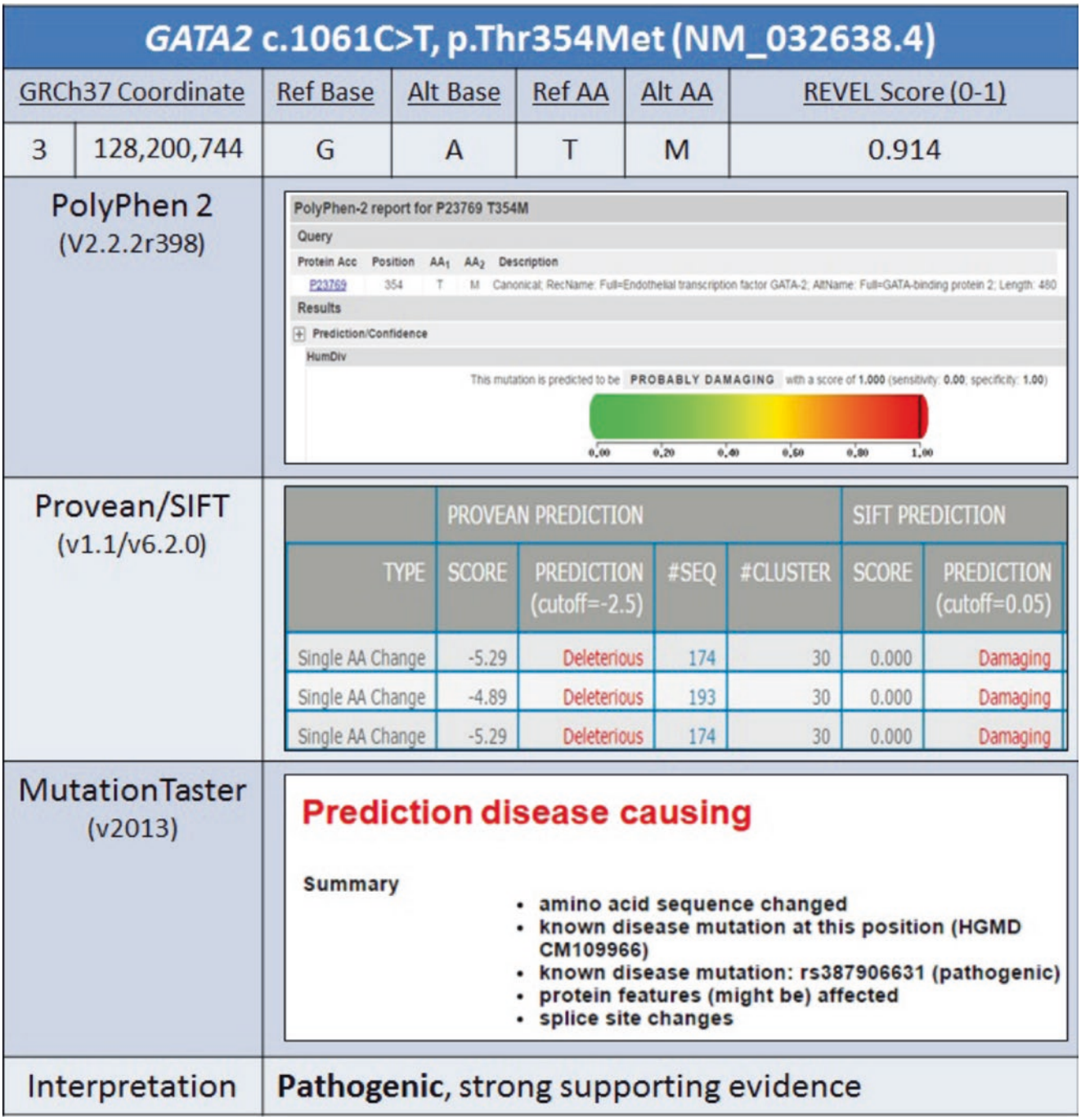


Fig. 14.7 Visualization of several in silico prediction scores for a *GATA2* germline missense variant with known pathogenicity. REVEL scores provide a weighted composite metric for missense variants drawn from 13 individual predictors. Scores are ranked from 0 to 1, with scores nearest 1 representing those most likely to be a disease-causing variant. In this example, all the predictors surveyed are in agreement supporting the variant classification as pathogenic [39, 41–43, 45, 47, 48, 74, 77, 82, 83, 86]

have been reported, while few in number, these algorithms are trained and tested on a well-defined disease setting with known genotype-phenotype outcomes on specific genes [108, 109]. Additionally, approaches that combine algorithm results such as REVEL, Condel, or Consensus have been reported, where benchmarking or ranking agreement of predicted phenotype severity across several complimentary algorithms may provide research priority for novel variants and VUS [86, 110, 111]. Figures 14.7 and 14.8 show the utility of REVEL scores which are based on a combination of scores from 13 computational predictors compacted into a single weighted scoring metric.

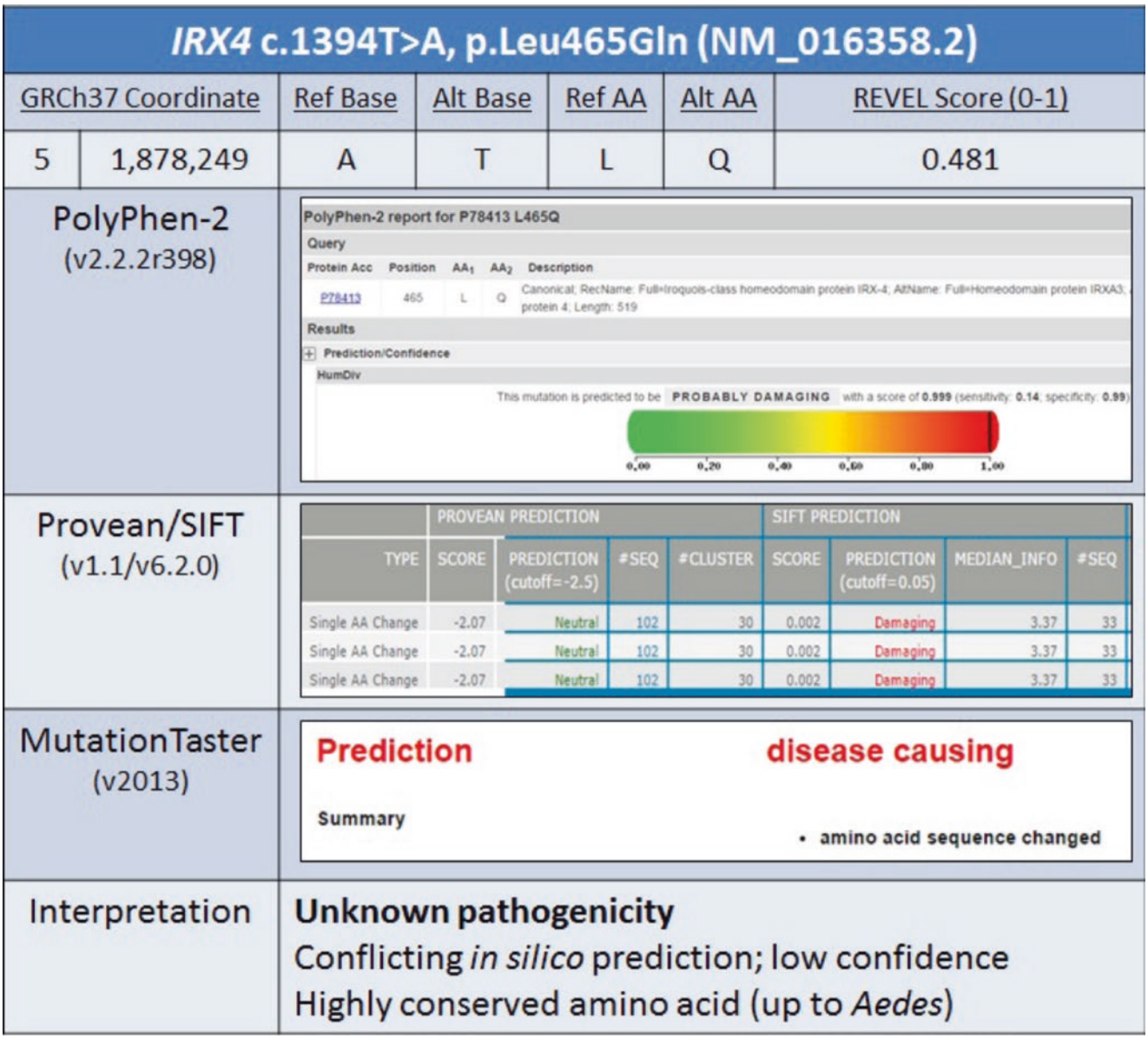


Fig. 14.8 Visualization of several in silico prediction scores for a *IRX4* germline missense variant with unknown significance. In this example, the REVEL score is near 0.5, and the prediction algorithms are conflicting with limited supportive evidence. Further review yields very little literature regarding germline disorders associated with this gene. In this example, the high degree of amino acid conservation and conflicting in silico prediction results would encourage classification of this germline variant as a VUS [39, 41–43, 45, 47, 48, 74, 77, 82, 83, 86]

Consensus guidelines provide laboratories with well-structured variant interpretation and classification schemes using a priority classification or tiered system. For example, Tier 1 somatic cancer variants represent definite disease-causing (pathogenic) alterations with expert agreement. Similarly, a germline variant meeting criterion for very strong evidence of pathogenicity, in the appropriate clinical setting, may be classified as PVS1 (pathogenic, very strong evidence). Subsequent classification levels traverse the spectrum from disease-causing toward benign (pathogenic, likely pathogenic, VUS, likely benign, benign) based on the strength of available evidence and predictive effect [105]. Comparative studies and training regarding these interpretation guidelines have significantly improved interlaboratory concordance (34% vs 71% over 99 variants) and reduced the need for reclassification [112]. Though not the focus of this section, recent guidelines regarding variant interpretation and reporting for somatic cancer NGS results have been published through joint consensus by Li and colleagues [113]. An added benefit of interpretation guidelines is greater harmony of variant classification within curated databases. This allows all laboratories, even those not specifically following ACMG guidelines, to derive more accurate interpretations when referencing these publically available resources. Over

time, previous classifications, particularly VUS, may need subsequent modification and reclassification. Many laboratories attempt to audit prior classification of variants in a cyclical nature. Clear policies regarding the reanalysis of genetic test results or variants should be provided by laboratories to ensure reasonable expectations. Ultimately, laboratories will need to maintain an adaptive and flexible outlook for approaching variant reclassification given the evolving nature of bioinformatics. Bioinformatics applications are likely to become vital to maintaining "real-time" adaptive variant classification by leveraging direct database interrogation, automatic secondary review prompts, and an interface with electronic record systems.

Conclusions

As bioinformatics moves into the mainstream of clinical laboratory workflow, several caveats may be appropriate based on key historical lessons from the genomics revolution. These include:

1. *Don't confuse more data with insight*: It can be difficult to extract clinically relevant conclusions from ever-increasing amounts of data in a reliable fashion.
2. *Don't confuse insight with value*: While many scientific findings may be interesting, they may do little to improve existing laboratory practices or to significantly improve current clinical outcomes.
3. *Don't overestimate one's ability to interpret the data*: Even the best data affords only limited insight into clinical health outcomes.
4. *Don't underestimate the implementation challenges*: Leveraging large data sets successfully requires a clinical laboratory system prepared to embrace and effectively handle new methodologies, requiring significant investment of time and capital, and the alignment of economic interests.
5. *Do thoroughly validate your bioinformatics pipeline*: NGS requires a purposeful effort to validate every stage of your pipeline, and established QC metrics with each run minimize the likelihood of missing or underestimating significant variants.

References

1. NHGRI Sequencing Costs. 2013. http://www.genome.gov/sequencingcosts/. Accessed 20 Feb 2013.
2. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.
3. Kircher M, Stenzel U, Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol. 2009;10(8):R83. https://doi.org/10.1186/gb-2009-10-8-r83.
4. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 2015;12(10):931–4. https://doi.org/10.1038/nmeth.3547.
5. Hogeweg P. The roots of bioinformatics in theoretical biology. PLoS Comput Biol. 2011;7(3):e1002021. https://doi.org/10.1371/journal.pcbi.1002021.
6. Huerta M, Haseltine F, Liu Y, Downing G, Seto B. NIH Working Definition of Bioinformatics and Computational Biology, The Biomedical Information Science and Technology Initiative Consortium (BISTIC) Definition Committee of National Institutes of Health (NIH); 2000.
7. Bioinformatics.org wiki. http://www.bioinformatics.org/wiki/Bioinformatics. Accessed 24 Feb 2013.
8. Moorthie S, Hall A, Wright CF. Informatics and clinical genome sequencing: opening the black box. Genet Med. 2013;15(3):165–71. https://doi.org/10.1038/gim.2012.116.
9. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198–203. https://doi.org/10.1038/nature09796.
10. Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Arch Pathol Lab Med. 2013;137(3):415–33. https://doi.org/10.5858/arpa.2012-0107-RA.
11. SEQanswers. http://seqanswers.com/. Accessed 12 Feb 2013.
12. Biostar - Bioinformatics Explained. http://www.biostars.org/. Accessed 12 Feb 2013.
13. BioIT-World. http://www.bio-itworld.com/. Accessed 12 Feb 2013.
14. Galaxy NGS server. https://main.g2.bx.psu.edu/. Accessed 12 Feb 2013.
15. NHGRI Short Course on NGS data. http://www.soph.uab.edu/ssg/courses/nhgri_r25. Accessed 12 Feb 2013.
16. AMIA 10x10 Courses. http://www.amia.org/education/10x10-courses. Accessed 12 Feb 2013.
17. Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. Brief Bioinform. 2011;12(5):489–97. https://doi.org/10.1093/bib/bbq077.
18. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8(3):175–85.
19. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8(3):186–94.
20. Ensembl. http://www.ensembl.org/Homo_sapiens/Info/Index. Accessed 30 May 2017.
21. NCBI Reference genome: Homo sapiens. https://www.ncbi.nlm.nih.gov/genome/?term=homo+sapiens. Accessed 30 May 2017.
22. UCSC Genome Bioinformatics. http://hgdownload.cse.ucsc.edu/downloads.html#human. Accessed 30 May 2017.
23. Burrows-Wheeler Aligner (BWA). http://bio-bwa.sourceforge.net. Accessed 15 Jan 2013.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:btp324 [pii]. https://doi.org/10.1093/bioinformatics/btp324.
25. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95. doi:btp698 [pii]. https://doi.org/10.1093/bioinformatics/btp698.
26. Novoalign. http://www.novocraft.com/main/index.php. Accessed 4 Mar 2013.
27. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11(5):473–83. doi:bbq015 [pii]. https://doi.org/10.1093/bib/bbq015.
28. Genome Analysis Toolkit (GATK). http://www.broadinstitute.org/gatk/. Accessed 4 Mar 2013.
29. SAMtools. http://samtools.sourceforge.net. Accessed 4 Mar 2013.

30. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6. doi:nbt.1754 [pii]. https://doi.org/10.1038/nbt.1754.
31. Integrative Genomics Viewer. http://software.broadinstitute.org/software/igv/. Accessed 22 June 2017.
32. Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. 2008;18(5):763–70. doi:gr.070227.107 [pii]. https://doi.org/10.1101/gr.070227.107.
33. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008;36(16):e105. doi:gkn425 [pii]. https://doi.org/10.1093/nar/gkn425.
34. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. Genome Res. 2011;21(9):1498–505. doi:gr.123638.111 [pii]. https://doi.org/10.1101/gr.123638.111.
35. de Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, Koudijs MJ, Nijman I, Hinrichs JW, Cuppen E, van Lieshout S, Loberg RD, de Jonge M, Voest EE, de Weger RA, Steeghs N, Langenberg MH, Sleijfer S, Willems SM, Lolkema MP. Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. PLoS One. 2016;11(2):e0149405. https://doi.org/10.1371/journal.pone.0149405.
36. Marchetti A, Del Grammastro M, Filice G, Felicioni L, Rossi G, Graziano P, Sartori G, Leone A, Malatesta S, Iacono M, Guetti L, Viola P, Mucilli F, Cuccurullo F, Buttitta F. Complex mutations & subpopulations of deletions at exon 19 of EGFR in NSCLC revealed by next generation sequencing: potential clinical implications. PLoS One. 2012;7(7):e42164. doi:PONE-D-12-12796 [pii]. https://doi.org/10.1371/journal.pone.0042164.
37. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, Ducar M, Van Hummelen P, Macconaill LE, Hahn WC, Meyerson M, Gabriel SB, Garraway LA. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer Discov. 2012;2(1):82–93. doi:2159-8290.CD-11-0184 [pii]. https://doi.org/10.1158/2159-8290.CD-11-0184.
38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. doi:gkq603 [pii]. https://doi.org/10.1093/nar/gkq603.
39. Annovar. http://www.openbioinformatics.org/annovar/. Accessed 4 Mar 2013.
40. SnpEff Annotation. http://www.broadinstitute.org/gatk/guide/tagged?tag=snpeff. Accessed 21 Nov 2012.
41. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812–4.
42. Sorting Intolerant from Tolerant (SIFT). http://sift.jcvi.org/. Accessed 4 Mar 2013.
43. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863–74. https://doi.org/10.1101/gr.176601.
44. PolyPhen2. http://genetics.bwh.harvard.edu/pph2/. Accessed 25 Feb 2013.
45. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20. https://doi.org/10.1002/0471142905.hg0720s76.
46. Mutation Taster. http://www.mutationtaster.org/index.html. Accessed 6 Dec 2012.
47. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575–6. doi:nmeth0810-575 [pii]. https://doi.org/10.1038/nmeth0810-575.
48. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361–2. https://doi.org/10.1038/nmeth.2890.
49. Torrent Suite Documentation. http://ioncommunity.lifetechnologies.com/docs/DOC-3343. Accessed 17 Jan 2013.
50. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn. 2017;19(3):341–65. https://doi.org/10.1016/j.jmoldx.2017.01.011.
51. Commission on Laboratory Accreditation CoAP. Commission on Laboratory Accreditation. Laboratory Accreditation Program: Molecular Pathology Checklist; 2016.
52. Schrijver I, Farkas DH, Gibson JS, Lyon E, Committee AMPE. The evolving role of the laboratory professional in the age of genome sequencing: a vision of the Association for Molecular Pathology. J Mol Diagn. 2015;17(4):335–8. https://doi.org/10.1016/j.jmoldx.2015.03.001.
53. Cucoranu IC, Parwani AV, West AJ, Romero-Lauro G, Nauman K, Carter AB, Balis UJ, Tuthill MJ, Pantanowitz L. Privacy and security of patient data in the pathology laboratory. J Pathol Inform. 2013;4:4. https://doi.org/10.4103/2153-3539.108542.
54. Yohe SL, Carter AB, Pfeifer JD, Crawford JM, Cushman-Vokoun A, Caughron S, Leonard DG. Standards for clinical grade genomic databases. Arch Pathol Lab Med. 2015;139(11):1400–12. https://doi.org/10.5858/arpa.2014-0568-CP.
55. Gargis AS, Kalman L, Bick DP, da Silva C, Dimmock DP, Funke BH, Gowrisankar S, Hegde MR, Kulkarni S, Mason CE, Nagarajan R, Voelkerding KV, Worthey EA, Aziz N, Barnes J, Bennett SF, Bisht H, Church DM, Dimitrova Z, Gargis SR, Hafez N, Hambuch T, Hyland FC, Luna RA, MacCannell D, Mann T, McCluskey MR, McDaniel TK, Ganova-Raeva LM, Rehm HL, Reid J, Campo DS, Resnick RB, Ridge PG, Salit ML, Skums P, Wong LJ, Zehnbauer BA, Zook JM, Lubin IM. Good laboratory practice for clinical next-generation sequencing informatics pipelines. Nat Biotechnol. 2015;33(7):689–93. https://doi.org/10.1038/nbt.3237.
56. Yandell M, Huff C, Hu H, Singleton M, Moore B, Xing J, Jorde LB, Reese MG. A probabilistic disease-gene finder for personal genomes. Genome Res. doi:gr.123158.111 [pii]. 2011; https://doi.org/10.1101/gr.123158.111.
57. Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, Swensen JJ, Johnson WE, Moore B, Huff CD, Bird LM, Carey JC, Opitz JM, Stevens CA, Jiang T, Schank C, Fain HD, Robison R, Dalley B, Chin S, South ST, Pysher TJ, Jorde LB, Hakonarson H, Lillehaug JR, Biesecker LG, Yandell M, Arnesen T, Lyon GJ. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. Am J Hum Genet. 2011;89:345. doi:S0002-9297(11)00210-2 [pii]. https://doi.org/10.1016/j.ajhg.2011.05.017.
58. Ionita-Laza I, Makarov V, Yoon S, Raby B, Buxbaum J, Nicolae DL, Lin X. Finding disease variants in Mendelian disorders by using sequence data: methods and applications. Am J Hum Genet. 2011;89(6):701–12. doi:S0002-9297(11)00481-2 [pii]. https://doi.org/10.1016/j.ajhg.2011.11.003.
59. Omicia. http://www.omicia.com/. Accessed 27 Feb 2013.
60. Ingenuity Systems. http://www.ingenuity.com/. Accessed 11 Mar 2013.
61. SV Bio. http://www.svbio.com/. Accessed 26 Feb 2013.
62. Agilent Technologies. http://www.agilent.com/en-us/solutions/clinical-grade-variant-assessment/cartagenia-bench-lab-for-clinical-genetics. Accessed 26 May 2017.
63. Pierian Dx. http://pieriandx.com. Accessed 26 May 2017.

64. Golden Helix Var Seq. http://goldenhelix.com/products/VarSeq/ Accessed 26 May 2017.
65. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet. 2012;49(6):353–61. https://doi.org/10.1136/jmedgenet-2012-100819.
66. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502–11. https://doi.org/10.1056/NEJMoa1306555.
67. Zhu X, Petrovski S, Xie P, Ruzzo EK, Lu YF, McSweeney KM, Ben-Zeev B, Nissenkorn A, Anikster Y, Oz-Levi D, Dhindsa RS, Hitomi Y, Schoch K, Spillmann RC, Heimer G, Marek-Yagel D, Tzadok M, Han Y, Worley G, Goldstein J, Jiang YH, Lancet D, Pras E, Shashi V, McHale D, Need AC, Goldstein DB. Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genet Med. 2015;17(10):774–81. https://doi.org/10.1038/gim.2014.191.
68. Hegde M, Santani A, Mao R, Ferreira-Gonzalez A, Weck KE, Voelkerding KV. Development and validation of clinical whole-exome and whole-genome sequencing for detection of germline variants in inherited disease. Arch Pathol Lab Med. 2017;141(6):798–805. https://doi.org/10.5858/arpa.2016-0622-RA.
69. Weinstein ND. What does it mean to understand a risk? Evaluating risk comprehension. J Natl Cancer Inst Monogr. 1999;1999(25):15–20.
70. Ensenauer RE, Michels VV, Reinke SS. Genetic testing: practical, ethical, and counseling considerations. Mayo Clin Proc. 2005;80(1):63–73.
71. Nowak R. Genetic testing set for takeoff. Science. 1994;265(5171):464–7.
72. Machens A, Gimm O, Hinze R, Hoppner W, Boehm BO, Dralle H. Genotype-phenotype correlations in hereditary medullary thyroid carcinoma: oncological features and biochemical properties. J Clin Endocrinol Metab. 2001;86(3):1104–9.
73. Wei Q, Wang L, Wang Q, Kruger WD, Dunbrack RL Jr. Testing computational prediction of missense mutation phenotypes: functional characterization of 204 mutations of human cystathionine beta synthase. Proteins. 2010;78(9):2058–74.
74. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073–81.
75. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009;25(21):2744–50.
76. Dorfman R, Nalpathamkalam T, Taylor C, Gonska T, Keenan K, Yuan XW, Corey M, Tsui LC, Zielenski J, Durie P. Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene? Clin Genet. 2010;77(5):464–73.
77. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002;30(17):3894–900.
78. Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics. 2005;21(14):3176–8.
79. Spencer DS, Stites WE. The M32L substitution of staphylococcal nuclease: disagreement between theoretical prediction and experimental protein stability. J Mol Biol. 1996;257(3):497–9.
80. Kang HH, Williams R, Leary J, Ringland C, Kirk J, Ward R. Evaluation of models to predict BRCA germline mutations. Br J Cancer. 2006;95(7):914–20.
81. Engelhardt BE, Jordan MI, Muratore KE, Brenner SE. Protein molecular function prediction by Bayesian phylogenomics. PLoS Comput Biol. 2005;1(5):e45.
82. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res. 2002;12(3):436–46.
83. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet. 2001;10(6):591–7.
84. Lopez-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpi JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic Acids Res. 2017;45:W222. https://doi.org/10.1093/nar/gkx313.
85. Pejaver VUJ, Lugo-Martinez J, Pagel KA, Lin GN, Nam H, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv. 2017; https://doi.org/10.1101/134981.
86. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877–85. https://doi.org/10.1016/j.ajhg.2016.08.016.
87. Zhang MQ. Statistical features of human exons and their flanking regions. Hum Mol Genet. 1998;7(5):919–32.
88. Splice Site Tools: A Comparative Analysis Report. http://www.ngrl.org.uk/Manchester/sites/default/files/publications/Informatics/NGRL_Splice_Site_Tools_Analysis_2009.pdf. Accessed 20 June 2017.
89. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 2001;29(5):1185–90.
90. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol. 1997;4(3):311–23. https://doi.org/10.1089/cmb.1997.4.311.
91. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67. https://doi.org/10.1093/nar/gkp215.
92. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res. 1996;24(17):3439–52.
93. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol. 1991;220(1):49–65.
94. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2–3):377–94. https://doi.org/10.1089/1066527041410418.
95. Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, Baralle D, Vorechovsky I. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. Hum Mutat. 2011;32(4):436–44. https://doi.org/10.1002/humu.21458.
96. Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res. 2004;32(Web Server issue):W187–90. https://doi.org/10.1093/nar/gkh393.
97. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Res. 2003;31(13):3568–3571kl.

98. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet. 2006;15(16):2490–508. https://doi.org/10.1093/hmg/ddl171.
99. Interactive Biosoftware Splicing Prediction Module. http://www.interactive-biosoftware.com/doc/alamut-visual/2.9/splicing.html. Accessed 20 June 2017.
100. Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, d'Enghien CD, Lauge A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat. 2008;29(7):975–82. https://doi.org/10.1002/humu.20765.
101. Javitt G, Katsanis S, Scott J, Hudson K. Developing the blueprint for a genetic testing registry. Public Health Genomics. 2010;13(2):95–105.
102. Bale S, Devisscher M, Van Criekinge W, Rehm HL, Decouttere F, Nussbaum R, Dunnen JT, Willems P. MutaDATABASE: a centralized and standardized DNA variation database. Nat Biotechnol. 2011;29(2):117–8.
103. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061–73.
104. Cotton RG, Al Aqeel AI, Al-Mulla F, Carrera P, Claustres M, Ekong R, Hyland VJ, Macrae FA, Marafie MJ, Paalman MH, Patrinos GP, Qi M, Ramesar RS, Scott RJ, Sijmons RH, Sobrido MJ, Vihinen M. Capturing all disease-causing mutations for clinical and research use: toward an effortless system for the Human Variome Project. Genet Med. 2009;11(12):843–9.
105. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24. https://doi.org/10.1038/gim.2015.30.
106. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5. https://doi.org/10.1038/ng.2892.
107. Salgado D, Bellgard MI, Desvignes JP, Beroud C. How to identify pathogenic mutations among all those variations: variant annotation and filtration in the genome sequencing era. Hum Mutat. 2016;37(12):1272–82. https://doi.org/10.1002/humu.23110.
108. Crockett DK, Piccolo SR, Ridge PG, Margraf RL, Lyon E, Williams MS, Mitchell JA. Predicting phenotypic severity of uncertain gene variants in the RET proto-oncogene. PLoS One. 2011;6(3):e18380. https://doi.org/10.1371/journal.pone.0018380.
109. Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubey IA, Li B, Bell R, Feng B, Mooney SD, Radivojac P, Sunyaev SR, Frebourg T, Hofstra RM, Sijmons RH, Boucher K, Thomas A, Goldgar DE, Spurdle AB, Tavtigian SV. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. Hum Mutat. 2013;34(1):255–65. https://doi.org/10.1002/humu.22214.
110. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011;88(4):440–9. https://doi.org/10.1016/j.ajhg.2011.03.004.
111. Crockett DK, Ridge PG, Wilson AR, Lyon E, Williams MS, Narus SP, Facelli JC, Mitchell JA. Consensus: a framework for evaluation of uncertain gene variants in laboratory test reporting. Genome Med. 2012;4(5):48. https://doi.org/10.1186/gm347.
112. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, Cooper GM, Dorschner MO, Dulik MC, Ghazani AA, Ghosh R, Green RC, Hart R, Horton C, Johnston JJ, Lebo MS, Milosavljevic A, Ou J, Pak CM, Patel RY, Punj S, Richards CS, Salama J, Strande NT, Yang Y, Plon SE, Biesecker LG, Rehm HL. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research consortium. Am J Hum Genet. 2016;99(1):247. https://doi.org/10.1016/j.ajhg.2016.06.001.
113. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, Tsimberidou AM, Vnencak-Jones CL, Wolff DJ, Younes A, Nikiforova MN. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn. 2017;19(1):4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.

Next-Generation Sequencing for Single-Gene Analysis

15

Hao Ho and Christopher D. Gocke

Introduction

The first generation of sequencing technologies was developed in the 1970s by Sanger [1, 2] and Maxam and Gilbert [3]. Frederick Sanger sequencing method is based on DNA synthesis making use of dideoxynucleotide analogues (radiolabeled or fluorescently labeled) to cause chain termination. By contrast, Allan Maxam and Walter Gilbert performed DNA sequencing through chemical degradation in which terminally labeled DNA fragments were chemically cleaved at specific bases and analyzed by gel electrophoresis. Since Maxam and Gilbert's method was more technically challenging and less amenable to being scaled up, Sanger sequencing ultimately prevailed and became the "gold standard" for decoding DNA sequence in the past four decades.

The first automated DNA sequencer, which performed partial automation of DNA sequence analysis through fluorescence detection of DNA fragments, was invented at Caltech in 1986 [4]. Subsequent improvement of the technology led to the introduction of the first commercial DNA sequencer by ABI in 1996, using slab-gel electrophoresis (ABI Prism 310). It was then improved upon 2 years later by the ABI Prism 3700 which utilized automated reloading of up to 96 capillaries filled with polymer matrix. While fully automated, the chief limitations of these capillary instruments are low throughput and high cost, resulting in it taking 13 years and nearly three billion dollars to complete the human genome sequencing project [5].

Due to the limitations of automated Sanger sequencing, new and improved technologies for sequencing large amounts of DNA have been developed in recent years, collectively referred to as next-generation sequencing (NGS). NGS, also known as massively parallel or high-throughput sequencing, relies on miniaturization of individual sequencing reactions by immobilizing spatially separated templates to a solid surface or support. This allows thousands to billions of individual sequencing reactions to be performed in parallel and be distinctly detected by digital imaging or electrical sensing, easily overcoming the limited scalability of "first-generation sequencing" by eliminating the electrophoresis step for sequence separation. The end results are vastly improved throughput at a fraction of the original cost, reducing the reagent costs from >$1000 down to 10 cents or less per megabase and putting the dream of whole-genome sequencing for $100–1000 within reach. The dramatic shifts in cost and accessibility brought by NGS not only have revolutionized the field of genomics but have also opened up a new world of medical diagnostics in which applications of NGS technology have a high impact now and in the foreseeable future [6–12].

Applications of Single-Gene NGS

Minimal Residual Disease Detection

For certain hematopoietic malignancies such as acute lymphoblastic leukemia (ALL), there is a strong correlation between the presence of posttreatment minimal residual disease (MRD) and adverse clinical outcomes [13–15]. As a result, accurate assessment of MRD is critical in risk stratification (standard, intermediate, or high risk) of ALL patients to predict clinical outcome and thus to provide guidance for proper management of the disease [16]. Currently, multiparameter flow cytometry (mpFC) and quantitative polymerase chain reaction (qPCR) are the main strategies for the assessment of MRD, but each method has its own drawbacks [17]. Multiparameter flow cytometry, which relies on the detection of a unique collection of antigens specific for leukemia, has sensitivity on the order of 1 cell in 10^4. However, data interpretation can be challenging and is operator- and/or laboratory-dependent and not infrequently is confounded by variable expression of leukemic antigens in the post-therapy setting. On the other hand, although PCR amplification of

H. Ho · C. D. Gocke (✉)
Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
e-mail: cgocke1@jhmi.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_15

immunoglobulin heavy chain (IGH) or T-cell receptor (TCR) genes or oncogenic fusion transcripts for MRD detection of leukemia may achieve higher sensitivity (1 cell in 10^5 or better), it often requires the use of patient-specific primers to assess the genetic rearrangement or translocation unique to each individual patient's disease—a task that can be expensive and labor-intensive with difficulty in achieving uniformity [18–20].

Theoretically, most, if not all, of the possible *IGH* or TCR rearrangement configurations can be sequenced by NGS using a set of consensus primers that are able to amplify all existing *IGH* or TCR segments. Automation of the procedure not only can eliminate operator dependency of data interpretation (as in mpFC) but also obviate the need to develop patient-specific reagents (for qPCR). In fact, Wu et al. [20] have shown that with targeted sequencing of *TCRB* and *TCRG* using an Illumina HiSeq NGS platform, they were able to detect MRD of T-ALL that was 10- to 100-fold lower than the limit of detection (LOD) of mpFC. However, the study was limited to the subtypes of T-ALL that had undergone TCR rearrangements (35 of 43 cases), and was not applicable to those at a more primitive stage (e.g., early thymic precursor immunophenotype). Additionally, Gawad et al. [21] showed that similar strategies can be applied in MRD monitoring of B-ALL by sequencing the *IGH* locus. Interestingly, their findings also provided new insight into the molecular mechanisms by which clonal evolution occurs in B-ALL patients. Single-gene targets for both solid tumor and hematopoietic neoplasms, used in the context of circulating tumor DNA (discussed elsewhere in the text), are also promising for MRD detection. In summary, although not yet a standard of care, the advent of NGS may present a more rapid, sensitive, informative, and cost-effective method for MRD testing in the future.

Oncologic Testing

In an era where target-based therapies are becoming a norm in the management of oncologic malignancies, pretreatment screening for predictive biomarkers is crucial in identification of cancer-specific genetic alteration(s) that are susceptible to available therapeutic modalities. Currently, this kind of "personalized" therapy is well established in the management of patients diagnosed with lung and colorectal cancers (e.g., EGFR and BRAF inhibitors), melanoma (e.g., BRAF inhibitors), and certain hematological malignancies (e.g., tyrosine kinase inhibitors in CML) but is expanding rapidly to other tumors as well. Traditional methods of mutational analysis such as Sanger sequencing, pyrosequencing, and allele-specific PCR have been widely used for this purpose. However, due to limited bandwidth and throughput of these older technologies, the depth of analysis has been confined to certain known mutational "hotspots" of individual genes—a practice that potentially can miss other significant genetic aberrations elsewhere. In order to supply the ever-growing breadth of information required to deliver truly personalized therapeutic interventions, increasingly the NGS approach has been utilized for cancer genomics analysis because it has the ability to simultaneously detect various genetic alterations in thousands of different genes in a single run [22–29].

Using the Roche GS Junior 454 NGS platform, a small pilot experiment by Borras et al. analyzing FFPE samples from colorectal and lung cancers showed that the approach is efficient and accurate in detecting all existing *KRAS* mutations [23]. Shindoh et al. used the SOLiD 4 platform to perform cDNA screening on primary specimens and cell lines derived from lung cancer, breast cancer, and melanoma and found that the system can efficiently identify various genetic alterations in *EGFR*, *KRAS*, *NRAS*, and *ERBB2* genes [24]. With targeted sequencing of *BRCA1* and *BRCA2* genes using two NGS platforms (SOLiD 4 and Ion Torrent PGM), Chan et al. reported that both systems are highly sensitive and specific for single nucleotide polymorphisms (SNPs), though the PGM platform lacked specificity in insertion/deletion (indel) calling [25]. It is clear that the single-gene approach is being built on by manufacturers and diagnostic labs, resulting in panels of various sizes directed at cancer-related genes broadly or at tumor-specific or drugable genes in particular. For example, a commercial "extended *RAS*" panel examines only six exons in the *NRAS* and *KRAS* genes in colorectal cancer for targeting with anti-*EGFR* antibody therapy [26]. A companion diagnostic NGS panel targeting only the *BRCA1* and *BRCA2* genes is available as a laboratory-developed test [27].

Besides solid tumors, NGS is also being applied in screening of individual genetic alterations and monitoring of disease progression in hematologic malignancies such as chronic myelomonocytic leukemia (e.g., testing the *TET2*, *CBL*, *RAS*, and *RUNX1* genes) [30, 31], myelodysplastic syndrome (e.g., *TP53*) [30], and myeloproliferative neoplasms (e.g., *JAK2*) [29]. Moreover, Grossmann et al. proved that NGS can be used successfully in the assessment of GC-rich genes such as *CEBPA* (in AML patients) and found it to be highly sensitive for mutation analysis of this gene [28].

The great depth of coverage inherent in NGS provides both benefits and drawbacks to single-gene or small panel testing. For example, most NGS tests have the ability to easily identify mutations in a target gene when they represent as little as 5% of the total. Molecular indexing or barcoding lowers this limit of detection by one to two orders of magnitude [32]. While this is useful in analyzing challenging clinical samples with a low-tumor burden, it makes initial validation or confirmation by some other methods with lower limits of detection (particularly Sanger sequencing) difficult or impossible. The detection of small numbers of

mutated molecules also challenges existing paradigms. For example, many patients are found to have low levels of the drug-resistant mutant T790M in the *EGFR* gene prior to therapy with small molecule inhibitors of the gene; the clinical significance of this finding, particularly with regard to therapy, remains unclear [33]. Similarly, gastrointestinal stromal tumors with *KIT* or *PDGFRA* activating mutations develop resistance to tyrosine kinase inhibitors such as imatinib, sometimes due to secondary mutations in *KIT*. The detection of such secondary resistance mutations may permit adjustment of therapy [34]. Tumor heterogeneity is being characterized in a number of other genes by sensitive methods. Will the detection of a minor subclone with a sensitizing mutation have the same import as the mutation in the bulk of the tumor? How hard will diagnostic laboratories have to look for such subclones? Studies with highly sensitive assays will be needed to answer these and similar questions.

Infectious Diseases

Infectious disease diagnosis and screening are additional areas in which NGS can have major impacts, gradually replacing the traditional molecular tests that are based mainly on Sanger sequencing. In the diagnosis and genotyping of hepatitis C (HCV) and human immunodeficiency viral (HIV) infections, NGS allows the completion of both steps in a single reaction [35, 36]. Once a diagnosis of infection is established, assessment of intra-patient viral genetic variation becomes crucial for evaluation of viral evolutionary dynamics and identifying emerging resistant strains not only in order to provide guidance for optimal antiviral therapy but also to serve as a valuable source of information for designing effective vaccines. In this regard, NGS has revolutionized the field by simplifying the once time-consuming and expensive assessment of intrahost viral genetic diversity of HCV and HIV into a cost-effective procedure at an unprecedented resolution [35, 37]. Moreover, similar to the oncologic target-based therapy, NGS is being applied to assess the co-receptor tropism of HIV-1 prior to treatment with CCR5 antagonist [38].

Not surprisingly, the use of NGS is being extended to the detection and classification of other known viruses (e.g., HPV genotyping) [39] and screening for unknown disease-causing microorganisms in pathology samples [40]. With its ability to simultaneously detect multiple infectious agents, NGS has been proven invaluable in the metagenomic analysis of infectious diseases during local outbreaks (e.g., norovirus), pandemics (e.g., avian influenza), or global epidemics (e.g., seasonal influenza virus) [41–43]. Infectious agents of global significance such as Zika virus may be identified by NGS, with subsequent epidemiologic analyses permitting case tracking [44]. The fields of biodefense against bioterrorism will also benefit from NGS as this new technology can not only rapidly detect the presence of specific pathogenic agent(s) but also perform subtyping/subclassification and drug resistance profiling at the same time, which can expedite implementation of counter-terrorism measures [45–47].

Inherited Diseases

Candidate genes responsible for inherited disorders have traditionally been identified through linkage studies [48]. Currently, a known genetic cause has been assigned to more than 6000 Mendelian disorders (http://www.ncbi.nlm.nih.gov/omim). Although classical genome-wide linkage studies are effective at elucidating causal variants for some inherited diseases, those that are sporadic or extremely rare or occur de novo are usually not amenable to this method. With the advent of NGS, whole-exome/whole-genome sequencing became feasible and served as a powerful tool for probing the genetic defects of those rare syndromes or complex diseases whose etiologies remained elusive [49, 50].

As mentioned above, the major contribution of NGS in medical genetics thus far lies not in screening of known single-gene mutations but rather in the discovery of allelic variants or novel genetic pathways associated with rare inherited syndromes that are beyond the reach of traditional linkage analysis. For instance, since its original description in 1981, the underlying cause of Kabuki syndrome—a rare, sporadic disorder with multiple congenital anomalies—had remained intractable to conventional approaches of gene discovery. Through massively parallel sequencing of the exomes of ten unrelated probands, Ng et al. were able to demonstrate that Kabuki syndrome is due to mutations in the *MLL2* gene [51]. Using similar strategies, the same group of scientists was also able to uncover the underlying cause of Miller syndrome, another rare inherited disease [52]. Likewise, Hoischen et al. were able to characterize Schinzel-Giedion syndrome as an entity secondary to de novo mutations in the *SETBP1* gene by analyzing the exomes from only four affected individuals [53]. Interesting inheritance patterns are also being discovered in genetic disorders such as Gillespie syndrome [54]. More clarification of the genetic basis of diseases will undoubtedly occur in the near future using genome and exome sequencing as a tool [49, 55].

As in other fields of study, the ability to decode multiple genes in parallel (e.g., in gene panels) allows NGS to be applied in the screening and monitoring of complex diseases such as inherited retinal degeneration (IRD) for which genetic testing has become increasingly important for proper diagnosis, prognosis, and development of personalized therapy [56].

Human Leukocyte Antigen Typing

The human leukocyte antigen (HLA) class I and class II gene loci consist of more than 7000 alleles, giving rise to >4600 distinct HLA proteins, and thus are the most polymorphic genes in the human genome known to date [57]. The *HLA-B7* gene [58] was the first HLA gene to be cloned (in 1980) and was subsequently used as a probe in Southern blot analyses to mark the advent of restriction fragment length polymorphism (RFLP) in the study of HLA genomic polymorphism [59]. This cumbersome method was later replaced by the sequence-specific oligonucleotide approach in combination with PCR [60–62] and the use of sequence-based typing for procedures requiring high-resolution HLA typing, such as hematopoietic stem cell transplantation [63]. Besides being labor-intensive, time-consuming, and expensive (due to the highly polymorphic nature of the HLA loci), typing ambiguity remains a critical challenge for the current methods secondary to their limitations in genomic coverage and the difficulties in determining the cis-trans relationships between variants [57].

A massively parallel sequencing approach can help to overcome HLA typing ambiguity by virtue of its ability to perform deep sequencing with high coverage of the entire HLA region, combined with clonal amplification to provide in-phase sequencing of linked polymorphisms [57, 64]. Lind et al. sequenced six known samples using NGS and obtained 100% concordance in all analyzed HLA loci [64]. In a double-blind study that enrolled 8 independent laboratories to genotype the same 20 samples for multiple HLA loci, using the 454 GS FLX platform coupled with CONEXIO ATF software, Holcomb et al. were able to achieve an overall concordance of 97.2% with the known genotypes [65], pointing to the interlaboratory reliability of this approach in high-resolution HLA genotyping. Moreover, Erlich et al. developed a novel NGS protocol for HLA class I typing and showed the superiority of this method relative to the current sequence-specific oligonucleotide-based gold standard in terms of typing accuracy while maximizing throughput and minimizing cost, providing concrete support for NGS as a reliable, efficient, and scalable approach for HLA typing [66]. More recently, massive genotyping projects for registry purposes have demonstrated the feasibility of NGS approaches, highlighting their ability to make novel findings while efficiently analyzing routine samples [67]. Although turnaround times are currently problematic for many transplant-related needs, this may be solved with higher throughput or more specialized, purpose-built equipment [68].

Archeological and Mitochondrial Studies

The study of ancient DNA began in the early 1980s with amplification of small DNA sequences using bacterial cloning followed by sequencing [69, 70]. This inefficient and labor-intensive technique became obsolete with the development of PCR a few years later [71, 72], which combined with cloning and Sanger sequencing to form the classical methodology in molecular archeology [73]. Traditionally, mitochondrial DNA (mtDNA) is used as a target primarily because it is present in many copies per cell in contrast to the two copies of nuclear DNA. Moreover, the much smaller mitochondrial genome size and lack of mtDNA recombination also simplify data analysis. Even so, only targeted regions of the mtDNA are used in most studies due to technical limitations (low throughput and high cost) of the classical methodology [74].

Development of NGS has opened up new possibilities in the field of archeology. The new technology not only can sequence the complete mtDNA genome with relative ease but also renders the previously unthinkable whole nuclear genome sequencing of an extinct species a distinct possibility. In fact, using massively parallel sequencing, the complete genomes of three long extinct hominid groups (Neanderthals, Denisovans, and Palaeo-Eskimo) were decoded in 2010 [75–77]. By uncovering the genetic diversity and composition of our ancient ancestors through whole-genome sequencing, NGS has helped to overcome a major restriction confronted by the classical methodology in human evolutionary study. In summary, the high sensitivity and efficiency of NGS have markedly enhanced our ability to generate vast amounts of high-quality data from ancient DNA in a relatively short time—a feat that in turn will help to unfold the evolutionary history of human and other species with unprecedented resolution and rapidity.

Mitochondrial disorders are also amenable to identification by NGS. These include neurologic and neuromuscular disorders such as MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes) and MERRF (myoclonic epilepsy with ragged red fibers). These conditions are often heterogeneous in their phenotype, in part because of the relative distributions of mutant and wild-type molecules in the affected tissues due to the cytoplasm mode of transmission [78]. The small size of the mitochondrial genome (about 16,000 bases, smaller than the span of most nuclear genes) permits easy targeting, and the quantitative nature of NGS permits accurate genotyping. In theory, scientists could also employ NGS to study mitochondrial genomes for forensic purposes, but this methodology may be too unproven to meet the required standards [79].

Other Clinical Applications

NGS has emerging applications in other areas such as forensic studies [80–82], post-bone marrow transplant engraftment testing [83], monitoring of transplanted organ rejection [84], and prenatal screening [85, 86]. With maturation and

continual improvement, there is no reason to doubt that NGS will eventually find its way into many other different fields of study and clinical applications as well.

Next-Generation Sequencing Platforms

Due to continual advancement of current technologies, new and improved NGS platforms are being introduced at a breakneck pace. We will only briefly touch upon the few currently commercially available NGS platforms here, with more details provided in Chaps. 1 and 2. The Ion Torrent Personal Genome Machine and Proton use sequencing-by-synthesis technology but detect pH changes inherent in nucleotide incorporation rather than light emitted from dye-containing nucleotide incorporation. A variety of Illumina products (MiSeq, HiSeq, NextSeq, and NovaSeq) use a now-standard fluorescent dye-based technology for detection. They differ in their capacity, run time, patterning of flow cells, dye compositions, and cost. Other emerging technologies have not yet made their way to clinical laboratories but remain promising because of special qualities such as the ability to perform very long reads or the avoidance of expensive dyes. These include nanopore sequencing technology, pyrosequencing, and the cyclic reversible termination (CRT) sequencing method.

Cost of Next-Generation Sequencing

The cost of DNA sequencing has taken roughly a four-order-of-magnitude plunge since the advent of NGS in 2005. According to data from the National Human Genome Research Institute (www.genome.gov), NGS has helped to reduce the raw cost of sequencing one megabase of DNA from $1000 in 2005 to a mere 1 cent in 2017. In 2005, sequencing a whole genome cost approximately 17 million dollars, while by 2017, it had dropped to slightly above $1000. It is important to note that raw sequencing costs do not include annotation and interpretation, which are now the most expensive part of the process. Joking reference to the "$1,000 genome/$1,000,000 interpretation" makes the point that the bioinformatic tools and medical infrastructure needed to place the data in context have not yet matured. The cost of DNA sequencing varies among the different NGS platforms. In a more realistic scenario within a regular DNA diagnostic lab in which smaller-scale DNA sequencing is the norm, the cost savings between NGS and Sanger sequencing are less dramatic. While comparing the cost of *BRCA* mutation screening using two NGS platforms (SOLiD 4 and Ion Torrent PGM), Chan et al. reported that NGS systems can afford more than twofold (Ion Torrent PGM) to more than fourfold (SOLiD 4) cost savings relative to Sanger sequencing. Moreover, turnaround time was reduced dramatically relative to Sanger sequencing [26]. This, however, very much depends on the exact application and will be quite different for single-gene applications compared to, for example, whole-genome sequencing.

Conclusions

NGS has helped to unwind, at an unprecedented pace, the mysteries embedded in the complicated genomes of human and other organisms. The efficiency, scalability, and affordability of NGS technologies will also turn whole-exome or whole-genome sequencing into a routine assay in clinical labs in the near future—a feat that was unthinkable just a few years ago with Sanger sequencing. Although promising, NGS is still in its infancy in the realm of clinical molecular diagnostics. With its impressive and ever-expanding range of applications, there is no doubt that NGS will have a tremendous impact on the future of personalized medicine.

References

1. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94:441–8.
2. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology. 1992;24:104–8.
3. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977;74:560–4.
4. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321:674–9.
5. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931–45.
6. Shendure J, Aiden EL. The expanding scope of DNA sequencing. Nat Biotechnol. 2012;30:1084–94.
7. Natrajan R, Reis-Filho JS. Next-generation sequencing applied to molecular diagnostics. Expert Rev Mol Diagn. 2011;11:425–44.
8. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet. 2010;19:R227–40.
9. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52:413–35.
10. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, Shi L. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn. 2011;11:333–43.
11. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010;11:31–46.
12. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
13. Flohr T, Schrauder A, Cazzaniga G, Panzer-Grumayer R, van der Velden V, Fischer S, Stanulla M, Basso G, Niggli FK, Schafer BW, Sutton R, Koehler R, Zimmermann M, Valsecchi MG, Gadner H, Masera G, Schrappe M, van Dongen JJ, Biondi A, Bartram CR. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. Leukemia. 2008;22:771–82.

14. Schrappe M, Valsecchi MG, Bartram CR, Schrauder A, Panzer-Grumayer R, Moricke A, Parasole R, Zimmermann M, Dworzak M, Buldini B, Reiter A, Basso G, Klingebiel T, Messina C, Ratei R, Cazzaniga G, Koehler R, Locatelli F, Schafer BW, Arico M, Welte K, van Dongen JJ, Gadner H, Biondi A, Conter V. Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: results of the AIEOP-BFM-ALL 2000 study. Blood. 2011;118:2077–84.
15. van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L, Stolz F, Schrappe M, Masera G, Kamps WA, Gadner H, van Wering ER, Ludwig WD, Basso G, de Bruijn MA, Cazzaniga G, Hettinger K, van der Does-van den Berg A, Hop WC, Riehm H, Bartram CR. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet. 1998;352:1731–8.
16. Willemse MJ, Seriu T, Hettinger K, d'Aniello E, Hop WC, Panzer-Grumayer ER, Biondi A, Schrappe M, Kamps WA, Masera G, Gadner H, Riehm H, Bartram CR, van Dongen JJ. Detection of minimal residual disease identifies differences in treatment response between T-ALL and precursor B-ALL. Blood. 2002;99:4386–93.
17. Campana D. Progress of minimal residual disease studies in childhood acute leukemia. Curr Hematol Malig Rep. 2010;5:169–76.
18. Thol F, Kolking B, Damm F, Reinhardt K, Klusmann JH, Reinhardt D, von Neuhoff N, Brugman MH, Schlegelberger B, Suerbaum S, Krauter J, Ganser A, Heuser M. Next-generation sequencing for minimal residual disease monitoring in acute myeloid leukemia patients with FLT3-ITD or NPM1 mutations. Genes Chromosomes Cancer. 2012;51:689–95.
19. Campana D. Minimal residual disease monitoring in childhood acute lymphoblastic leukemia. Curr Opin Hematol. 2012;19:313–8.
20. Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML, Greisman HA, Sabath DE, Wood BL, Robins H. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci Transl Med. 2012;4:134ra63.
21. Gawad C, Pepin F, Carlton VE, Klinger M, Logan AC, Miklos DB, Faham M, Dahl G, Lacayo N. Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. Blood. 2012;120:4407–17.
22. Cronin M, Ross JS. Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology. Biomark Med. 2011;5:293–305.
23. Borras E, Jurado I, Hernan I, Gamundi MJ, Dias M, Marti I, Mane B, Arcusa A, Agundez JA, Blanca M, Carballo M. Clinical pharmacogenomic testing of KRAS, BRAF and EGFR mutations by high resolution melting analysis and ultra-deep pyrosequencing. BMC Cancer. 2011;11:406.
24. Shindoh N, Yoda A, Yoda Y, Sullivan TJ, Weigert O, Lane AA, Kopp N, Bird L, Rodig SJ, Fox EA, Weinstock DM. Next-generation cDNA screening for oncogene and resistance phenotypes. PLoS One. 2012;7:e49201.
25. Chan M, Ji SM, Yeo ZX, Gan L, Yap E, Yap YS, Ng R, Tan PH, Ho GH, Ang P, Lee AS. Development of a next-generation sequencing method for BRCA mutation screening: a comparison between a high-throughput and a benchtop platform. J Mol Diagn. 2012;14:602–12.
26. https://www.accessdata.fda.gov/cdrh_docs/pdf16/p160038c.pdf. Accessed 27 Apr 2018.
27. https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160018B.pdf. Accessed 27 Apr 2018.
28. Grossmann V, Schnittger S, Schindela S, Klein HU, Eder C, Dugas M, Kern W, Haferlach T, Haferlach C, Kohlmann A. Strategy for robust detection of insertions, deletions, and point mutations in CEBPA, a GC-rich content gene, using 454 next-generation deep-sequencing technology. J Mol Diagn. 2011;13:129–36.
29. Kohlmann A, Grossmann V, Haferlach T. Integration of next-generation sequencing into clinical practice: are we there yet? Semin Oncol. 2012;39:26–36.
30. Kohlmann A, Grossmann V, Klein HU, Schindela S, Weiss T, Kazak B, Dicker F, Schnittger S, Dugas M, Kern W, Haferlach C, Haferlach T. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. J Clin Oncol. 2010;28:3858–65.
31. Delhommeau F, Dupont S, Della VV, James C, Trannoy S, Masse A, Kosmider O, Le Couedic JP, Robert F, Alberdi A, Lecluse Y, Plo I, Dreyfus FJ, Marzac C, Casadevall N, Lacombe C, Romana SP, Dessen P, Soulier J, Viguie F, Fontenay M, Vainchenker W, Bernard OA. Mutation in TET2 in myeloid cancers. N Engl J Med. 2009;360:2289–301.
32. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci. 2011;108:9530–5. https://doi.org/10.1073/pnas.1105422108.
33. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene. 2009;28(Suppl 1):S24–31.
34. Chen W, Kuang Y, Qiu HB, Cao Z, Tu Y, Sheng Q, Eilers G, He Q, Li HL, Zhu M, Wang Y, Zhang R, Wu Y, Meng F, Fletcher JA, Ou WB. Dual targeting of insulin receptor and KIT in imatinib-resistant gastrointestinal stromal tumors. Cancer Res. 2017;77:5107–17.
35. Escobar-Gutierrez A, Vazquez-Pichardo M, Cruz-Rivera M, Rivera-Osorio P, Carpio-Pedroza JC, Ruiz-Pacheco JA, Ruiz-Tovar K, Vaughan G. Identification of hepatitis C virus transmission using a next-generation sequencing approach. J Clin Microbiol. 2012;50:1461–3.
36. Redd AD, Collinson-Streng A, Martens C, Ricklefs S, Mullis CE, Manucci J, Tobian AA, Selig EJ, Laeyendecker O, Sewankambo N, Gray RH, Serwadda D, Wawer MJ, Porcella SF, Quinn TC. Identification of HIV superinfection in seroconcordant couples in Rakai, Uganda, by use of next-generation deep sequencing. J Clin Microbiol. 2011;49:2859–67.
37. Beerenwinkel N, Gunthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol. 2012;3:329.
38. Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, Paxinos E, Arts EJ, Robertson DL, Mimms L, Quinones-Mateu ME. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One. 2012;7:e49602.
39. Meiring TL, Salimo AT, Coetzee B, Maree HJ, Moodley J, Hitzeroth II, Freeborough MJ, Rybicki EP, Williamson AL. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. Virol J. 2012;9:164.
40. Katano H, Sato S, Sekizuka T, Kinumaki A, Fukumoto H, Sato Y, Hasegawa H, Morikawa S, Saijo M, Mizutani T, Kuroda M. Pathogenic characterization of a cervical lymph node derived from a patient with Kawasaki disease. Int J Clin Exp Pathol. 2012;5:814–23.
41. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One. 2009;4:e4219.
42. Nakamura S, Nakaya T, Iida T. Metagenomic analysis of bacterial infections by means of high-throughput DNA sequencing. Exp Biol Med (Maywood). 2011;236:968–71.
43. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Curr Opin Virol. 2012;2:63–77.
44. Gu SH, Song DH, Lee D, Jang J, Kim MY, Jung J, Woo KI, Kim M, Seog W, Oh HS, Choi BS, Ahn JS, Park Q, Jeong ST. Whole-genome sequence analysis of Zika virus, amplified from urine of traveler from the Philippines. Virus Genes. 2017;53:918–21.
45. Serizawa M, Sekizuka T, Okutani A, Banno S, Sata T, Inoue S, Kuroda M. Genomewide screening for novel genetic variations

associated with ciprofloxacin resistance in Bacillus anthracis. Antimicrob Agents Chemother. 2010;54:2787–92.
46. Wright AM, Beres SB, Consamus EN, Long SW, Flores AR, Barrios R, Richter GS, Oh SY, Garufi G, Maier H, Drews AL, Stockbauer KE, Cernoch P, Schneewind O, Olsen RJ, Musser JM. Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. Arch Pathol Lab Med. 2011;135:1447–59.
47. Kuroda M, Sekizuka T, Shinya F, Takeuchi F, Kanno T, Sata T, Asano S. Detection of a possible bioterrorism agent, Francisella sp., in a clinical specimen by use of next-generation direct DNA sequencing. J Clin Microbiol. 2012;50:1810–2.
48. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003;33(Suppl):228–37.
49. Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV. Developing genome and exome sequencing for candidate gene identification in inherited disorders. Arch Pathol Lab Med. 2013;137:415–33.
50. Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparso T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jorgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang X, Li R, Yang H, Wang J, Hansen T, Pedersen O, Nielsen R, Wang J. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet. 2010;42:969–72.
51. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42:790–3.
52. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42:30–5.
53. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet. 2010;42:483–5.
54. Gerber S, Alzayady KJ, Burglen L, Bremond-Gibnac D, Marchesin V, Roche O, Rio M, Funalot B, Calmon R, Durr A, Gil-da-Silva-Lopes VL, Ribeiro Bittar MF, et al. Recessive and dominant de novo ITPR1 mutations cause Gillespie syndrome. Am J Hum Genet. 2016;98:971–80.
55. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. Hum Genet. 2011;129:351–70.
56. Shanks ME, Downes SM, Copley RR, Lise S, Broxholme J, Hudspith KA, Kwasniewska A, Davies WI, Hankins MW, Packham ER, Clouston P, Seller A, Wilkie AO, Taylor JC, Ragoussis J, Nemeth AH. Next-generation sequencing (NGS) as a diagnostic tool for retinal degeneration reveals a much higher detection rate in early-onset disease. Eur J Hum Genet. 2013;21:274–80.
57. Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens. 2012;80:1–11.
58. Sood AK, Pereira D, Weissman SM. Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. Proc Natl Acad Sci U S A. 1981;78:616–20.
59. Erlich HA, Stetler D, Sheng-Dong R, Ness D, Grumet C. Segregation and mapping analysis of polymorphic HLA class I restriction fragments: detection of a novel fragment. Science. 1983;222:72–4.
60. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature. 1986;324:163–6.
61. Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc Natl Acad Sci U S A. 1989;86:6230–4.
62. Blake E, Mihalovich J, Higuchi R, Walsh PS, Erlich H. Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: casework experience. J Forensic Sci. 1992;37:700–26.
63. Santamaria P, Lindstrom AL, Boyce-Jacino MT, Myster SH, Barbosa JJ, Faras AJ, Rich SS. HLA class I sequence-based typing. Hum Immunol. 1993;37:39–50.
64. Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, Walker R, Hsiao T, McLaughlin L, D'Arcy M, Gai X, Goodridge D, Sayer D, Monos D. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum Immunol. 2010;71:1033–42.
65. Holcomb CL, Hoglund B, Anderson MW, Blake LA, Bohme I, Egholm M, Ferriola D, Gabriel C, Gelber SE, Goodridge D, Hawbecker S, Klein R, Ladner M, Lind C, Monos D, Pando MJ, Proll J, Sayer DC, Schmitz-Agheguian G, Simen BB, Thiele B, Trachtenberg EA, Tyan DB, Wassmuth R, White S, Erlich HA. A multi-site study using high-resolution HLA genotyping by next generation sequencing. Tissue Antigens. 2011;77:206–17.
66. Erlich RL, Jia X, Anderson S, Banks E, Gao X, Carrington M, Gupta N, DePristo MA, Henn MR, Lennon NJ, de Bakker PI. Next-generation sequencing for HLA typing of class I loci. BMC Genomics. 2011;12:42.
67. Schöfl G, Lang K, Quenzel P, Böhme I, Sauter J, Hofmann JA, Pingel J, Schmidt AH, Lange V. 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. BMC Genomics. 2017;18:161.
68. Bravo-Egana V, Monos D. The impact of next-generation sequencing in immunogenetics: current status and future directions. Curr Opin Organ Transplant. 2017;22:400–6.
69. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312:282–4.
70. Paabo S. Molecular cloning of Ancient Egyptian mummy DNA. Nature. 1985;314:644–5.
71. Paabo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A. 1989;86:1939–43.
72. Paabo S, Higuchi RG, Wilson AC. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. J Biol Chem. 1989;264:9709–12.
73. Cooper A, Poinar HN. Ancient DNA: do it right or not at all. Science. 2000;289:1139.
74. Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: new perspectives on old samples. Genet Sel Evol. 2012;44:21.
75. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la RM, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Paabo S. A draft sequence of the Neandertal genome. Science. 2010;328:710–22.
76. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good

JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Paabo S. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468:1053–60.

77. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Gronnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Ponten T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010;463:757–62.
78. Sproule DM, Kaufmann P. Mitochondrial encephalopathy, lactic acidosis, and strokelike episodes. Ann N Y Acad Sci. 2008;1142:133–58.
79. Bandelt HJ, Salas A. Current next generation sequencing technology may not meet forensic standards. Forensic Sci Int Genet. 2012;6:143–5.
80. Zietkiewicz E, Witt M, Daca P, Zebracka-Gala J, Goniewicz M, Jarzab B, Witt M. Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet. 2012;53:41–60.
81. Berglund EC, Kiialainen A, Syvanen AC. Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet. 2011;2:23.
82. Alvarez-Cubero MJ, Saiz M, Martinez-Gonzalez LJ, Alvarez JC, Eisenberg AJ, Budowle B, Lorente JA. Genetic identification of missing persons: DNA analysis of human remains and compromised samples. Pathobiology. 2012;79:228–38.
83. Tiercy JM. Immunogenetics of hematopoietic stem cell transplantation: the contribution of microsatellite polymorphism studies. Int J Immunogenet. 2011;38:365–72.
84. Snyder TM, Khush KK, Valantine HA, Quake SR. Universal noninvasive detection of solid organ transplant rejection. Proc Natl Acad Sci U S A. 2011;108:6229–34.
85. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, van den BD, Bombard AT, Deciu C, Grody WW, Nelson SF, Canick JA. DNA sequencing of maternal plasma to detect down syndrome: an international clinical validation study. Genet Med. 2011;13:913–20.
86. Palomaki GE, Deciu C, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, van den BD, Bombard AT, Grody WW, Nelson SF, Canick JA. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as down syndrome: an international collaborative study. Genet Med. 2012;14:296–305.

Next-Generation Sequencing for Gene Panels

16

Michael O. Dorschner

Introduction

Next-generation sequencing (NGS) technologies have revolutionized molecular diagnostics over the past several years. These rapidly evolving platforms are moving quickly from the research bench to the clinical laboratory. The enormous outputs of these massively parallel technologies have dramatically reduced the cost of DNA sequencing [1]. A single sequencing run can generate from a few gigabases to over half a terabase of data. Analysis and storage of data on this scale are no small undertaking, requiring specialized bioinformaticists, computational biologists, and substantial information technology (IT) infrastructure. To remain competitive, molecular diagnostic laboratories have converted conventional Sanger sequencing-based assays to NGS-based gene panels.

Molecular diagnostic laboratories often begin implementing NGS with indication or disease-targeted multigene panels [2]. Panel-based approaches have been applied to the diagnosis of heterogeneous disorders with overlapping, difficult-to-distinguish phenotypes. Multigene panels for cardiomyopathies [3], cancer predisposition [4], and X-linked intellectual disability [5] were among the first NGS assays to be launched by clinical laboratories (Table 16.1). With disease-targeted panels, all of the clinically relevant genes, from 10 or more to well over 500, can be examined concurrently, putting an end to the serial gene-by-gene diagnostic odysseys imposed by traditional Sanger-based approaches. NGS gene panels offer several advantages by (1) reducing the time a clinician spends on test selection, (2) reducing the collective turnaround time from test initiation to the reporting of results to patients, and (3) limiting testing to only those genes with proven clinical utility for a given phenotype.

M. O. Dorschner (✉)
Department of Pathology, UW Medicine Center for Precision Diagnostics, Northwest Clinical Genomics Laboratory, Seattle, WA, USA
e-mail: mod@uw.edu

A number of academic and commercial laboratories are offering clinical exome and genome sequencing. Exome and genome sequencing come with a variety of issues to consider: (1) the likelihood of incidental findings, not relevant to the indication for which the test was ordered and whether the lab is responsible for reporting these variants; (2) a lack of sequence coverage for disease-relevant loci, which are likely to be better targeted with panel testing; and (3) the increased likelihood of identifying and reporting more variants of unknown significance, as laboratories are typically not genome-wide experts. Laboratories that offer disease-targeted panels often have expertise relevant to the interpretation of the variants detected by the assay. A growing body of evidence suggests diagnostic yield can be improved by using whole-genome sequencing (WGS) as a first-tier test. WGS has the capacity to uncover a wider spectrum of sequence variants, ranging from single-nucleotide variants to larger structural rearrangements which may make cytogenetic microarrays redundant [6]. The American College of Medical Genetics and Genomics currently recommends exome or genome sequencing only for cases in which a disease-targeted panel is likely to produce a negative result or for disorders for which a targeted test is not available [7].

Table 16.1 Next-generation sequencing gene panels offered by clinical laboratories

Next-generation sequencing gene panel	Number of genes
Hereditary cancer	10–100
Somatic/tumor cancer	10–200
Cardiomyopathies (dilated, hypertrophic)	50–70
Arrhythmias	10–30
Hearing loss (syndromic, non-syndromic)	23–72
Neurodegenerative (dementia, Parkinson's, amyotrophic lateral sclerosis, dystonia)	4–75
X-linked mental retardation	30–150
RASopathies	10–20
Mitochondrial disorders	35–400

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_16

Clinical laboratories typically offer focused panels for phenotypes caused by 250 or fewer known genes. Developing and validating iterative panels can be cumbersome and expensive. Some laboratories have chosen to use an exome as a backbone for many tests, simply annotating only the genes relevant to a specific condition, thereby creating virtual panels. If a panel is negative, the provider can reflex to the exome and add family members to the investigation when needed. Virtual panels can be updated as new genes gain evidence for causality, without the need to revalidate the entire test. Such a strategy allows the lab to remain current with their test offerings. Several labs, including GeneDx, PreventionGenetics and the University of Washington Northwest Clinical Genomics lab, enable providers to create their own panels through Xome Slice, PGxome Custom Panel, and Panel-on-Demand assays. As panels become increasingly larger for phenotypes such as intellectual disability, epilepsy, and autism spectrum disorder, an exome-based approach becomes more feasible.

Along with the commercialization of several NGS platforms came technologies for performing target enrichment [8]. Polymerase chain reaction (PCR), the mainstay of Sanger-based diagnostics, is not practical for processing multigene panels at an appreciable scale. Technologies based on highly multiplexed PCR, such as TruSeq Custom Amplicon Assays™ (Illumina), AmpliSeq™ (Life Technologies), and droplet PCR (RainDance Technologies), have recently been launched. However, these assays remain more expensive than hybridization-based enrichment strategies. Customizable, in-solution hybridization-based technologies include SureSelect™ (Agilent Technologies), SeqCap EZ™ (NimbleGen), TargetSeq™ (Life Technologies), and xGen™ Lockdown™ Probes (Integrated DNA Technologies). These platforms, based on hybridization of biotinylated RNA or DNA probes to sample fragment libraries, are used to capture from 1 kilobase sequences to genomic regions of more than 24 megabases [9–11] (Fig. 16.1). Users can choose

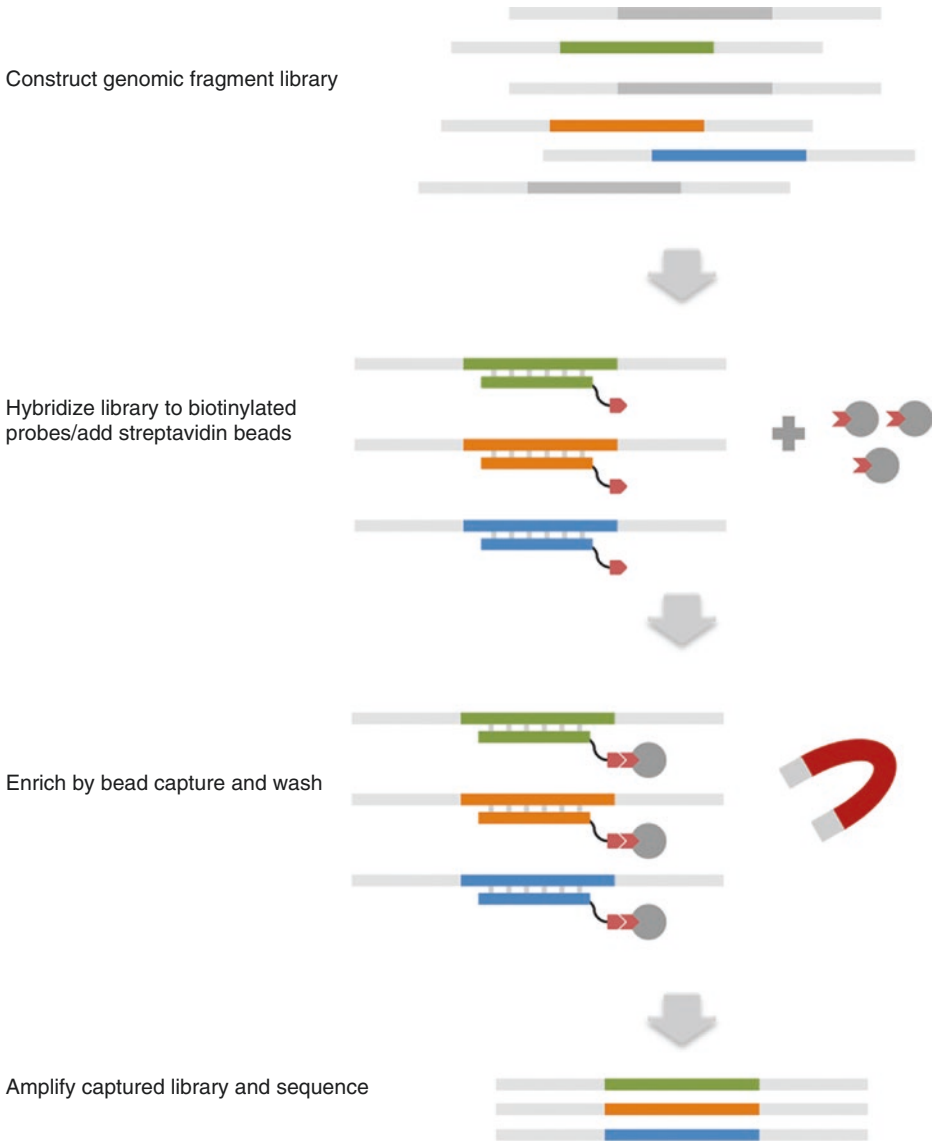


Fig. 16.1 In-solution sequence capture for target enrichment. After the construction of a genomic fragment library, the library is hybridized to capture probes and enriched. Subsequently, the captured library can be amplified and sequenced

to target coding segments or entire transcribed regions, within the limitations of the capture and sequencing technologies. In-solution, hybridization-based approaches are widely used among molecular diagnostic laboratories. Many clinical laboratories have tested multiple platforms, ultimately focusing on one enrichment strategy, in an effort to streamline laboratory workflows.

Panel-based testing can be performed with gene-specific capture probes or by tailored informatic analysis of exome or genome data. This chapter focuses on the development of individual panels through the design of custom target enrichment probes, restricted to specific genes; however, a growing number of clinical laboratories have begun using the exome as a backbone for a multitude of panels, ranging from ten genes to hundreds. By limiting variant annotation to only those genes on a specified list, a "virtual" panel can be analyzed. Providers can also order custom panels, "on demand," to fit their patient's needs or when no such test is clinically available (i.e., UW Medicine Center for Precision Diagnostics (Panel-on-Demand), PreventionGenetics (PGxome Custom Panel), or Greenwood Genetic Center (Focused Exome)). Utilization of the exome for panel test sequence data allows providers to reflex to the remainder of the exome should the panel not reveal a causal variant. Using the exome as a backbone for panel testing also allows laboratories to update panels more rapidly without the need for constant assay revalidation. While this approach is gaining momentum, many labs continue to implement multigene panels with separate assays.

In this chapter, an example of a procedure for NGS-based gene panel testing is provided. This is only one of the many possible examples, using equipment and reagents that could be exchanged for others, but it provides insight into the conceptual workflow of NGS-based testing of gene panels in a laboratory-developed assay setting. The example procedure was developed using the Agilent SureSelect in-solution sequence capture system for target enrichment, followed by sequencing with an Illumina HiSeq or MiSeq system. The procedure commences with the construction of sample-specific fragment libraries with platform-specific indexed adapters, followed by enrichment of the libraries for sequence targets of interest, sequencing of those targets, and analysis (Fig. 16.2).

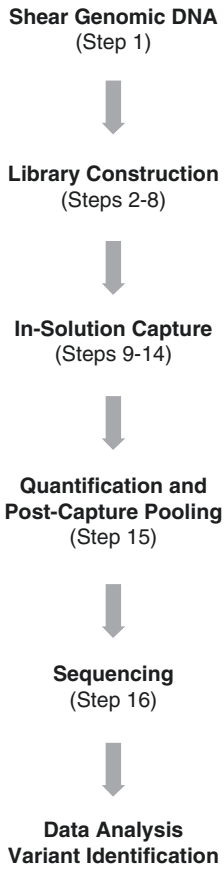


Fig. 16.2 Step-by-step workflow. This example of a laboratory procedure for an NGS-based gene panel assay begins with the construction of sample-specific fragment libraries with platform-specific indexed adapters, followed by enrichment of the libraries for sequence targets of interest, sequencing of those targets, and data analysis

Assay Design and Considerations for Developing an NGS Gene Panel

After selecting a panel of genes for assay development, the user of our example wet-bench procedure will need to design enrichment probes. Custom probe designs can be easily generated using, for example, the online Agilent SureDesign software (https://earray.chem.agilent.com/suredesign). Depending on the target of interest, probes can be designed to capture only exons or entire genomic regions. If copy number variants (CNVs) will need to be ascertained, it is advised to design probes across the entire genomic region, starting 10 kilobases upstream of the first exon and ending 10 kilobases downstream of the final exon. This should provide sufficient read depth to call small insertions/deletions using some of the CNV detection algorithms [12].

In addition to selection of the target regions, the user must also consider the existence of repetitive regions, such as segmental duplications or paralogs and highly repetitive elements, including short interspersed elements (SINEs) and long interspersed elements (LINEs). Depending on the sequence similarity of these loci, they can be difficult to enrich and sequence. It can also be challenging to align reads accurately to the reference. With the Illumina platform, the current maximum read length on a HiSeq is 100 and 250 bp on a MiSeq. Longer paired-end reads can dramatically improve the accuracy of read mapping, but this will depend on the length of identity between or among paralogous regions of the genome.

GC content can influence the ability to capture and sequence targets. The first exons of many genes are GC rich. Despite the inclusion of probes to enrich libraries for these regions, read coverage is often low and sometimes absent over these segments. The user may need to include additional probes over these regions to enhance sequence capture.

After the design is completed, the user can review the results by downloading what is called a BED file. The BED file can be uploaded to the UCSC Genome Browser (http://genome.ucsc.edu/) to view probe positions and target coverage. If the desired targets are not adequately covered by probes, the user can redesign the probe set by modifying design parameters in the SureDesign software. When the custom design has been finalized, the user can request an online quote and consider proceeding with placing an order.

Example Procedure

Materials

1. Qubit dsDNA HS Quantification Kit (Life Technologies, P/N (part number) Q32851)
2. P20, P200, P1000 filtered pipet tips (Rainin, P/N RT-L10F, RT-L200F, and RT-L1000)
3. Covaris microTUBE AFA Fiber Screw-Cap 6 × 16 mm (P/N 520096) for the M220 series, Covaris microTUBE AFA Fiber Pre-Slit Snap-Cap 6 × 16 mm (P/N 520045) for the S220 series, *or* Covaris 96 microTUBE Plate (P/N 520078) for the E220 series
4. Ultrapure Distilled Deionized Water (Life Technologies, P/N 10977-015)
5. Eppendorf LoBind Microcentrifuge Tubes 1.5 mL (Fisher, P/N 13-6987-91)
6. 0.2 mL PCR Tube Strips, 8-Tube, full-height with domed caps (Bio-Rad, P/N TBC-0201)
7. 0.2 mL PCR Tube Strips, 8-Tube, low profile (Bio-Rad, P/N TLS-0801)
8. Optical Flat 8-Cap Strips (Bio-Rad, P/N TCS-0803)
9. Domed 8-Cap Strips (Bio-Rad, P/N TCS-0801)
10. Minicentrifuge, with tube strip rotor (Fisher, P/N 05-090-100)
11. NEB end-repair kit (NEB, P/N E6050L)
12. NEB Next A-tailing Kit (NEB, P/N E6053L)
13. Agencourt AMPure XP Beads, 60 mL (Beckman Coulter, P/N A63881)
14. 70% Ethanol (see section "Reagent and Oligonucleotide Preparation")
15. 20% PEG/2.5 M NaCl solution (see section "Reagent and Oligonucleotide Preparation")
16. T4 Ultrapure Ligase and 10× T4 UltraPure Buffer (Enzymatics Inc., P/N L603-HC-L)
17. KAPA Library Amplification Readymix (KAPA Biosystems, P/N KK2612)
18. DNA 1000 kit (Agilent, P/N 5067-1504)
19. SureSelect XT Custom Library (Agilent, part numbers vary depending on the size of the target region)
20. Dynabeads M-270 Streptavidin (Life Technologies, P/N 65305)
21. Library Quantification Kit/Illumina/Universal (Kapa Biosystems, P/N KK4824)
22. M Tris–HCl, pH 7.5 (Fisher, P/N BP1758-100)
23. Tween 20 (Fisher, P/N BP337-100)
24. Oligonucleotides (Integrated DNA Technologies or other commercial vendors) (Tables 16.2 and 16.3)

Equipment

1. P10, P20, P200, and P1000 pipets (Rainin, P/N L-10XLS, L-20XLS, L-200XLS, and L-1000XLS)
2. Covaris® M220, S220, or E220 Focused-Ultrasonicator, connected to a computer with SonoLab™ software
3. Microcentrifuge (Eppendorf 5430)
4. One 24 position Tube IsoRack with 0 °C IsoPack (Eppendorf, P/N 22510053)
5. Dyna-Mag 2 Magnetic Stand (Life Technologies, P/N 12321D)

Table 16.2 Oligonucleotide sequences

Oligo name	Sequence (5′–3′)
Universal	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
Index	/5P/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC [idx] ATCTCGTATGCCGTCTTCTGCTTG
Block-U	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Block-U-RC	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
Block-I	CAAGCAGAAGACGGCATACGAGAT [idx] GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
Block-I-RC	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC [idx] ATCTCGTATGCCGTCTTCTGCTTG
TS-PCR-1	AATGATACGGCGACCACCGAGA
TS-PCR-2	CAAGCAGAAGACGGCATACGAG

U universal, *RC* reverse complement, *idx* index, *5P* 5′ phosphorylation, *asterisk* phosphorothioate bond, *I* index-containing, *TS* TruSeq™, *PCR* polymerase chain reaction

Table 16.3 Index sequences

ID	Sequence
1	ATCACG
2	CGATGT
3	TTAGGC
4	TGACCA
5	ACAGTG
6	GCCAAT
7	CAGATC
8	ACTTGA
9	GATCAG
10	TAGCTT
11	GGCTAC
12	CTTGTA
13	AGTCAA
14	AGTTCC
15	ATGTCA
16	CCGTCC
18	GTCCGC
19	GTGAAA
20	GTGGCC
21	GTTTCG
22	CGTACG
23	GAGTGG
25	ACTGAT
27	ATTCCT

ID identification number

6. 2100 Bioanalyzer (Agilent Technologies)
7. Speed vacuum with tube or plate adaptor (Eppendorf Vacufuge Plus)
8. HiSeq 1000/2000, HiSeq 1500/2500, or MiSeq system (Illumina)
9. CFX96 Touch Real-Time PCR Detection System (Bio-Rad, P/N 185-5484)
10. T100 Thermal Cycler (Bio-Rad, P/N 186-1096)
11. Qubit 2.0 Fluorometer (Life Technologies, P/N Q32866)

Reagent and Oligonucleotide Preparation

1. 100 mL 20% PEG/NaCl solution
   (a) 50 mL 5 M NaCl (Life Technologies, P/N AM9759)
   (b) 20 g PEG 8000 (Fisher, P/N BP233-100)
   (c) Water, up to 100 mL total volume
   Autoclave for 15 min. Allow the solution to cool because separation may occur upon heating.
2. 100 mL 70% ethanol:
   In a graduated cylinder, measure 70 mL 100% ethanol. Add 30 mL of molecular grade water, and mix.
3. Generation of 50 μM adapter stocks:
   (a) Resuspend the universal adapter oligo and indexed oligos at 100 μM in 10 mM Tris pH 8.0 and 50 mM NaCl.
   (b) For each indexed adaptor, combine 25 μL of index oligo and 25 μL of universal oligo in a clean 0.2 mL PCR tube.
   (c) Heat to 95 °C for 5 min in a thermal cycler.
   (d) Remove the tubes from the thermal cycler, and briefly spin to ensure that contents are at the bottoms of the tubes. Allow to cool at room temperature for 30 min.
4. Prepare index-specific hybridization blockers (ISHB):
   (a) Resuspend the Block-U, Block-U-RC, Block-I, and Block-I-RC oligos at 1000 μM in 10 mM Tris pH 8.0 and 50 mM NaCl.
   (b) For each index used, combine 0.5 μL of each oligo Block-U, Block-U-RC, Block-I, and Block-I-RC. 0.6 μL of this stock will be used during the hybridization reaction setup in Step 9. This blocker solution can be used in place of the Agilent-provided blockers.
5. Prepare library amplification oligonucleotides:
   (a) Resuspend TS-PCR-1 and TS-PCR-2 oligos at 100 μM in 10 mM Tris pH 8.0.

Procedure

This procedure provides step-by-step instructions for generating genomic fragment libraries and subsequent target enrichment, using the Agilent SureSelect in-solution capture system. With minor modification, the libraries generated by this protocol could be enriched with any other hybridization-based approach. This protocol is designed for processing batches of eight samples. Volumes and quantities can be scaled according to the needs of the user.

Step 1: Shear DNA

Covaris provides a shearing guide for each instrument model with recommended settings to generate specific fragment sizes. The user should test these settings prior to shearing valuable samples. The optimal insert size for 2 × 100 bp paired-end sequencing is 250 bp (225–275 bp). The average insert size will be larger than the total of the paired-end reads to avoid generating overlapping sequence at the 3′ ends.

1. Quantify DNA using the Qubit dsDNA quantification kit according to the manufacturer's instructions.
2. Prepare the Covaris instrument. Ensure that the reservoir is filled with deionized water and the water temperature has equilibrated and degassed, prior to use.
3. Dilute 1 μg of high-quality gDNA with 1× low TE buffer in a 1.5 mL LoBind tube to a total volume of 130 μL.
4. Transfer the 130 μL of DNA sample to the proper Covaris microTUBE, making sure not to introduce bubbles.

5. Secure the microTUBE in the tube holder, and shear the DNA using the appropriate settings to generate fragments of 150–200 bp (or other fragment size ranges, depending on specific needs).
6. Optional: Repeat DNA quantification prior to proceeding to Step 2.

Step 2: End Repair

Shearing will create double-stranded fragments with overhangs. A combination of T4 polynucleotide kinase and T4 DNA polymerase will convert these to 5′-phosphorylated, blunt ends.

1. Prepare a master mix from the components listed in Table 16.4 in a clean 1.5 mL LoBind tube on ice or a cooling rack.
2. Dispense 11 µL of master mix into each 0.2 mL PCR tube.
3. Dispense 50 µL of sheared DNA to each master mix-containing PCR tube. Mix carefully and thoroughly by pipetting up and down ten times.
4. Briefly spin to bring contents to the bottoms of the tubes.
5. Incubate tubes in a thermal cycler for 30 min at 20 °C.
6. Remove AMPure XP beads from the refrigerator, and gently shake to resuspend beads. The beads must be at room temperature prior to use.
7. After the 20 °C incubation, shake the AMPure XP beads to ensure complete resuspension, and transfer 110 µL of bead solution to each end-repaired DNA.
8. Mix and incubate for 5 min at room temperature.
9. Place tubes on a strip/plate magnet for 5 min to separate beads from solution.
10. Visually confirm that the beads have moved to the side of the tube and the solution is clear. Aspirate 171 µL of clear solution from each tube without disturbing the beads, and discard.
11. Dispense 180 µL of freshly prepared 70% ethanol to each tube, and incubate for 30 s at room temperature. Aspirate ethanol, and repeat for a total of two washes. Let beads dry for 5 min. Under- or overdrying beads will reduce yield.
12. Remove tubes from the magnet, add 43 µL of nuclease-free water, and mix. Do not elute DNA off of the beads. The beads and DNA will be carried forward into subsequent reactions.
13. Samples can be stored at −20 °C if not proceeding to the next step.

Step 3: Adenylate End-Repaired DNA

1. Prepare a master mix from the components listed in Table 16.5 in a clean 1.5 mL LoBind tube on ice or cooling rack.
2. Dispense 8 µL of adenylation master mix into each of the 0.2 mL PCR tubes.
3. Add 42 µL of end-repaired DNA (including beads) to each master mix-containing PCR tube. Mix carefully and thoroughly by pipetting up and down ten times.
4. Briefly spin to bring contents to the bottoms of the tubes.
5. Incubate in a thermal cycler for 30 min at 37 °C.
6. Vortex the 20% PEG/2.5 M NaCl solution. The solution must be at room temperature before use.
7. Transfer 90 µL of PEG solution to each tube containing the A-tailed DNA, and mix 20+ times. Incubate for 5 min at room temperature.
8. Place the tube on a magnet for 2 min to separate the beads from the solution.
9. Visually confirm that the beads have moved to the side of the tube and the solution is clear. Aspirate 140 µL of clear solution from each reaction tube, and discard.
10. Dispense 200 µL of freshly prepared 70% ethanol to each tube, and incubate for 30 s at room temperature. Aspirate the ethanol, and discard. Repeat for a total of two washes. Let the beads dry for 5 min. Under- or overdrying beads will reduce yield.
11. Remove the tubes from the magnet, add 38 µL of nuclease-free water, and mix.
12. Place the tube back on the magnet. Wait for the solution to clear, and remove 1 µL of solution for quantification. Remove the tube from the magnet, and mix briefly. Do not elute DNA off of the beads.
13. Samples can be stored at −20 °C if not proceeding to the next step.

Table 16.4 End-repair master mix components

Reagent	Per reaction (µL)	For 8 reactions (µL)
10× End-repair buffer	6	52.8
End-repair enzyme mix	5	44
Total	11	96.8

Table 16.5 Adenylation master mix components

Reagent	Per reaction (µL)	For 8 reactions (µL)
10× dA-tailing buffer	5	44
Klenow fragment	3	26.4
Total	8	70.4

Step 4: Quantification of Adenylated DNA

1. Using the 1 µL of solution retrieved from Step 12 in section "Step 3: Adenylate End-Repaired DNA," quantify the DNA using the Qubit dsDNA HS Quantification Kit.
2. Note any samples that are less than 10 ng/µL. Samples at lower concentrations may exhibit reduced library complexity, potentially impacting variant detection.

Step 5: Ligate Indexed Adapters to Adenylated DNA

Prior to performing the ligation, determine how the samples will be sequenced. If the samples will be pooled, each will require a specific indexed adaptor so that reads can be demultiplexed after sequencing.

1. Remove T4 Ultrapure Ligase, 10× T4 UltraPure Buffer, and indexed adapters (50 µM) from the freezer. Prepare ligation master mix by combining the reagents listed in Table 16.6.
2. Remove Agencourt AMPure XP beads and 20% PEG solution from the refrigerator, and equilibrate to room temperature for at least 30 min. The beads and PEG solution must be at room temperature before use. Vortex very well to resuspend all beads.
3. Vortex and centrifuge the adenylated DNA from *Step 3* and place on ice or cold block.
4. Prepare ligation master mixes by combining the following for each sample:
   (a) Dilute the indexed 50 µM adapters 1:3 with ultrapure distilled water into clean 1.5 mL microfuge tubes to make a 16.7 µM solution.
   (b) Add 5 µL of diluted adaptor to each sample.
5. Add 10 µL of ligation master mix to each sample tube, and mix thoroughly by gently pipetting up and down ten times. Avoid bubble formation.
6. Briefly spin to bring contents to the bottoms of the tubes.
7. Incubate at room temperature (25 °C) for 15 min.
8. Briefly spin to bring contents to the bottoms of the tubes, and proceed with bead cleanup.
9. Vortex the 20% PEG/2.5 M NaCl solution.
10. Transfer 90 µL of PEG solution to the tube containing the ligated DNA.
11. Mix 20 times, and incubate for 5 min at room temperature.
12. Place the tube on a strip/plate magnet for 2 min to separate the beads from solution.
13. Visually confirm that the beads have moved to the side of the tube and the solution is clear.
14. Aspirate 140 µL of clear solution from the reaction tube, and discard.
15. Dispense 180 µL of freshly prepared 70% ethanol to each well of the tube, and incubate for 30 s at room temperature. Aspirate the ethanol, and discard. Repeat for a total of two washes. Let the beads dry for 5 min. Under- or overdrying beads will reduce yield.
16. Remove the tube from the magnet, add 50 µL of nuclease-free water, and mix.
17. Place tube on magnet for 1 min to separate beads.
18. Aspirate and transfer 50 µL eluent to a new labeled tube.

Step 6: Amplify Pre-capture Library

1. Thaw PCR primers (TS-PCR-1 and TS-PCR-2) at room temperature, and keep KAPA Library Amplification Mix on ice.
2. Label two 0.2 mL PCR strip tubes per sample to prepare for two 50 µL PCRs each.
3. Spin down TS-PCR-1 and TS-PCR-2 (20 µM) primer tubes and the indexed libraries.
4. Divide each library prepared in the previous step into the two labeled PCR tubes (22.5 µL/tube).
5. Prepare 50 µL of master mix for each library/sample by combining the reagents listed in Table 16.7.
6. Add 27.5 µL of amplification master mix to each library-containing tube.
7. Seal with PCR tube caps, and centrifuge the PCR tubes briefly, and place in a thermal cycler.
8. Amplify libraries with the thermal cycling profile in Table 16.8.
9. Store at 4 °C or continue with Step 7.

Table 16.6 Ligation master mix components

Reagent	Per reaction (µL)	For 8 reactions (µL)
10× T4 UltraPure Buffer	5	44
Ultrapure Ligase	5	44
Total	10	88

Table 16.7 Pre-capture library amplification components

Reagent	Per reaction (µL)	For 8 reactions (µL)
TS-PCR-1 (20 µM)	2.5	44
TS-PCR-2 (20 µM)	2.5	44
KAPA Library Amplification Mix	50	880
Total	55	968

Table 16.8 Pre-capture library amplification thermal cycling protocol

Step	Temperature/duration	Cycles
Activation/denaturation	98 °C, 30 s	1
Amplification	98 °C, 10 s	8
	60 °C, 30 s	
	72 °C, 30 s	
Final extension	72 °C, 5 min	1
Hold	16 °C	

Step 7: Purify Amplified Pre-capture Libraries

1. Remove AMPure XP beads from 4 °C and keep at room temperature for at least 30 min. Vortex generously to resuspend all beads.
2. Pool the 2 × 50 µL PCR reactions (per library) into a 1.5 mL tube.
3. Shake the Agencourt AMPure XP beads to resuspend.
4. Dispense 180 µL of beads into each pooled PCR reaction.
5. Mix well, and incubate for 5 min at room temperature.
6. Place the reaction tube onto a Dyna-Mag 2 rack for 5 min to separate beads from the solution.
7. Visually confirm that the beads have moved to the side of the tube and the solution is clear.
8. Aspirate 280 µL of clear solution from the tubes, and discard.
9. Dispense 200 µL of 70% ethanol to each tube, and incubate for 30 s at room temperature. Aspirate the ethanol, and discard. Repeat for a total of two washes. Let the beads dry for 5 min. Under- and overdrying beads will reduce yield.
10. Take the tubes off the magnet, add 50 µL water to each reaction tube, and mix. Place the reaction plate onto a magnet for 1 min to separate the beads.
11. Transfer 50 µL eluent to a new labeled tube.

Step 8: Pre-capture Library Assessment

1. Assess the fragment size distribution and concentration by running a sample of each library on Bioanalyzer 2100 (DNA 1000 Chip).
2. When assessing the fragment size distribution, keep in mind that the average fragment size should be the size of the sheared DNA from Step 1 plus 100 bp to account for the ligation of the adapters. If your target insert size is 200 bp, you should expect to see an average post-capture fragment size of 300 bp (±25 bp). If the fragment size distribution is significantly above or below the expected size, i.e., 100 bp smaller or larger, you should consider repeating the procedure.

Step 9: Hybridization with Sequence Capture Probes

Blockers are added to inhibit the adapters from cross hybridizing and "daisy chaining" during the hybridization. Full-length index-specific blockers are more efficient at inhibiting cross hybridization than blockers targeted to the common portions of the adapters. The Agilent-provided blockers are replaced with the blockers described in section "Reagent and Oligonucleotide Preparation."

1. Pre-capture indexed library must be at a concentration of at least 147–221 ng/µL. Use a vacuum concentrator to concentrate the samples, if needed. Do not heat above 45 °C.
2. Combine the reagents listed in Table 16.9 to generate the hybridization buffer.
3. Warm the hybridization buffer to 65 °C if a precipitate forms.
4. Label three clean 8 PCR Tube Strips: *A*, *B*, and *C*.
5. Dispense 40 µL of hybridization buffer into each tube of strip *A*.
6. Prepare a SureSelect RNase block dilution according to Table 16.10. Note: The amount of RNAse block solution needed will depend on the size of the capture target.
7. Combine the appropriate amount of diluted RNase block solution and SureSelect capture library according to Table 16.11 in strip *C*. Mix by pipetting. Each tube should contain 7 µL of RNase block/capture library mixture.
8. Prepare SureSelect Block Mix by combining the components listed in Table 16.12. The SureSelect Index Block #3 is replaced with the ISHB assembled in section "Reagent and Oligonucleotide Preparation."

Table 16.9 Hybridization buffer components

Reagent	Per reaction (µL)	For 8 reactions (µL)
SureSelect Hyb #1	25	200
SureSelect Hyb #2 (red cap)	1	8
SureSelect Hyb #3 (yellow cap)	10	80
SureSelect Hyb #4	13	104
Total	49 (40 µL/rxn)	392 (40 µL/rxn)

Rxn reaction

Table 16.10 RNase block dilution

	Per reaction (µL)		For 8 reactions (µL)	
Capture size	RNase	Water	RNase	Water
<3.0 Mb	1	9	8	72
≥3.0 Mb	1	3	8	24

Table 16.11 Combining RNase block dilution with the SureSelect library

	Per reaction (μL)		For 8 reactions (μL)	
Capture size	RNase	Library	RNase	Library
<3.0 Mb	5	2	40	16
≥3.0 Mb	2	5	16	40

Table 16.12 SureSelect adapter block mix

Reagent	Per reaction (μL)
SureSelect indexing block #1	2.5
SureSelect block #2	2.5
Index-specific block	0.6
Total	5.6

An individual SureSelect Block Mix will be prepared for each index used.

9. In strip *B*, prepare the pre-capture indexed libraries for target enrichment:
   (a) Add 3.4 μL of 147–221 ng/μL of indexed library to each tube.
   (b) Add 5.6 μL of the corresponding SureSelect Block Mix (containing ISHB).
   (c) Mix thoroughly by pipetting up and down.
   (d) Cap the tubes and place in the thermal cycler.
   (e) Heat the library/adapter block mix to 95 °C for 5 min; then hold at 65 °C. Use a heated lid set at 105 °C.
10. Equilibrate hybridization buffer-containing strip *A* to 65 °C for 5 min before proceeding to the next step.
11. Place the strip into the thermal cycler and incubate at 65 °C for 2 min.
12. Maintain all three strips at 65 °C while transferring 13 μL of the hybridization buffer from each tube of strip *A* to the SureSelect capture library tubes in strip *C*.
13. Transfer 9 μL of indexed library/adapter block mix from each tube of strip *B* to the corresponding tube of strip *C*.
14. Seal strip *C*, now containing 29 μL of solution, tightly with a new strip cap.
15. Incubate the hybridization strip at 65 °C for 24 h.

Step 10: Bead Capture and Post-hybridization Washes

Prior to beginning the post-hybridization washes, prepare a clean, RNase-free workspace. Examine the volume of each hybridization reaction. If greater than 5 μL has evaporated, repeat the hybridization. Prewarm SureSelect wash buffer 2–65 °C.

1. Bead preparation:
   (a) Resuspend the Dynabeads by vortexing.
   (b) To each of 8 clean 1.5 mL LoBind tubes, add 50 μL Dynabeads.
   (c) To each tube of beads:
   - Add 200 μL of SureSelect binding buffer, and vortex for 5 s.
   - Place tubes on a magnetic stand until the solution clears (1–2 min).
   - Aspirate and discard the solution. Do not disturb beads.
   - Repeat for a total of three washes.

   (d) Resuspend beads in 200 μL of SureSelect binding buffer per hybridization.
2. Capture hybrid library:
   (a) Add the contents of each hybridization reaction to a tube of prepared beads.
   (b) Mix beads and hybrid library by pipetting up and down slowly ten times.
   (c) Place each tube on a tube rotator and mix for 30 min at room temperature. Make sure that the solution is being adequately mixed.
   (d) Centrifuge briefly and place tubes on magnetic stand. After the solution has cleared (1–2 min), aspirate the solution and discard.
   (e) Resuspend the beads in 500 μL of SureSelect buffer #1 by briefly vortexing for 5 s.
   (f) Centrifuge briefly and place tubes on magnetic separator. After the solution has cleared (1–2 min), aspirate the solution and discard.
3. Stringency washes:
   (a) Resuspend beads in 500 μL of prewarmed SureSelect wash buffer #2.
   (b) Vortex briefly for 5 s.
   (c) Incubate tubes at 65 °C for 10 min, mixing periodically.
   (d) Centrifuge briefly and place the tubes on the magnetic separator. After the solution has cleared (1–2 min), aspirate the solution and discard.
   (e) Repeat (a)–(d) for a total of three washes.
   (f) Make sure that all wash buffer has been removed.
4. Add 50 μL of SureSelect elution buffer, and vortex for 5 s to resuspend the beads.
5. Incubate tubes for 10 min at room temperature. Mix periodically.
6. Centrifuge briefly and place the tubes on the magnetic separator. After the solution has cleared (1–2 min), aspirate the solution, and transfer it to clean 1.5 mL LoBind tubes.
7. Add 50 μL SureSelect neutralization buffer to the captured DNA. Mix briefly.

Step 11: Post-hybridization Library Cleanup

1. Remove AMPure XP Beads from refrigerator, and gently shake to resuspend the beads. The beads must be at room temperature before use.
2. Resuspend and transfer 180 μL of the beads to each 100 μL tube of captured DNA.
3. Mix, and incubate for 5 min at room temperature.
4. Place the tubes on a magnetic stand for 5 min to separate the beads from the solution.
5. Visually confirm that the beads have moved to the side of the tube and that the solution is clear.
6. Aspirate ~280 μL of clear solution from the tube without disturbing the beads, and discard.
7. Dispense 500 μL of freshly prepared 70% ethanol to each tube, and incubate for 30 s at room temperature. Aspirate ethanol, and repeat for a total of two washes. Let beads dry for 5 min. Under- and overdrying beads will reduce yield.
8. Add 15 μL of molecular biology-grade water, and incubate at room temperature for 2 min.
9. Place the tube on a magnetic stand. After the solution clears (2–3 min), aspirate 15 μL of water, and transfer to a clean 1.5 mL LoBind tube.
10. Repeat Steps 8 and 9 above, adding the second eluent to the first for a total of 40 μL of captured DNA.

Step 12: Amplification of Post-capture Library

The goal of this step is to generate enough material for sequencing. Only the minimum number of cycles to generate sufficient material should be performed. Minimize the number of cycles to ensure maintenance of library complexity.

1. Prepare post-capture amplification mix by combining the reagents listed in Table 16.13.
2. Prepare four PCR reaction tubes for each library. Dispense 40 μL of amplification master mix into each tube; add 10 μL of library to each tube for a total of 50 μL per reaction.

Table 16.13 Post-capture amplification mix

Reagent	Per reaction (×4) (μL)	For 8 reactions (×4) (μL)
PCR-grade water	50	420
TS-PCR-1 (20 μM)	5.0	42
TS-PCR-1 (20 μM)	5.0	42
KAPA Library Amplification Mix	100	840
Total	160	1344 (160 μL/rxn)

Rxn reaction

Table 16.14 Post-capture library amplification thermal cycling protocol

Step	Temperature/duration	Cycles
Activation/denaturation	98 °C, 30 s	1
Amplification	98 °C, 10 s	6
	60 °C, 30 s	
	72 °C, 30 s	
Final extension	72 °C, 5 min	1
Hold	16 °C	

3. Place sealed reactions in a thermal cycler. Amplify the captured DNA using the thermal cycling protocol in Table 16.14.
4. When amplifications are complete, proceed with Step 13 or store at 4 °C for up to 72 h.

Step 13: Purify Amplified Post-capture Libraries

1. Pool the 4 × 50 μL PCR reactions into a single clean 1.5 mL microcentrifuge tube.
2. Gently shake the AMPure XP bottle to resuspend any beads that may have settled. The beads must be at room temperature before use.
3. Transfer 360 μL of beads to each tube containing pooled reactions.
4. Mix well, and incubate for 5 min at room temperature.
5. Place the tubes onto a Dyna-Mag 2 rack for 5 min to separate the beads from solution.
6. Visually confirm that the beads have moved to the side of the tube and the solution is clear.
7. Aspirate 560 μL of clear solution from each tube, and discard.
8. Dispense 300 μL of 70% ethanol to each tube, and incubate for 30 s at room temperature. Aspirate the ethanol, and discard. Repeat for a total of two washes. Dry the beads for 5 min. Under- or overdrying of the beads will reduce yield.
9. Take the tubes off the magnet, add 20 μL of water to each tube, and mix.
10. Place the tubes on the Dyna-Mag 2 rack for 2 min to separate the beads.
11. Transfer 20 μL eluent to a new tube.
12. Add 2 μL of 1% Tween (final 0.1%). The samples are now ready for Q/C.

Step 14: Quality Control Analysis of Captured Libraries

1. Assess the fragment size distribution and concentration by running a sample of each library on Bioanalyzer 2100 (DNA 1000 Chip) as performed in Step 8.

2. Record the average fragment size for each library. The proper size will be important for accurately calculating the library concentrations in Step 15.

Step 15: Library Quantification by Real-Time Quantitative PCR

1. Quantify each library using the KAPA Biosystems Library Quantification Kit according to the manufacturer's specification. Run 1:50,000 and 1:100,000 dilutions for each library in triplicate.
2. Make sure to take into account the difference in size between the library fragment size and the standards included in the kit. Use the average fragment size determined in Step 14. This will have an impact on the final molar concentration.
3. Samples should be adjusted to a final concentration of 10 nM and pooled (if desired).
4. *Equimolar library pooling*. Precise pooling of libraries can be challenging. The reproducibility of the qPCR-based assay described in Step 15 is very high but is very dependent on the pipetting skills of the operator. Make sure to run qPCR reactions in triplicate. One should expect to see variances between libraries (as measured by sequence read output) of less than 0.5×.

Step 16: Illumina Sequencing

1. Sequence libraries according to the manufacturer's instructions.
2. *Minimal read depth for accurate variant calling*. As a general rule of thumb, a minimum of 20–30× coverage over each target base will ensure that both alleles are detected, if the individual is heterozygous at a given nucleotide. Clinical laboratories will typically set the coverage threshold at 100× or greater for targeted gene panels, as it is not difficult (or expensive) to obtain coverages at or above this level.
3. *Sequence run evaluation*. Accuracy of NGS has improved greatly over the past several years. For 2 × 100 bp sequencing run on a HiSeq instrument, greater than 85% of bases should yield Q30 or greater. For clinical work, it may be advisable to sacrifice read length in favor of accuracy. Some laboratories still perform 2 × 50 bp or 2 × 75 bp—as the first portion of the read is the most accurate on a per nucleotide basis.
4. *Library complexity*. If your library does not possess adequate complexity, your results may not reveal the full complement of expected variants, as the template has undergone a bottleneck during the procedure. As you become more familiar with the process, you can add various Q/C cutoffs or metrics to reduce the likelihood of low-complexity data. Low complexity will also lower the sensitivity of the assay.

Data Analysis

A wide variety of software tools are available for analysis of next-generation sequence data. Many laboratories use the software pipeline provided by Illumina or other commonly used tools, such as the Burrows–Wheeler Aligner [13] and Genome Analysis Tool Kit (GATK) [14, 15]. Each laboratory will need to evaluate analytical tools and determine how to set software parameters needed for their specific application(s). Most software tools are optimized for the detection of germline variants. If users need to identify somatic mutations found in tumors, they will need to optimize the specific parameters of each application to ensure robust detection. Specific software applications have become available, such as Mutect [16], SomaticSniper [17], or SNVmix [18, 19], just for this purpose. One would need to test their analysis pipeline empirically with a rigorous validation protocol, before applying their use to clinical diagnostic work.

In some cases, it may be useful to verify alignments and variant detection visually. The Integrative Genomics Viewer [20] can be used to examine a wide array of sequence data, including read alignments, variants, and coverage data.

Conclusion

NGS gene panels provide comprehensive, rapid, and cost-effective technology for clinical genetic testing. Single-gene testing only identifies the causative variant in 10–20% of clinically diagnosed complex genetic diseases. By combining all of the known genes for a given phenotype, testing can be performed in a more effective manner, thereby reducing the time required to make a molecular diagnosis. Currently, the optimal platform for development and implementation of NGS gene panels is an in-solution, hybridization-based enrichment system as described in the example provided in this chapter. These assays are automatable, reproducible, and highly sensitive for the detection of single-nucleotide variants, small insertions/deletions, and copy number variants. As the cost of exome sequencing declines, more labs will begin using the exome as a backbone for multigene testing, in particular for large panels, such as those focused on movement disorders, neuromuscular disorders, and intellectual disability. NGS gene panels will likely continue to be the preferred testing modality for many applications for the foreseeable future, whether through the development of separate assays or exome-based protocols.

References

1. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010;11:31–46.
2. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet. 2013;14:295–300. https://doi.org/10.1038/nrg3463.
3. Teekakirikul P, Kelly MA, Rehm HL, Lakdawala NK, Funke BH. Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era. J Mol Diagn. 2013;15(2):158–70.
4. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, Gulden C, Kupfer SS, Swisher EM, Bennett RL, Novetsky AP, Jarvik GP, Olopade OI, Goodfellow PJ, King MC, Tait JF, Walsh T. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn. 2012;14(4):357–66.
5. Gecz J, Shoubridge C, Corbett M. The genetic landscape of intellectual disability arising from chromosome X. Trends Genet. 2009;25(7):308–16.
6. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, Thiruvahindrapuram B, Merico D, Jobling R, Nalpathamkalam T, Pellecchia G, Sung WWL, Wang Z, Bikangaga P, Boelman C, Carter MT, Cordeiro D, Cytrynbaum C, Dell SD, Dhir P, Dowling JJ, Heon E, Hewson S, Hiraki L, Inbar-Feigenberg M, Klatt R, Kronick J, Laxer RM, Licht C, MacDonald H, Mercimek-Andrews S, Mendoza-Londono R, Piscione T, Schneider R, Schulze A, Silverman E, Siriwardena K, Snead OC, Sondheimer N, Sutherland J, Vincent A, Wasserman JD, Weksberg R, Shuman C, Carew C, Szego MJ, Hayeems RZ, Basran R, Stavropoulos DJ, Ray PN, Bowdin S, Meyn MS, Cohn RD, Scherer SW, Marshall CR. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med. 2018 Apr;20(4):435–43. https://doi.org/10.1038/gim.2017.119.
7. ACMG Board of Directors. Points to consider in the clinical application of genomic sequencing. Genet Med. 2012;14:759–61.
8. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7(2):111–8.
9. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182–9.
10. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. 2011;12(1):R1.
11. Shearer AE, Hildebrand MS, Smith RJ. Solution-based targeted genomic enrichment for precious DNA samples. BMC Biotechnol. 2012;12:20.
12. Nord AS, Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. BMC Genomics. 2011;12:184.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
15. DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
16. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213–9.
17. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012;28(3):311–7.
18. Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman D, Murphy KP, Aparicio S, Shah SP. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics. 2010;26(6):730–6.
19. Roth A, Ding J, Morin R, Crisan A, Ha G, Giuliany R, Bashashati A, Hirst M, Turashvili G, Oloumi A, Marra MA, Aparicio S, Shah SP. JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. Bioinformatics. 2012;28(7):907–13.
20. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2012;14:178–92. https://doi.org/10.1093/bib/bbs017.

Implementation of Exome Sequencing Assay

17

Samya Chakravorty, Arunkanth Ankala, and Madhuri R. Hegde

Introduction

The paradigm shift from genetics to genomics was put in motion by a revolutionary study that described sequencing of the entire genome of *Mycoplasma genitalium* in a single run on a Roche 454 instrument [1]. The study revealed a highly parallel sequencing-by-synthesis method capable of sequencing 25 million bases, at 99% or more accuracy, in a single 4 h run. Subsequently, several high-throughput flow cell-based sequencing methods became commercially available from Illumina (San Diego, California), Roche (454 Life Sciences Corporation, Branford, Connecticut), and Life Technologies (Carlsbad, California). These developments marked the beginning of a new era based on next-generation sequencing (NGS). Simultaneously, several sequence capture or target enrichment methods were evolving to improve the throughput and specificity of sequencing technology [2–7]. With the rapid development of these advanced sequencing technologies, per-base sequencing costs are declining drastically, to a level at which almost complete resequencing of the human genome is becoming affordable, even in clinical settings [7–10]. Nevertheless, the infrastructure requirements, analysis burden, and turnaround time requirements involved in clinically interpreting entire patient genomes for mutation detection bear significant issues. Whole-exome sequencing (WES), in contrast, which interrogates the roughly 1% of the human genome that represents the entire coding region and harbors 85% or more of causative mutations, is quite feasible and much more affordable in a clinical setting.

The successful implementation of NGS technology in clinical laboratories for diagnostic purposes began with gene panels designed to specifically target and sequence multiple genes related to a particular disorder. Soon several disease-specific or phenotype-specific gene panels became clinically available [11–18]. These included highly heterogeneous disorders, such as congenital disorders of glycosylation (CDGs), congenital muscular dystrophies (CMDs), limb girdle muscular dystrophies (LGMDs), dilated cardiomyopathy, and mitochondrial disorders, each with several subtypes of overlapping phenotypes and associated with a large number of causative genes [16, 19]. Traditional molecular diagnostic approaches for such diseases followed a sequential, Sanger sequencing-based gene-by-gene analysis of known disease-associated genes. However, with the advent of NGS technologies and the decline in per-base sequencing cost, the NGS panel approach has become a significantly cheaper and quicker option, available as a single test. Subsequently, with the availability of better sequence chemistries and easier workflows, NGS technology moved into other clinical arenas, including cancer diagnosis [20], human leukocyte antigen locus characterization [21, 22], and pathogen genome sequencing for the purpose of evaluating resistance [23]. Rapid identification of novel disease genes and the revealing of locus and allelic heterogeneity of inherited genetic disorders, both Mendelian and complex, have established WES as a comprehensive clinical test.

In this chapter, we discuss the various roles of WES in clinical medicine and provide an overview of how WES has transformed the diagnostic outlook on genetic disorders. We highlight the major successes and challenges of implementing WES assays in clinical genetics, concluding with a note on the future of whole-exome assays.

Whole-Exome Sequencing: Methodology

Exome Capture and Next-Generation Sequencing

WES refers to sequencing of the entire protein-coding region of the human genome. This is achieved by parallel sequencing of all targeted regions (exons) using NGS technologies.

S. Chakravorty · A. Ankala · M. R. Hegde (✉)
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA
e-mail: mhegde@emory.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_17

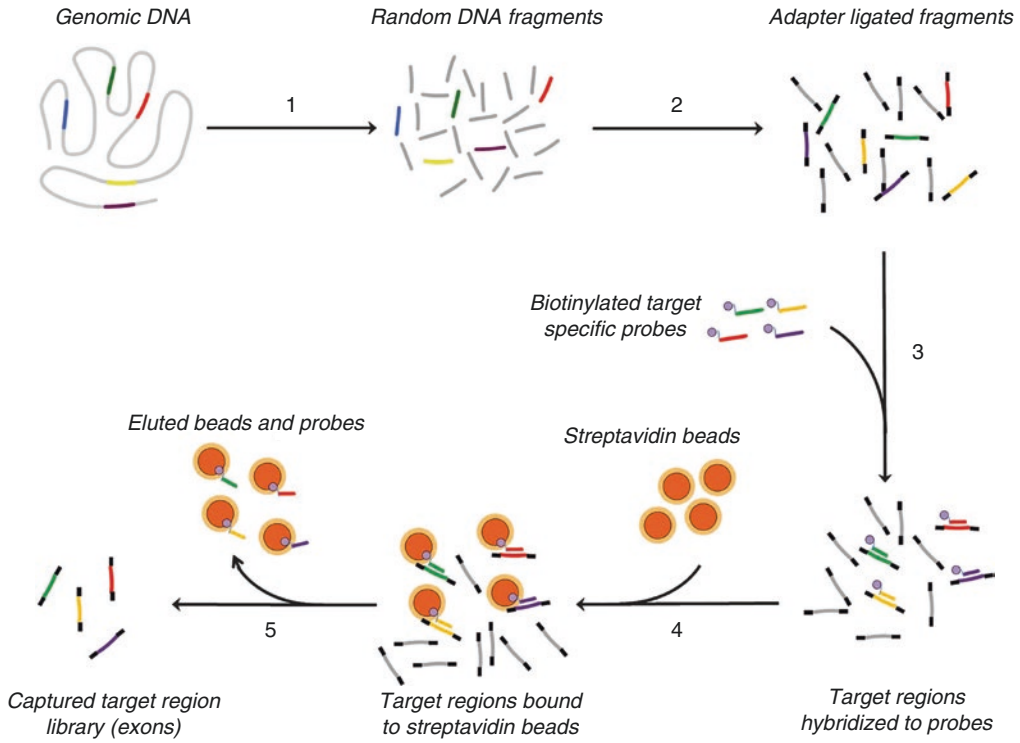


Fig. 17.1 Basic methodology of exome capture or target enrichment for whole-exome sequencing. The various steps involved are indicated by numbers. In step 1, genomic DNA is randomly fragmented into more or less uniform shorter segments, either by ultrasonication or restriction digestion with enzymes. In step 2, adapters with sequencing motifs and indices are ligated to the fragments. In step 3, biotinylated probes that are specific for target regions (exons) are added and allowed to hybridize. Step 4 involves addition of streptavidin beads to selectively capture all target regions by binding biotin. While the streptavidin beads (with bound target regions) are held by a magnet, unbound nonspecific DNA fragments are separated and washed away. Finally, in step 5, target regions are eluted by denaturation from the biotinylated probes. Although alternative methods for adapter ligation may be available, the basic concept for target (exon) capture is similar

Irrespective of the manufacturer and sequencing platform, the basic methodology or principles involved in WES are similar (Fig. 17.1). First, genomic DNA is fragmented either by optimized sonication or by restriction digestion to generate uniform libraries of DNA strands. This fragmented DNA is then enriched for protein-coding regions of the genome (exons), using unique adapter ligation chemistry that is proprietary to each individual commercial manufacturer [24]. Adapter-ligated DNA fragments are captured and amplified either on a solid surface (bridge amplification on a glass slide) or in solution (emulsion PCR on micro-beads). Finally, different massively parallel sequencing technologies are used to sequence all target DNA regions and produce what are called sequence reads, of different lengths, depending on the technology used. Sequence reads are computationally aligned to a reference exome and analyzed for sequence variations. The experimental design allows for each nucleotide to be represented in a large number of reads, which is referred to as "read depth" or "coverage." Variant annotation using analytical pipelines helps filter false-positive and noncontributive calls to identify causal mutations. WES therefore serves as a comprehensive method for rapid identification of exonic mutations, such as missense, nonsense, splice site, and small deletion and insertion mutations (indels); however, detection of copy number variations (CNVs) and structural variations (SVs) is still an issue.

Sequence Analysis and Variant Detection

Massively parallel sequencing of the entire exome generates terabytes of information. Sorting through and making sense of such massive volumes of data to identify causative genes and mutations require multistep bioinformatics analysis. Upon initial generation of sequence base call files, they are converted into the more commonly used FASTQ file format for storage and later analysis [24]. Several open-source and in-house-developed software programs can be used to align sequence reads to a best-match location of a reference sequence and stored in what is called the BAM (binary alignment) file format [25]. These aligned reads are then processed to call out sequence variants depending on the presence and zygosity of variants. Information from this analysis, which includes inferred single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) along with base coverage, quality, and score, is stored in a different

file format termed variant call format (VCF) [26, 27]. Finally, each single call in the entire set of variants is annotated with a variety of customizable information, including gene name, genomic and cDNA coordinates, amino acid change, and functional classification, to help with the interpretation of causative variants [28].

Variant Analysis and Molecular Diagnosis

Analysis of the variants and identification of the disease causative gene and mutations in WES are daunting tasks compared to the traditional single-gene sequencing approach. To this end, the American College of Medical Genetics and Genomics (ACMG) has released guidelines [29] to help clinical diagnostic laboratories with interpretation of NGS data and classification of observed variants. The clinical significance of each variant is determined based on the available evidence and can be classified as either benign, likely benign, pathogenic, likely pathogenic or as a variant of uncertain clinical significance (VUS). However, when gathering such evidence from available variant databases, caution needs to be exercised as the variant classification or interpretation in these databases may be outdated or incorrect [30]. During this process, variants in genes that are not relevant to the patient's clinical phenotypes are filtered out. The caveat to such filtering is potential incidental findings, which are reported based on ACMG guidelines [31]. In the filtering process, variants in genes that are not relevant to the patient's clinical phenotypes are filtered in the clinical setting using population-wide minor allele frequency, predicted effect on protein function or splicing, literature evidence, and disease-variant databases such as the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk), ClinVar (http://www.clinvar.com), Online Mendelian Inheritance of Man (https://www.ncbi.nlm.nih.gov/omim), 1000 Genome, ExAC, genome Aggregation Database (gnomAD), genome-wide association study (GWAS) catalog, GWAS Central, Human Genome Variation Society (HGVS), the locus-specific databases of NHLBI Exome Variant Server (EVS), and Pharmacogenomics Knowledgebase (PharmGKB). Nevertheless, caution should be taken while filtering as pathogenic variants occurring at a high frequency in certain populations can be mistakenly filtered out. Frequency cutoffs to determine variant pathogenicity can be arbitrary; however, recently a more robust statistical framework for determining likely pathogenicity of variant based on population frequency using the Exome Aggregation Consortium (ExAC) dataset was reported [32]. This framework of filtering candidate variants was based on disease prevalence, genetic and allelic heterogeneity, inheritance patterns, penetrance, and variation due to sampling in different datasets that resulted in a low (<0.001) false-positive rate. In addition, new algorithms are under development to assist in the variant-disease phenotype relationship interpretation, including those which compare cross-species phenotypic similarity in order to prioritize genes in the exome data for a given disease [33].

Several predictive algorithms are being developed and made commercially available, but their reliability and interpretative ability are not well established. Repetitive regions such as trinucleotide repeats and genes with pseudo sequences remain problematic to detect or interpret. Novel algorithms and databases are in development to better identify and annotate these sequences and their variations [34–38]. Moreover, there lies marked difference in the predictive algorithms used for exome variant interpretation in different diagnostic clinics. To overcome that, standardization and sharing of the diagnostic pipeline and data are being proposed [39–41] such as the InterVar (http://wintervar.wglab.org/) [42]. Importantly, it was found recently that some of the newer algorithms show higher and more robust predictive power by integrating disease mechanism, gene constraint, and inheritance mode than the older ones in current clinical use [43, 44], which are REVEL [45], VEST3 [46], MetaLR and MetaSVM [47], and hEAt [48]. This implies the need for revising the guidelines for usage of bioinformatics algorithms in the clinic for exome variant inference. Evaluating the limitations of each algorithm and making such algorithms and the data widely available to the scientific community through open source, such as the newly developed Personal Cancer Genome Reporter (PCGR) (https://github.com/sigven/pcgr) [49], are critical to achieving a more uniform clinical interpretation of exome sequencing data.

Most of the clinical laboratories that offer WES assays currently include various parameters, such as the functional effect of the observed variant, relevance of the gene to the clinical presentation, and mode of inheritance, to filter variant calls through in-house-validated pipelines and algorithms (Fig. 17.2). Clinical correlation analysis is important to perform on the filtered variants to further short-list the candidate variants based on the initial clinical diagnosis or clinical data [50]. Finally, short-listed candidate variants are confirmed by the gold standard Sanger sequencing. Confirmed variants may fall into different categories based on previous association and functional effects of the variant (Table 17.1). In the event a new disease gene is identified, disease association requires further evidence. In silico analysis by prediction algorithms based on evolutionary conservation of the amino acid or nucleotide may increase confidence in an association, but is not definitive [24]. Segregation of mutations in the gene with presence of disease among family members may also provide additional evidence, but does not necessarily or fully associate the gene with disease. Functional studies are best, when available, because they may not only establish disease association but also provide insight into disease

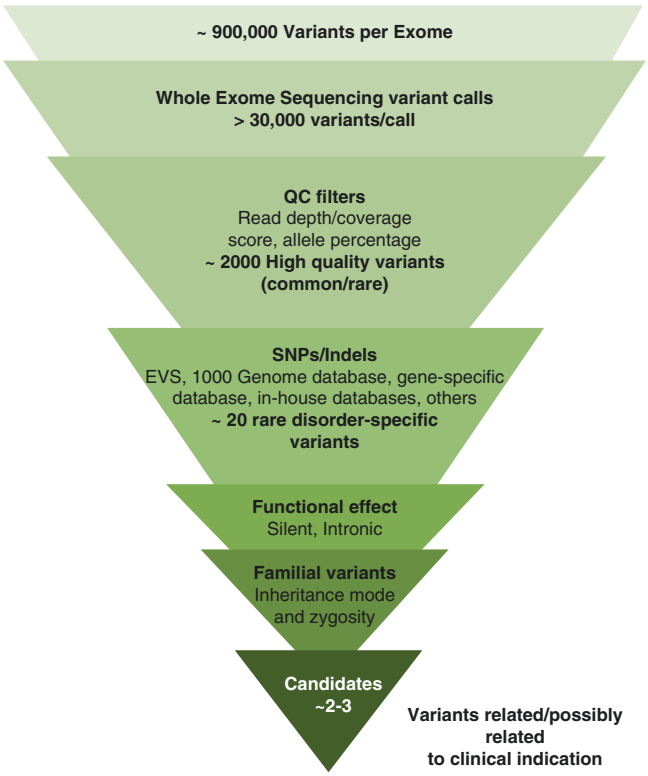


Fig. 17.2 Basic pipeline for variant filtration in whole-exome sequencing analysis. Various parameters are included in WES algorithms to filter and remove nonpathogenic, false-positive, and false-negative variants from whole-exome variant data to create a manageable dataset (from >30,000 variants per variant call to about 2000 high-quality variants and then filtered further using databases to150–250 variants or sometimes about 20 variants depending on the type of database filtering) that includes the candidate causative mutations. As indicated in the data filtration funnel, variants that do not meet QC metrics, such as those with poor coverage (<20×), are considered less likely to be real, treated as false positives, and therefore filtered. Variants with a minor allele frequency of >0.01 are polymorphisms by definition and less likely to be pathogenic. Silent changes and intronic variants beyond the consensus splice donor/acceptor sequences are less likely to be pathogenic and are often filtered in initial rounds of analysis. Familial variants may also be carefully filtered based on zygosity and segregation pattern. Though the basic parameters followed are common to all commercial and laboratory-developed algorithms, the thresholds and ranges for acceptability may vary. The narrowed down variants after all the filtering process is then analyzed in conjunction with specific individual's initial clinical diagnosis or other clinical information, which leads to about 2–3 candidate variants that are related or possibly related to the individual's clinical indication. EVS = Exome Variant Server (NHLBI Exome Sequencing Project)

pathogenesis and treatment options, as per ACMG guidelines [29]. Alternatively, identification of mutations in the novel gene in unrelated individuals with similar phenotypes by rapid targeted single-gene testing may establish disease association, as well. While diagnostic laboratories focus on finding a pathogenic change in a known disease-associated gene, research testing of exomes is driven by the additional goal of new gene discovery and may include extensive functional analysis to establish disease association with the gene.

Table 17.1 Predictive value and significance of confirmed whole-exome sequencing variants

Variant category	Clinical diagnostic value	Functional value	Further action
Previously reported mutation	Establish diagnosis and disease subtype; recruitment to disease registries or clinical trials	Understand clinical spectrum	–
Novel mutation in known disease-associated gene	Establish familial mutation; correlation of inheritance pattern and disease association in family	Expand allelic and phenotypic heterogeneity and variant spectrum	Functional studies, characterize and classify the novel variant
Mutation in known disease-unrelated gene	Establish disease diagnosis	Expand locus heterogeneity of disease	Functional studies, characterize disease subtype
Potential pathogenic variant/mutation in previously unknown gene	Not predictable or actionable	Hypothesize new disease/gene, gene discovery	Characterize disease type, functional studies

Exome Sequencing: A Transformative Technology

NGS approaches, and especially WES, have created hope for patients who may have already undergone a diagnostic odyssey of invasive approaches and clinical tests and yet remain in the dark as to the underlying genetic cause of their condition. The potential of WES to provide molecular diagnoses by screening nearly all human exons for mutations was recognized early on, and attempts to explore its diagnostic potential were soon underway, heralding a new era in clinical and medical genetics.

WES as a Diagnostic Assay: Proven Potential

WES has facilitated characterization of several recessive as well as dominant diseases, revealing associations with new disease genes. Recessive traits, which are more commonly highlighted in consanguineous families, are comparatively easier to diagnose and implicate through WES because affected individuals within the family carry causative mutations in segments that are homozygous by descent. For example, in the case of first cousin mating, these regions account for approximately 10% of the entire exome, thereby restricting the search to this small region. For dominant

traits, however, the process is less straightforward. Molecular characterization of dominant traits is complicated by several factors, including reduced penetrance for certain genes, locus heterogeneity, and alleles that affect reproductive fitness. In such scenarios, the finding of independent de novo variants in the same gene among multiple unrelated affected individuals provides considerable evidence for disease association irrespective of allelic heterogeneity. The first successful demonstration of the potency of WES for rare variant identification and disease diagnosis came from an unexpected diagnosis of a patient referred for possible Bartter syndrome [51]. Due to an inconclusive clinical presentation, WES was performed for this individual, and informed variant analysis led to the identification of a homozygous mutation in the *SLC26A3* gene. This study provided the first proof of concept of the application of WES for genetic disease diagnosis. Even though the gene was previously known to be disease causing (congenital chloride-losing diarrhea, CLD), the clinical overlap of the patient's phenotype with that of Bartter syndrome [52] obviated suspicion of the gene. Substantial family information, including that of consanguinity, inheritance mode of the disease, and regions of excessive homozygosity due to identity by descent, helped with the molecular characterization. Moreover, reevaluation of additional study subjects with a presumptive diagnosis of Bartter syndrome identified mutations in *SLC26A3*. These findings not only established the diagnostic ability of WES but also expanded the phenotypic variability of *SLC26A3*-associated CLD.

Whole-Exome Sequencing Facilitates Gene Discovery

Traditional gene mapping tools, such as homozygosity mapping, linkage analysis, karyotyping, and copy number variation (CNV) analysis, have led to the identification of new disease genes [53–56]; however, these methods require analysis of a cohort of multiple unrelated affected individuals to narrow down genomic regions of interest, before finally zeroing in on the candidate gene. In contrast, WES of a single family or a parent–proband trio can result in rapid gene identification. This was first reported approximately 5 years after the launch of the technology in 2005 [57]. Using WES, two potentially pathogenic variants were identified in a novel candidate gene, *DHODH*, thus implicating the gene in the autosomal recessive Miller syndrome. This condition is characterized by severe micrognathia, cleft lip or palate, limb defects, coloboma, and supernumerary nipples [58]. Even though the disease had been described several decades ago, not much about the causal gene or mode of inheritance was known until this study. Despite little understanding of how *DHODH* mutations cause Miller syndrome, the subsequent identification of mutations in additional patients by targeted gene sequencing confirmed disease association without functional analysis. Shortly thereafter, another novel disease gene association was reported by the same group, which identified *MLL2* (*KMT2D*) to be the causative gene for Kabuki syndrome [59]. These findings strongly suggested that exome sequencing of a small number of affected individuals from unrelated kindred, or of multiple individuals from a single affected family, could be a powerful and efficient strategy for the identification of rare disease genes.

From Medical Genetics to Medical Genomics: A Shift in Paradigm

Beginning in early 2008, the NIH's Undiagnosed Diseases Program (UDP) began offering clinical WES as a pilot program, with initial funds totaling $280,000 [60]. UDP's explicit objectives were to provide molecular diagnosis to patients who remained undiagnosed despite thorough workup and to discover novel disease genes and disorders to gain insight into the pathogenesis of the clinical manifestations. After receiving several thousand applications from prospective participants, 160 individuals were enrolled, and the huge task of deciphering the underlying genetic causes began. Included was a healthy Colombian couple with two sons affected with an uncharacterized neurological illness, presenting with seizures, tremors, and several other complications. When one of the sons succumbed to the disease, the second son of the family was enrolled in the abovementioned multi-institute initiative in hopes of identifying the underlying cause. After collaborative efforts for more than a year, a definitive diagnosis came from WES analysis. Furthermore, the molecular diagnosis was also established for almost 25% (39/160) of the enrolled individuals overall. Novel disease genes, including *NT5E*, associated with arterial calcification disorder [61], and *HINT3*, an aprataxin-related gene causative of a familial distal myopathy [62], were identified as well. Most of the diagnoses made, however, included known rare (≤1 in 10,000) or ultra-rare (<60 cases reported) diseases in individuals who had previously undergone multiple molecular and/or biochemical genetic tests. UDP's experience suggested that, with comprehensive phenotypic information, accurate bioinformatics tools, and a methodological approach, WES can be an economical single test for disease diagnosis.

Implementation of Exome Sequencing in Clinical Medicine

Whereas the suitability of WES for clinical medicine was initially debated, the emerging consensus is that the future of diagnostic exome sequencing has already begun [63, 64]. As new genes and diseases are identified through clinical WES,

the test is gaining popularity. Expected reductions in cost and improved reimbursement are also likely to lead to wider implementation of WES in clinical medicine.

Mendelian Disorders and Exome Sequencing

The conventional approach, still widely in practice, for molecular diagnosis of single-gene Mendelian disorders follows serial interrogation of all exons and exon–intron boundaries of known disease-specific genes via traditional polymerase chain reaction (PCR) amplification and the gold standard Sanger sequencing. Unlike complex traits and disorders such as autism and intellectual disability, which can involve several causative genes and variants, Mendelian disorders are generally associated with mutations in a single gene. With the utilization of clinical genetics and molecular diagnosis, however, locus heterogeneity and overlapping disease phenotypes have shown that, even for Mendelian disorders, making a molecular diagnosis is less straightforward than previously thought. This notion favored the application of multigene panels in which all common disease-related genes are interrogated simultaneously through NGS. Consequently, there are now several individual disease gene panels available [12, 16, 19]. Even though the panel approach has reduced the diagnostic odyssey for patients and boosted diagnostic capacity, a substantial fraction of patients still remain without a molecular diagnosis. This can be attributed, in part, to the inability to detect mutations in regulatory and intronic regions. Nevertheless, most such cases are believed to be due to the involvement of previously unknown disease genes. One important feature in support of this is the occurrence of more than 85% of causative mutations for Mendelian disorders in exonic regions of the genome [65]. This percentage, together with the growing potential of WES as a diagnostic tool, makes it a preferred approach for rare Mendelian disorders with genetic and phenotypic heterogeneity. Notably, however, causative variants detectable by a combination of conventional methodologies, including homozygosity mapping and candidate gene selection, may be missed by WES [66, 67]. Bloch-Zupan et al. [66] report a case of homozygous mutations in the *SMOC2* gene, responsible for dental developmental defects, which were initially missed by WES due to poor coverage [66]. Overall, however, whereas homozygosity mapping or linkage analysis may be preferred for consanguineous and large pedigrees, WES is proving to be the most informative of these diagnostic tests [68–71]. In some cases, WES has provided an accurate molecular diagnosis in patients previously diagnosed with a different disease, further cementing the value of this assay in clinically heterogeneous Mendelian disorders [72]. Besides establishing a molecular diagnosis in patients and providing carrier testing opportunities for family members, the identification of causative mutations in Mendelian diseases also guides patient management and family counseling [73] and opens up opportunities for therapeutic intervention and participation in clinical studies [74]. Finally, the identification of new disease genes and causative mutations contributes to our understanding of disease phenotype, pathogenesis, and gene function [75].

Complex Disorders and Exome Sequencing

Common complex diseases constitute a major part of overall disease burden in the general population. Most common diseases are complex, with extensive genetic heterogeneity resulting in clinically indistinguishable phenotypes. This includes conditions such as autism, intellectual disability, cardiac disease, and diabetes. X-chromosome-linked intellectual disability alone has been associated with more than 100 different genes. Similarly, autism spectrum disorders are linked to multiple genes, with no single gene accounting for more than 1% of cases [76]. It is obvious that, even more so than for single-gene Mendelian disorders, the WES approach is advantageous for multifactorial and multigenic complex disease characterization. Recently, one single WES study investigating the genetic etiology of autosomal recessive forms of intellectual disability identified 50 novel candidate genes [77]. These include genes encoding proteins involved in transcription, translation, cell cycle control, and fatty acid and energy metabolism critical for normal brain development and function. The discovery of such novel disease-associated genes not only improves our understanding of the underlying cause of disease manifestations but can also suggest novel targets for therapy and management.

Unlike most Mendelian disorders, diseases with complex genetic etiologies involve coding variants that present as risk factors rather than direct causes of disease. Such risk factors found by traditional methods to date include an *APOE* genotype that plays a role in late-onset Alzheimer's disease, complement factor H polymorphism in age-related macular degeneration, and an *LRRK2* risk variant in Parkinson's disease [78–80]. The application of WES to complex disease diagnosis will enable the identification of similar common protein-coding risk alleles, as well as rare risk alleles. Genome-wide association studies (GWAS) have been revolutionary in terms of uncovering common variants associated with complex disorders, but have not satisfactorily explained the heritability of these traits [81–84]. With the advent of WES, the focus of complex trait genetics has shifted toward low-frequency and rare variants [85, 86], and the link between variants and complex traits is on its way to becoming clearer [87–91]. The routine use of WES in clinical laboratories will most likely identify more and more rare variants that have a strong causative effect on phenotype, unlike the common variants that, individually, contribute only minimally [92, 93].

Application of WES to Neoplastic Diseases

Historically, pathologists have relied on histomorphology to classify and diagnose neoplasms [94, 95]. Recent progress in cancer genomics, however, has pointed toward the utility of a more granular approach through the identification of genetic alterations common to morphologically diverse tumor types and through the discrimination of subgroups within what was thought to be a single tumor type [96]. Consequently, WES has been applied to tumor diagnostics to obtain a comprehensive picture of copy number alterations (CNAs) and of pathogenic mutations [97]. The potential of WES to detect somatic CNAs in cancer syndromes has been explored, as well [97, 98]. In a study involving 17 matched tumor and normal tissues from patients with metastatic castrate-resistant prostate cancer, targeted WES analysis successfully identified various common CNAs, such as androgen receptor (*AR*) gain and *PTEN* loss [97]. This study and others suggest that somatic CNAs that involve the amplification of oncogenes or deletion of tumor suppressors and are significant contributors to cancer etiology can now be monitored more comprehensively using WES than array-based technologies [99]. Unlike germ line mutations, somatic mutation and CNA detection in cancer are performed by simultaneous exome sequencing of normal and tumor tissue from the same individual, followed by a comparison of copy number ratios of exonic regions in the two sample types [97]. This approach of analyzing the relative coverage (of tumor versus normal sample) distinguishes a true chromosomal deletion from a lack of coverage due to technical limitations. WES thus offers the combined efficiency of both array comparative genomic hybridization (aCGH), which detects CNAs by relative probe frequency [100], and single nucleotide polymorphism (SNP) array, which detects loss of heterozygosity (LOH) and absence of heterozygosity (AOH) by zygosity changes at known SNP loci [101]. Whereas the prohibitive cost and analysis burden of whole-genome sequencing (WGS) have limited its clinical application thus far, successful detection of somatic *DNMT3A* mutations in acute monocytic leukemia [102], *PBRM1* mutations in renal carcinoma [103], *BAP1* mutations in metastasizing uveal melanomas [34], and *AR*, *NCOA2*, *PTEN*, *RB1*, and *TP53* CNAs in prostate cancer [104] by WES is confirming it as a cancer diagnostic and monitoring assay option.

There are several advantages to using WES for cancer genomics. First, it provides an exon-level resolution of CNAs. Second, the vast data available through comprehensive sequencing projects such as The Cancer Genome Atlas (TCGA) can be leveraged because whole-exome data for thousands of cancer cases from multiple studies are publicly available [105]. This makes integrative cancer detection strategies possible and drives personalized medicine approaches. Genotype-directed therapies are transforming cancer care, as seen with several drugs and target inhibitors in various cancer types, including chronic myeloid leukemia, colorectal adenocarcinoma, and melanoma [106–108]. The role of coexisting or co-occurring passenger mutations, separate from the driver mutations that actually cause the clonal expansion of cancer cells, is also being investigated so the two can be distinguished [109, 110]. Comparison of WES data across multiple patients is expected to contribute to the teasing out of the two, which could in turn translate into new drug targets. Despite these advantages, WES still has some limitations. These are primarily pertaining to coverage of certain exons and of genes with complex sequence context; as a result of which, some mutations and CNAs may be missed. Additionally, CNAs involving gene-poor regions may not be detected due to assay design. Gene fusion events or chimeric gene products unique to cancer etiology and the more frequent large chromosomal aberration events, such as translocations, large deletions, or inversions, are not detected by WES. A comprehensive approach of various NGS technologies including WES, WGS, and transcriptome analysis is being explored, but clinical applicability is still rudimentary [105, 111–113].

From Diagnosis to Therapy: Advances in Clinical Care

Despite the proven potential of WES for clinical diagnostic purposes, one common criticism of the technology is the lack of evidence for its clinical usefulness. Pharmacogenomics is one area in which WES is expected to play a major role, especially by identifying variants that contribute to genotype-specific responses to drugs. One such example is related to the substitution of glutamic acid for valine at position 600 (p.V600E) in the *BRAF* gene in individuals with malignant melanoma [114, 115]. This specific mutation acts by conferring a constant flux through the mitogen-activated protein kinase (MAPK) pathway, thereby promoting malignancy. The genotype-specific drug vemurafenib (PLX4032) is used for targeted intervention of metastatic melanoma [116, 117]. Eventually, however, tumor cells were found to develop resistance to the drug over time, but in a cohort of 20 melanoma patients treated with vemurafenib, WES identified the underlying cause for the development of drug resistance: a gain in copy number (by 2–13 times) of the mutant p.V600E *BRAF* allele [118].

Several other targeted therapies, such as imatinib for chronic myeloid leukemia, trastuzumab for breast cancer, irinotecan and panitumumab for colorectal cancer, and erlotinib for lung cancer, may all be monitored for their treatment effect and resistance development using WES. Implementation of WES in the context of personalized medicine is highlighted by a recent study reporting a novel genetic

risk factor linked to the VACTERL association [119]. A heterozygous mutation in the *CPSI* gene, identified by WES in monozygotic twins, is suspected of being the risk factor associated with the severe pulmonary artery hypertension observed post-surgery in the twin who underwent surgery. Generally, homozygous or compound heterozygous mutations in *CPSI* are associated with a rare urea cycle disorder; however, through WES analysis, the authors clarified that there were no discordant de novo mutations between the two twins and that the observed complication must have been due to the combination of the observed heterozygous variant and an environmental trigger: in this case, surgery.

Limitations and Challenges of Implementing Exome Sequencing Assays

Despite being quite comprehensive, WES has yet to overcome several technical and analytic challenges before it can replace the current gold standard of Sanger sequencing or even targeted NGS panels. These challenges are summarized here. The first and foremost technical challenge is the inefficiency to capture and sequence all target exons. Contrary to what is suggested by its name, WES currently misses around 5–8% of the human exome because of low or no coverage [51]. Most of this is explained by sequence context, such as with high or low GC content or the presence of highly homologous pseudogenes [120]. Capture of all target exons is, of course, essential to avoid false-negative interpretations due to the presence of potentially causative mutations in missed exons. Highly repetitive sequences, which include interspersed repeats and tandem repeats, constitute more than half of the human genome. These highly homologous regions are co-enriched and co-sequenced along with the target regions [121]. This challenge may be countered by increasing the sequence read size, which is still limited with current NGS technologies. However, several alternative approaches, such as paired end sequencing and correlation of average read depth differences to detect repeat regions, are being explored [121]. A second challenge is storage and management of the vast amount of sequencing data generated by the technology. This demands a large investment in infrastructure and technology, which is a major strain for diagnostic laboratories. A third limitation is the variant detection capability of WES. With high coverage and read depth, point mutations and small indels in exonic regions can be detected with high efficiency, but those in regulatory regions are not. In addition, larger multi-exon or multigene deletions and duplications, which contribute to a significant proportion of the mutation spectrum for several genes, as well as gene fusion or chimeric events common in cancer, are not efficiently detected. Besides variant detection capability, another major challenge of the test involves assessment of the clinical implications of variants identified. Most of the observed variants may not be clinically predictable or actionable due to lack of sufficient evidence. However, with the routine practice of WES and accumulation of relevant information, this concern would gradually be reduced. The fifth challenge to implementing WES assays in clinical care is the requirement of additional training for physicians to help them interpret test results and reports. With a more comprehensive set of variants available for consideration in the patient's clinical context, clinicians who see the patient, if trained in this area, would be able to make the optimal interpretation as to the causative gene. Alternatively or ideally simultaneously, extensive phenotypic information may be collected beforehand and made available to the pathologists and laboratorians interpreting the data. Finally, a considerable challenge facing the clinics and laboratories that offer these tests is the constantly changing technology. Recently, members of the Standardization of Clinical Testing workgroup (Nex-StoCT) have laid out guidelines for the validation and implementation of NGS-based tests [122]. With NGS technology changing all the time, however, these aspects also change and can become a hurdle to implementation.

Despite the challenges and limitations, WES and WGS have stirred tremendous interest, with the future of clinical care promising expedited diagnosis and more personalized medicine. Moreover, implementation of WES in medical practice will potentially aid the advancement of our understanding of human biology and pathogenesis.

Comparison of Gene Panel Sequencing, WES, and WGS: What Have We Learnt?

NGS-based approaches have been widely used in research and clinical laboratories for discovery and diagnosis of Mendelian disorders. These approaches such as gene panel sequencing and WES target a subset of genomic region known to be associated with a disease type and the entire coding region, respectively, having a higher depth of coverage compared to WGS. In both WES and gene panels, the target selection approaches such as "amplicon-based selection" and "hybridization-based selection" can introduce bias due to complexity of the sequence such as GC content and repetitive sequence which results in lack of optimal targeting. WGS also suffer from the same bias till date (see Table 17.2 for comparison) but with lesser depth of coverage than WES.

Large populations of children for pediatric disease including childhood cancer can be screened using data from WES to catalog germline variants in genes that may be involved in cancer predisposition, allowing better disease management [123]. Both WES and WGS have been used to detect

Table 17.2 Comparison between NGS-based targeted gene panel sequencing, WES, and WGS

Categories	Targeted panels	WES	WGS
Genomic target region	Smaller target regions/genes associated based on disease type or initial clinical diagnosis	Entire coding region (all exons) of the genome (~95% coverage)	Entire genome
Depth of coverage	Highest since specific regions are targeted. Sanger sequencing fill-in for drop out regions/exons is possible. Libraries can be boosted by targeting specific noncoding regions, introns with known disease-associated variants	100 X Sanger sequencing fill-in possibility for a few targeted exons that drop out or with low coverage in genes related to phenotype is possible	30 X Sanger fill-in possibility for a few target exons in genes related to phenotype possible
Sequencing bias	Based on target method	Based on target selection method, not including noncoding regions	Least
Variant detection	Difficult for SVs such as CNVs, indels, trinucleotide repeats, pseudogenes	Difficult for SVs such as CNVs, indels, trinucleotide repeats, pseudogenes	Allows for SVs such as CNVs, indels; difficult for trinucleotide repeats, pseudogenes
Incidental findings	Incidental findings (allelic changes; extended pathogenic variant clinical spectrum)	Need to address incidental findings	Need to address incidental findings

molecular biomarkers, allowing a merging of research and clinical therapeutics such as in cancer [124, 125]. Current exome capture kits can capture 95% of the coding regions [126]. The clinical diagnostic yield of WES highly varies, ranging from 25% to as high as 60% of selected cases depending on number, ethnicity, trio (both parents and proband), or singleton sequencing and disease cohort, with exceptional rate of finding de novo variants by trio sequencing across both rare and more common diseases [127–134]. WES can also detect carrier status of recessive diseases and variants that may affect the patient's response to various pharmaceutical drugs [135]. For diseases with high genetic heterogeneity, WES and also WGS may be more efficient than gene panel testing – for example, Leslie et al. [136] used WES in combination with Sanger sequencing to elucidate the broader genotypic and phenotypic spectra of popliteal pterygium disorders. Recently in multiple cohort studies, WES has been shown to have higher clinical sensitivity in detecting pathogenic variants in undiagnosed patients compared to conventional Sanger sequencing or traditional clinical diagnosis which strengthens its position as an important genetic diagnostic strategy for Mendelian and complex disorders [137–139].

A Look to the Future of Whole-Exome Assays

Current commercially available NGS technologies have already revolutionized the diagnostic capacity of modern clinical genetics. Nevertheless, advanced so-called "third-generation" sequencing technologies, such as Helicos HeliScope (Helicos Biosciences Corporation, Cambridge, MA), PacBio SMRT (Pacific Biosciences, California), and nanopore sequencers (Oxford Nanopore Technologies, Oxford, UK), are being actively developed to further improve genomic sequencing applications [140]. These third-generation sequencing platforms differ from the current technologies in that the initial target capture and enrichment step, which involves DNA amplification, are no longer required. The input patient DNA is sequenced and analyzed at the single-molecule level with the help of engineered protein polymerases [140]. This will not only cut cost and turnaround time but also have the added advantage of avoiding any in vitro amplification bias. Upon thorough validation and optimization of their diagnostic ability, these future technologies promise to move today's medical practice to the anticipated next level of care.

Currently, even more so than the sequencing technology and needed coverage improvements, the progress in data analysis tools and candidate variant filtration is of major concern. WES alone, which interrogates about 1% of the human genome, returns a list of about 20,000 variant calls [141]. Family information, such as the mode of inheritance within a family, linkage analysis, or variant data, i.e., the WES profile of unaffected family members, helps eliminate familial normal variations and track down disease-causing mutations [60], but performing additional tests including WES on multiple family members increases diagnostic costs and is not ideal for a variety of reasons. As more exomes are analyzed and sequence variants reported in publicly available databases, however, variant analysis and disease diagnosis by WES will certainly become easier and faster.

Meanwhile, with the implementation of WES and NGS technologies in clinical pathology becoming more common, the need for trained pathologists capable of interpreting the data and assessing the potential impact on an individual's

health is growing. The training of future pathologists is now under discussion, and teaching curricula in genomics and personalized medicine are being actively developed for residents [142, 143]. A national committee of Pathology Program Directors and other experts has also recently formed to develop model curricula and promote their widespread implementation [142, 144]. The implementation of WES and WGS in clinical practice has, therefore, added a new dimension to the already multifaceted roles of pathologists.

Conclusions

With more than 85% of causative mutations harbored in as little as 1% of the entire human genome, the use of WES as the most efficient strategy for disease diagnosis seems well justified. Even though WGS has the potential to identify CNVs and point mutations in exons, as well as in regulatory regions of the introns, the cost, time, and the analysis burden currently involved has meant WGS is on hold for clinical implementation, at least for now. Substantial proof-of-principle studies and evidence of diagnostic capability, affordability, and feasibility in the clinical setting have supported the use of WES. Currently, it is offered for clinical diagnosis by multiple major clinical laboratories across the USA, and as the technology improves and becomes less expensive, more laboratories are beginning to develop the test.

Clinicians who contemplate ordering a WES assay should first consider other available tests, such as relatively comprehensive gene panels. Gene panels, which interrogate only a limited number of genes, each more or less associated with the patient's clinical presentation, more completely retain the integrity of the individual's genetic information. Appropriate ethical guidelines and data-masking features during data analysis will likely overcome this difference eventually and make WES widely acceptable for rare diseases, cancer, and prenatal and infectious disease diagnosis. Finally, reductions in cost, more robust technologies, and improved data storage processes will soon make clinical WGS feasible, as well. The future of medical care can be envisioned as an integrated approach, with pathologists, geneticists, and other physicians all contributing to make informed decisions about patient management and treatment.

References

1. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80. https://doi.org/10.1038/nature03959.
2. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4(11):903–5. https://doi.org/10.1038/nmeth1111.
3. Bashiardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M. Direct genomic selection. Nat Methods. 2005;2(1):63–9. https://doi.org/10.1038/nmeth0105-63.
4. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182–9. https://doi.org/10.1038/nbt.1523.
5. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007;39(12):1522–7. https://doi.org/10.1038/ng.2007.42.
6. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7(2):111–8. https://doi.org/10.1038/nmeth.1419.
7. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359.
8. Ansorge WJ. Next-generation DNA sequencing techniques. New Biotechnol. 2009;25(4):195–203. https://doi.org/10.1016/j.nbt.2008.12.009.
9. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010;11(1):31–46. https://doi.org/10.1038/nrg2626.
10. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135–45. https://doi.org/10.1038/nbt1486.
11. Gowrisankar S, Lerner-Ellis JP, Cox S, White ET, Manion M, LeVan K, Liu J, Farwell LM, Iartchouk O, Rehm HL, Funke BH. Evaluation of second-generation sequencing of 19 dilated cardiomyopathy genes for clinical applications. J Mol Diagn. 2010;12(6):818–27. https://doi.org/10.2353/jmoldx.2010.100014.
12. Jones MA, Bhide S, Chin E, Ng BG, Rhodenizer D, Zhang VW, Sun JJ, Tanner A, Freeze HH, Hegde MR. Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. Genet Med. 2011;13(11):921–32. https://doi.org/10.1097/GIM.0b013e318226fbf2.
13. Lemke JR, Riesch E, Scheurenbrand T, Schubach M, Wilhelm C, Steiner I, Hansen J, Courage C, Gallati S, Burki S, Strozzi S, Simonetti BG, Grunt S, Steinlin M, Alber M, Wolff M, Klopstock T, Prott EC, Lorenz R, Spaich C, Rona S, Lakshminarasimhan M, Kroll J, Dorn T, Kramer G, Synofzik M, Becker F, Weber YG, Lerche H, Bohm D, Biskup S. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia. 2012;53(8):1387–98. https://doi.org/10.1111/j.1528-1167.2012.03516.x.
14. Lin X, Tang W, Ahmad S, Lu J, Colby CC, Zhu J, Yu Q. Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities. Hear Res. 2012;288(1–2):67–76. https://doi.org/10.1016/j.heares.2012.01.004.
15. Neveling K, Collin RW, Gilissen C, van Huet RA, Visser L, Kwint MP, Gijsen SJ, Zonneveld MN, Wieskamp N, de Ligt J,

Siemiatkowska AM, Hoefsloot LH, Buckley MF, Kellner U, Branham KE, den Hollander AI, Hoischen A, Hoyng C, Klevering BJ, van den Born LI, Veltman JA, Cremers FP, Scheffer H. Next-generation genetic testing for retinitis pigmentosa. Hum Mutat. 2012;33(6):963–72. https://doi.org/10.1002/humu.22045.

16. Valencia CA, Ankala A, Rhodenizer D, Bhide S, Littlejohn MR, Keong LM, Rutkowski A, Sparks S, Bonnemann C, Hegde M. Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. PLoS One. 2013;8(1):e53083. https://doi.org/10.1371/journal.pone.0053083.
17. Valencia CA, Rhodenizer D, Bhide S, Chin E, Littlejohn MR, Keong LM, Rutkowski A, Bonnemann C, Hegde M. Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy. J Mol Diagn. 2012;14(3):233–46. https://doi.org/10.1016/j.jmoldx.2012.01.009.
18. Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. Genome Med. 2009;1(10):100. https://doi.org/10.1186/gm100.
19. Ankala A, Kohn JN, Dastur R, Gaitonde P, Khadilkar SV, Hegde MR. Ancestral founder mutations in calpain-3 in the Indian Agarwal community: historical, clinical, and molecular perspective. Muscle Nerve. 2013;47(6):931–7. https://doi.org/10.1002/mus.23763.
20. Mitra AP, Pagliarulo V, Yang D, Waldman FM, Datar RH, Skinner DG, Groshen S, Cote RJ. Generation of a concise gene panel for outcome prediction in urinary bladder cancer. J Clin Oncol. 2009;27(24):3929–37. https://doi.org/10.1200/JCO.2008.18.5744.
21. Holcomb CL, Hoglund B, Anderson MW, Blake LA, Bohme I, Egholm M, Ferriola D, Gabriel C, Gelber SE, Goodridge D, Hawbecker S, Klein R, Ladner M, Lind C, Monos D, Pando MJ, Proll J, Sayer DC, Schmitz-Agheguian G, Simen BB, Thiele B, Trachtenberg EA, Tyan DB, Wassmuth R, White S, Erlich HA. A multi-site study using high-resolution HLA genotyping by next generation sequencing. Tissue Antigens. 2011;77(3):206–17. https://doi.org/10.1111/j.1399-0039.2010.01606.x.
22. Proll J, Danzer M, Stabentheiner S, Niklas N, Hackl C, Hofer K, Atzmuller S, Hufnagl P, Gully C, Hauser H, Krieger O, Gabriel C. Sequence capture and next generation resequencing of the MHC region highlights potential transplantation determinants in HLA identical haematopoietic stem cell transplantation. DNA Res. 2011;18(4):201–10. https://doi.org/10.1093/dnares/dsr008.
23. Serizawa M, Sekizuka T, Okutani A, Banno S, Sata T, Inoue S, Kuroda M. Genomewide screening for novel genetic variations associated with ciprofloxacin resistance in Bacillus anthracis. Antimicrob Agents Chemother. 2010;54(7):2787–92. https://doi.org/10.1128/AAC.01405-09.
24. Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Arch Pathol Lab Med. 2013;137(3):415–33. https://doi.org/10.5858/arpa.2012-0107-RA.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
26. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8. https://doi.org/10.1038/ng.806.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603.
29. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24. https://doi.org/10.1038/gim.2015.30.
30. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, Kim JH, Rosenthal EA, Kim DS, National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project, Tabor HK, Bamshad MJ, Motulsky AG, Scott CR, Pritchard CC, Walsh T, Burke W, Raskind WH, Byers P, Hisama FM, Nickerson DA, Jarvik GP. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet. 2013;93(4):631–40. https://doi.org/10.1016/j.ajhg.2013.08.006.
31. Hegde M, Bale S, Bayrak-Toydemir P, Gibson J, Jeng LJ, Joseph L, Laser J, Lubin IM, Miller CE, Ross LF, Rothberg PG, Tanner AK, Vitazka P, Mao R. Reporting incidental findings in genomic scale clinical sequencing--a clinical laboratory perspective: a report of the Association for Molecular Pathology. J Mol Diagn. 2015;17(2):107–17. https://doi.org/10.1016/j.jmoldx.2014.10.004.
32. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med. 2017;19(10):1151–8. https://doi.org/10.1038/gim.2017.26.
33. Robinson PN, Kohler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, Krawitz P, Gilissen C, Haendel M, Smedley D. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res. 2014;24(2):340–8. https://doi.org/10.1101/gr.160325.113.
34. Frankish A, Harrow J. GENCODE pseudogenes. Methods Mol Biol. 2014;1167:129–55. https://doi.org/10.1007/978-1-4939-0835-6_10.
35. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, Reymond A, Hubbard TJ, Harrow J, Gerstein MB. The GENCODE pseudogene resource. Genome Biol. 2012;13(9):R51. https://doi.org/10.1186/gb-2012-13-9-r51.
36. Gymrek M, Golan D, Rosset S, Erlich Y. lobSTR: a short tandem repeat profiler for personal genomes. Genome Res. 2012;22(6):1154–62. https://doi.org/10.1101/gr.135780.111.
37. Cao MD, Balasubramanian S, Boden M. Sequencing technologies and tools for short tandem repeat variation detection. Brief Bioinform. 2015;16(2):193–204. https://doi.org/10.1093/bib/bbu001.
38. Gelfand Y, Hernandez Y, Loving J, Benson G. VNTRseek-a computational tool to detect tandem repeat variants in high-throughput sequencing data. Nucleic Acids Res. 2014;42(14):8884–94. https://doi.org/10.1093/nar/gku642.
39. Bean LJ, Hegde MR. Gene variant databases and sharing: creating a global genomic variant database for personalized medicine. Hum Mutat. 2016;37(6):559–63. https://doi.org/10.1002/humu.22982.
40. Bean LJ, Hegde MR. Gene variant databases and sharing: creating a global genomic variant database for personalized medicine. Hum Mutat. 2017;38(1):122. https://doi.org/10.1002/humu.23064.

41. Garber KB, Vincent LM, Alexander JJ, Bean LJH, Bale S, Hegde M. Reassessment of genomic sequence variation to harmonize interpretation for personalized medicine. Am J Hum Genet. 2016;99(5):1140–9. https://doi.org/10.1016/j.ajhg.2016.09.015.
42. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017;100(2):267–80. https://doi.org/10.1016/j.ajhg.2017.01.004.
43. Bean LJH, Hegde MR. Clinical implications and considerations for evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Med. 2017;9(1):111. https://doi.org/10.1186/s13073-017-0508-z.
44. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. 2017;18(1):225. https://doi.org/10.1186/s13059-017-1353-5.
45. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877–85. https://doi.org/10.1016/j.ajhg.2016.08.016.
46. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics. 2013;14(Suppl 3):S3. https://doi.org/10.1186/1471-2164-14-S3-S3.
47. Dong L, Shi B, Tian G, Li Y, Wang B, Zhou M. An accurate de novo algorithm for glycan topology determination from mass spectra. IEEE/ACM Trans Comput Biol Bioinform. 2015;12(3):568–78. https://doi.org/10.1109/TCBB.2014.2368981.
48. Katsonis P, Koire A, Wilson SJ, Hsu TK, Lua RC, Wilkins AD, Lichtarge O. Single nucleotide variations: biological impact and theoretical interpretation. Protein Sci. 2014;23(12):1650–66. https://doi.org/10.1002/pro.2552.
49. Nakken S, Fournous G, Vodak D, Aasheim LB, Myklebost O, Hovig E. Personal cancer genome reporter: variant interpretation report for precision oncology. Bioinformatics. 2017;34(10):1778–80. https://doi.org/10.1093/bioinformatics/btx817.
50. Chakravorty S, Hegde M. Gene and variant annotation for Mendelian disorders in the era of advanced sequencing technologies. Annu Rev Genomics Hum Genet. 2017;18:229–56. https://doi.org/10.1146/annurev-genom-083115-022545.
51. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009;106(45):19096–101. https://doi.org/10.1073/pnas.0910672106.
52. Hebert SC. Bartter syndrome. Curr Opin Nephrol Hypertens. 2003;12(5):527–32. https://doi.org/10.1097/01.mnh.0000088732.87142.43.
53. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245(4922):1073–80.
54. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N. Haploinsufficiency of NSD1 causes Sotos syndrome. Nat Genet. 2002;30(4):365–6. https://doi.org/10.1038/ng863.
55. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science. 1987;236(4808):1567–70.
56. Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet. 2004;36(9):955–7. https://doi.org/10.1038/ng1407.
57. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42(1):30–5. https://doi.org/10.1038/ng.499.
58. Miller M, Fineman R, Smith DW. Postaxial acrofacial dysostosis syndrome. J Pediatr. 1979;95(6):970–5.
59. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42(9):790–3. https://doi.org/10.1038/ng.646.
60. Maxmen A. Exome sequencing deciphers rare diseases. Cell. 2011;144(5):635–7. https://doi.org/10.1016/j.cell.2011.02.033.
61. St Hilaire C, Ziegler SG, Markello TC, Brusco A, Groden C, Gill F, Carlson-Donohoe H, Lederman RJ, Chen MY, Yang D, Siegenthaler MP, Arduino C, Mancini C, Freudenthal B, Stanescu HC, Zdebik AA, Chaganti RK, Nussbaum RL, Kleta R, Gahl WA, Boehm M. NT5E mutations and arterial calcifications. N Engl J Med. 2011;364(5):432–42. https://doi.org/10.1056/NEJMoa0912923.
62. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, Carlson-Donohoe H, Gropman A, Pierson TM, Golas G, Wolfe L, Groden C, Godfrey R, Nehrebecky M, Wahl C, Landis DM, Yang S, Madeo A, Mullikin JC, Boerkoel CF, Tifft CJ, Adams D. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. Genet Med. 2012;14(1):51–9. https://doi.org/10.1038/gim.0b013e318232a005.
63. Majewski J, Rosenblatt DS. Exome and whole-genome sequencing for gene discovery: the future is now! Hum Mutat. 2012;33(4):591–2. https://doi.org/10.1002/humu.22055.
64. Mefford HC. Diagnostic exome sequencing--are we there yet? N Engl J Med. 2012;367(20):1951–3. https://doi.org/10.1056/NEJMe1211659.
65. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003;33(Suppl):228–37. https://doi.org/10.1038/ng1090.
66. Bloch-Zupan A, Jamet X, Etard C, Laugel V, Muller J, Geoffroy V, Strauss JP, Pelletier V, Marion V, Poch O, Strahle U, Stoetzel C, Dollfus H. Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects. Am J Hum Genet. 2011;89(6):773–81. https://doi.org/10.1016/j.ajhg.2011.11.002.
67. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet. 2012;49(6):353–61. https://doi.org/10.1136/jmedgenet-2012-100819.
68. Cabral RM, Kurban M, Wajid M, Shimomura Y, Petukhova L, Christiano AM. Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. Genomics. 2012;99(4):202–8. https://doi.org/10.1016/j.ygeno.2012.01.005.

69. Campeau PM, Lu JT, Sule G, Jiang MM, Bae Y, Madan S, Hogler W, Shaw NJ, Mumm S, Gibbs RA, Whyte MP, Lee BH. Whole-exome sequencing identifies mutations in the nucleoside transporter gene SLC29A3 in dysosteosclerosis, a form of osteopetrosis. Hum Mol Genet. 2012;21(22):4904–9. https://doi.org/10.1093/hmg/dds326.
70. Dundar H, Ozgul RK, Yalnizoglu D, Erdem S, Oguz KK, Tuncel D, Temucin CM, Dursun A. Identification of a novel Twinkle mutation in a family with infantile onset spinocerebellar ataxia by whole exome sequencing. Pediatr Neurol. 2012;46(3):172–7. https://doi.org/10.1016/j.pediatrneurol.2011.12.006.
71. Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, Frye M, Al-Gazali L, Gleeson JG. Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. J Med Genet. 2012;49(6):380–5. https://doi.org/10.1136/jmedgenet-2011-100686.
72. Leidenroth A, Sorte HS, Gilfillan G, Ehrlich M, Lyle R, Hewitt JE. Diagnosis by sequencing: correction of misdiagnosis from FSHD2 to LGMD2A by whole-exome analysis. Eur J Hum Genet. 2012;20(9):999–1003. https://doi.org/10.1038/ejhg.2012.42.
73. Antonarakis SE, Beckmann JS. Mendelian disorders deserve more attention. Nat Rev Genet. 2006;7(4):277–82. https://doi.org/10.1038/nrg1826.
74. Oti M, Brunner HG. The modular nature of genetic diseases. Clin Genet. 2007;71(1):1–11. https://doi.org/10.1111/j.1399-0004.2006.00708.x.
75. Peltonen L, Perola M, Naukkarinen J, Palotie A. Lessons from studying monogenic disease for common disease. Hum Mol Genet. 2006;15(1):R67–74. https://doi.org/10.1093/hmg/ddl060.
76. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 2011;1380:42–77. https://doi.org/10.1016/j.brainres.2010.11.078.
77. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Puttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wieczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJ, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature. 2011;478(7367):57–63. https://doi.org/10.1038/nature10423.
78. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993;261(5123):921–3.
79. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385–9. https://doi.org/10.1126/science.1109557.
80. Tan EK. Identification of a common genetic risk variant (LRRK2 Gly2385Arg) in Parkinson's disease. Ann Acad Med Singap. 2006;35(11):840–2.
81. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010;11(6):415–25. https://doi.org/10.1038/nrg2779.
82. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53. https://doi.org/10.1038/nature08494.
83. McClellan J, King MC. Genetic heterogeneity in human disease. Cell. 2010;141(2):210–7. https://doi.org/10.1016/j.cell.2010.03.032.
84. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19(3):212–9. https://doi.org/10.1016/j.gde.2009.04.010.
85. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, Rickaby J, Lu C, Szatmari P, Roberts W, Fernandez BA, Marshall CR, Hatchwell E, Eis PS, Scherer SW. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. G3 (Bethesda). 2012;2(12):1665–85. https://doi.org/10.1534/g3.112.004689.
86. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, Ogilvie C, Ahn JW, Drmic I, Senman L, Chrysler C, Thompson A, Russell C, Prasad A, Walker S, Pinto D, Marshall CR, Stavropoulos DJ, Zwaigenbaum L, Fernandez BA, Fombonne E, Bolton PF, Collier DA, Hodge JC, Roberts W, Szatmari P, Scherer SW. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet. 2012;90(1):133–41. https://doi.org/10.1016/j.ajhg.2011.11.025.
87. Bogardus C. Missing heritability and GWAS utility. Obesity (Silver Spring). 2009;17(2):209–10. https://doi.org/10.1038/oby.2008.613.
88. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol. 2010;8(1):e1000294. https://doi.org/10.1371/journal.pbio.1000294.
89. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11(6):446–50. https://doi.org/10.1038/nrg2809.
90. Nielsen R. Genomics: in search of rare human variants. Nature. 2010;467(7319):1050–1. https://doi.org/10.1038/4671050a.
91. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 2001;69(1):124–37. https://doi.org/10.1086/321272.
92. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. Hum Mol Genet. 2012;21(R1):R1–9. https://doi.org/10.1093/hmg/dds387.
93. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Lehner T, Shugart YY, Price AL, de Bakker PI, Purcell SM, Sunyaev SR. Exome sequencing and the genetic basis of complex traits. Nat Genet. 2012;44(6):623–30. https://doi.org/10.1038/ng.2303.
94. Berman JJ. Tumor taxonomy for the developmental lineage classification of neoplasms. BMC Cancer. 2004;4:88. https://doi.org/10.1186/1471-2407-4-88.
95. de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI thesaurus: using science-based terminology to integrate cancer research results. Stud Health Technol Inform. 2004;107.(Pt 1:33–7.
96. Berman J. Modern classification of neoplasms: reconciling differences between morphologic and molecular approaches. BMC Cancer. 2005;5:100. https://doi.org/10.1186/1471-2407-5-100.
97. Lonigro RJ, Grasso CS, Robinson DR, Jing X, Wu YM, Cao X, Quist MJ, Tomlins SA, Pienta KJ, Chinnaiyan AM. Detection of somatic copy number alterations in cancer using targeted exome capture sequencing. Neoplasia. 2011;13(11):1019–25.
98. Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S, Quackenbush J, Nelson SF. Exome sequencing-based

copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics. 2011;27(19):2648–54. https://doi.org/10.1093/bioinformatics/btr462.
99. Chang H, Jackson DG, Kayne PS, Ross-Macdonald PB, Ryseck RP, Siemers NO. Exome sequencing reveals comprehensive genomic alterations across eight cancer cell lines. PLoS One. 2011;6(6):e21097. https://doi.org/10.1371/journal.pone.0021097.
100. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet. 1998;20(2):207–11. https://doi.org/10.1038/2524.
101. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008;40(10):1166–74. https://doi.org/10.1038/ng.238.
102. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet. 2011;43(4):309–15. https://doi.org/10.1038/ng.788.
103. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, Tuveson DA, Perez-Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell PJ, Teh BT, Stratton MR, Futreal PA. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469(7331):539–42. https://doi.org/10.1038/nature09639.
104. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18(1):11–22. https://doi.org/10.1016/j.ccr.2010.05.026.
105. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15. https://doi.org/10.1038/nature10166.
106. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, Sawyers CL. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med. 2001;344(14):1031–7. https://doi.org/10.1056/NEJM200104053441401.
107. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129–39. https://doi.org/10.1056/NEJMoa040938.
108. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500. https://doi.org/10.1126/science.1099314.
109. Barton NH. Genetic hitchhiking. Philos Trans R Soc Lond Ser B Biol Sci. 2000;355(1403):1553–62. https://doi.org/10.1098/rstb.2000.0716.
110. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57–70.
111. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010;42(2):181–5. https://doi.org/10.1038/ng.518.
112. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009;461(7265):809–13. https://doi.org/10.1038/nature08489.
113. Taylor BS, Ladanyi M. Clinical cancer genomics: how soon is now? J Pathol. 2011;223(2):318–26. https://doi.org/10.1002/path.2794.
114. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949–54. https://doi.org/10.1038/nature00766.
115. Millington GW. Mutations of the BRAF gene in human cancer, by Davies et al. (Nature 2002; 417: 949-54). Clin Exp Dermatol. 2013;38(2):222–3. https://doi.org/10.1111/ced.12015.
116. Lee JT, Li L, Brafford PA, van den Eijnden M, Halloran MB, Sproesser K, Haass NK, Smalley KS, Tsai J, Bollag G, Herlyn M. PLX4032, a potent inhibitor of the B-Raf V600E oncogene, selectively inhibits V600E-positive melanomas. Pigment Cell Melanoma Res. 2010;23(6):820–7. https://doi.org/10.1111/j.1755-148X.2010.00763.x.
117. Yang H, Higgins B, Kolinsky K, Packman K, Go Z, Iyer R, Kolis S, Zhao S, Lee R, Grippo JF, Schostack K, Simcox ME, Heimbrook D, Bollag G, Su F. RG7204 (PLX4032), a selective BRAFV600E inhibitor, displays potent antitumor activity in preclinical melanoma models. Cancer Res. 2010;70(13):5518–27. https://doi.org/10.1158/0008-5472.CAN-10-0646.
118. Shi H, Moriceau G, Kong X, Lee MK, Lee H, Koya RC, Ng C, Chodon T, Scolyer RA, Dahlman KB, Sosman JA, Kefford RF, Long GV, Nelson SF, Ribas A, Lo RS. Melanoma whole-exome sequencing identifies (V600E)B-RAF amplification-mediated acquired B-RAF inhibitor resistance. Nat Commun. 2012;3:724. https://doi.org/10.1038/ncomms1727.
119. Solomon BD, Pineda-Alvarez DE, Hadley DW, Program NCS, Teer JK, Cherukuri PF, Hansen NF, Cruz P, Young AC, Blakesley RW, Lanpher B, Mayfield Gibson S, Sincan M, Chandrasekharappa SC, Mullikin JC. Personalized genomic medicine: lessons from the exome. Mol Genet Metab. 2011;104(1–2):189–91. https://doi.org/10.1016/j.ymgme.2011.06.022.
120. Hoischen A, Gilissen C, Arts P, Wieskamp N, van der Vliet W, Vermeer S, Steehouwer M, de Vries P, Meijer R, Seiqueros J, Knoers NV, Buckley MF, Scheffer H, Veltman JA. Massively parallel sequencing of ataxia genes after array-based enrichment. Hum Mutat. 2010;31(4):494–9. https://doi.org/10.1002/humu.21221.

121. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011;13(1):36–46. https://doi.org/10.1038/nrg3117.
122. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, Lu F, Lyon E, Voelkerding KV, Zehnbauer BA, Agarwala R, Bennett SF, Chen B, Chin EL, Compton JG, Das S, Farkas DH, Ferber MJ, Funke BH, Furtado MR, Ganova-Raeva LM, Geigenmuller U, Gunselman SJ, Hegde MR, Johnson PL, Kasarskis A, Kulkarni S, Lenk T, Liu CS, Manion M, Manolio TA, Mardis ER, Merker JD, Rajeevan MS, Reese MG, Rehm HL, Simen BB, Yeakley JM, Zook JM, Lubin IM. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30(11):1033–6. https://doi.org/10.1038/nbt.2403.
123. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR, Vadodaria B, Chen X, McGee RB, Hines-Dowell S, Nuccio R, Quinn E, Shurtleff SA, Rusch M, Patel A, Becksfort JB, Wang S, Weaver MS, Ding L, Mardis ER, Wilson RK, Gajjar A, Ellison DW, Pappo AS, Pui CH, Nichols KE, Downing JR. Germline mutations in predisposition genes in pediatric cancer. N Engl J Med. 2015;373(24):2336–46. https://doi.org/10.1056/NEJMoa1508054.
124. Helleday T. Cancer phenotypic lethality, exemplified by the non-essential MTH1 enzyme being required for cancer survival. Ann Oncol. 2014;25(7):1253–5. https://doi.org/10.1093/annonc/mdu158.
125. Offit K. Decade in review--genomics: a decade of discovery in cancer genomics. Nat Rev Clin Oncol. 2014;11(11):632–4. https://doi.org/10.1038/nrclinonc.2014.170.
126. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Hum Mutat. 2015;36(8):815–22. https://doi.org/10.1002/humu.22813.
127. Dragojlovic N, Elliott AM, Adam S, van Karnebeek C, Lehman A, Mwenifumbo JC, Nelson TN, du Souich C, Friedman JM, Lynd LD. The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. Genet Med. 2018; https://doi.org/10.1038/gim.2017.226.
128. Long PA, Evans JM, Olson TM. Diagnostic yield of whole exome sequencing in pediatric dilated cardiomyopathy. J Cardiovasc Dev Dis. 2017;4(3):11. https://doi.org/10.3390/jcdd4030011.
129. Parsons DW, Roy A, Yang Y, Wang T, Scollon S, Bergstrom K, Kerstein RA, Gutierrez S, Petersen AK, Bavle A, Lin FY, Lopez-Terrada DH, Monzon FA, Hicks MJ, Eldin KW, Quintanilla NM, Adesina AM, Mohila CA, Whitehead W, Jea A, Vasudevan SA, Nuchtern JG, Ramamurthy U, McGuire AL, Hilsenbeck SG, Reid JG, Muzny DM, Wheeler DA, Berg SL, Chintagumpala MM, Eng CM, Gibbs RA, Plon SE. Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. JAMA Oncol. 2016; https://doi.org/10.1001/jamaoncol.2015.5699.
130. Rossi M, El-Khechen D, Black MH, Farwell Hagman KD, Tang S, Powis Z. Outcomes of diagnostic exome sequencing in patients with diagnosed or suspected autism spectrum disorders. Pediatr Neurol. 2017;70(34–43):e32. https://doi.org/10.1016/j.pediatrneurol.2017.01.033.
131. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH Jr, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014;46(9):944–50. https://doi.org/10.1038/ng.3050.
132. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Koster J, Marais A, Paknia O, Schroder R, Garcia-Aznar JM, Werber M, Brandau O, Calvo Del Castillo M, Baldi C, Wessel K, Kishore S, Nahavandi N, Eyaid W, Al Rifai MT, Al-Rumayyan A, Al-Twaijri W, Alothaim A, Alhashem A, Al-Sannaa N, Al-Balwi M, Alfadhel M, Rolfs A, Abou Jamra R. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017;25(2):176–82. https://doi.org/10.1038/ejhg.2016.146.
133. Vissers L, van Nimwegen KJM, Schieving JH, Kamsteeg EJ, Kleefstra T, Yntema HG, Pfundt R, van der Wilt GJ, Krabbenborg L, Brunner HG, van der Burg S, Grutters J, Veltman JA, Willemsen M. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. Genet Med. 2017;19(9):1055–63. https://doi.org/10.1038/gim.2017.1.
134. Yavarna T, Al-Dewik N, Al-Mureikhi M, Ali R, Al-Mesaifri F, Mahmoud L, Shahbeck N, Lakhani S, AlMulla M, Nawaz Z, Vitazka P, Alkuraya FS, Ben-Omran T. High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. Hum Genet. 2015;134(9):967–80. https://doi.org/10.1007/s00439-015-1575-0.
135. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med. 2015;17(6):444–51. https://doi.org/10.1038/gim.2014.122.
136. Leslie EJ, O'Sullivan J, Cunningham ML, Singh A, Goudy SL, Ababneh F, Alsubaie L, Ch'ng GS, van der Laar IM, Hoogeboom AJ, Dunnwald M, Kapoor S, Jiramongkolchai P, Standley J, Manak JR, Murray JC, Dixon MJ. Expanding the genetic and phenotypic spectrum of popliteal pterygium disorders. Am J Med Genet A. 2015;167A(3):545–52. https://doi.org/10.1002/ajmg.a.36896.
137. Harris E, Topf A, Barresi R, Hudson J, Powell H, Tellez J, Hicks D, Porter A, Bertoli M, Evangelista T, Marini-Betollo C, Magnusson O, Lek M, MacArthur D, Bushby K, Lochmuller H, Straub V. Exome sequences versus sequential gene testing in the UK highly specialised Service for Limb Girdle Muscular Dystrophy. Orphanet J Rare Dis. 2017;12(1):151. https://doi.org/10.1186/s13023-017-0699-9.
138. Schofield D, Alam K, Douglas L, Shrestha R, MacArthur DG, Davis M, Laing NG, Clarke NF, Burns J, Cooper ST, North KN, Sandaradura SA, O'Grady GL. Cost-effectiveness of massively parallel sequencing for diagnosis of paediatric muscle diseases. NPJ Genom Med. 2017;2:4. https://doi.org/10.1038/s41525-017-0006-7.
139. Johnson K, Topf A, Bertoli M, Phillips L, Claeys KG, Stojanovic VR, Peric S, Hahn A, Maddison P, Akay E, Bastian AE, Lusakowska A, Kostera-Pruszczyk A, Lek M, Xu L, MacArthur DG, Straub V. Identification of GAA variants through whole exome sequencing targeted to a cohort of 606 patients with unexplained limb-girdle muscle weakness. Orphanet J Rare Dis. 2017;12(1):173. https://doi.org/10.1186/s13023-017-0722-1.
140. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. J Pathol Inform. 2012;3:40. https://doi.org/10.4103/2153-3539.103013.
141. Stower H. The exome factor. Genome Biol. 2011;12(9):407. https://doi.org/10.1186/gb-2011-12-9-407.
142. Haspel RL, Arnaout R, Briere L, Kantarci S, Marchand K, Tonellato P, Connolly J, Boguski MS, Saffitz JE. A call to action: training pathology residents in genomics and personalized medicine. Am J Clin Pathol. 2010;133(6):832–4. https://doi.org/10.1309/AJCPN6Q1QKCLYKXM.
143. Schrijver I, Natkunam Y, Galli S, Boyd SD. Integration of genomic medicine into pathology residency training: the stanford open curriculum. J Mol Diagn. 2013;15(2):141–8. https://doi.org/10.1016/j.jmoldx.2012.11.003.
144. Wall DP, Tonellato PJ. The future of genomics in pathology. F1000 Med Rep. 2012;4:14. https://doi.org/10.3410/M4-14.

Implementation of Genome Sequencing Assays

18

Tina M. Hambuch, Keith Nykamp, and Carri-Lyn Rebecca Mead

Introduction

The rapid adoption of genetic sequencing in clinical laboratories has been largely driven by the evolution of more efficient sequencing methods, enabling physicians' access to more detailed genetic information about their patients at costs and turnaround times that can be integrated into medical care. The rapid increase in availability of genetic sequence information has also enabled clinical discovery, which then forms the basis of new clinical tests; according to the Genetic Testing Registry (http://www.genetests.org) [1], as of 2017, more than 5000 genes are available for clinical testing, making genetic testing increasingly more relevant for more patients suspected of having a genetic disease. As our knowledge of disease biology and genetics increases, the reach and utility of clinical genetic testing will only continue to expand and improve. The implementation of NGS will undoubtedly further accelerate both discovery and testing. In this chapter, we focus on the implementation of whole-genome sequencing (WGS) as a clinical laboratory test. This chapter is organized according to the workflow, and sections are arranged in terms of pre-analytic, analytic, and post-analytic considerations (Fig. 18.1).

Whereas WGS may appear to be a single test, it has many possible indications for use, and each requires different handling throughout the process. Therefore, we discuss the possible clinical indications for testing and the pre-analytical, analytical, and post-analytical requirements for each of these applications. These issues are addressed with regard to current professional and regulatory best practices, guidelines, and resources [2, 3]. However, this field is evolving rapidly, and whereas the principles in this chapter are likely to remain consistent, many details such as specific resources or databases that are discussed are likely to change; therefore, we recommend that additional resources be consulted when implementing WGS in a clinical laboratory. It is an exciting time to be involved in clinical genetic testing, as there is an opportunity to help drive important advances in medical care. However, WGS, as a diagnostic test, is still in the nascent stages, straddling the line between clinical research and clinical testing. As a result, the diagnostic potential, utility, and challenges have not been entirely characterized. Until WGS diagnostic standards have been fully fleshed out and while inconsistencies exist across different laboratory offerings, open communication between laboratory professionals and the ordering physician, rigorous and transparent analytical and bioinformatic processes, and thoughtful policy development are critical when offering WGS as a clinical test.

Pre-analytical Considerations: Test Definition, Physician Support, and Process Development

The pre-analytical phase encompasses all steps taken prior to the actual testing of the sample. The introduction of any new test in a clinical laboratory requires several considerations prior to the physical launch of this clinical test. Several guidelines have been published to aid clinical laboratorians with the evaluation of when, how, and why to implement a new test ([4]; CLSI publications (multiple); CAP checklists). These guidelines include discussions of assessing clinical and regulatory concerns as well as financial and workflow considerations. Additional guidelines and recommendations specific to the implementation and offering of genomic sequencing testing have also recently been published [5–8]. The principles established in previous guidelines and best practice recommendations are still applicable and certainly should be included in the planning process. However, when the test involves a relatively new methodology that can be applied in a number of different ways, these multiple considerations must be refined and developed by the individual laboratories offering the testing. Newer guidelines, such as the ACMG (American College of Medical Genetics

T. M. Hambuch (✉) · K. Nykamp
Genetics, Invitae, San Francisco, CA, USA
e-mail: tina.hambuch@invitae.com

C.-L. R. Mead
Emerging Technologies, Illumina, San Diego, CA, USA

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_18

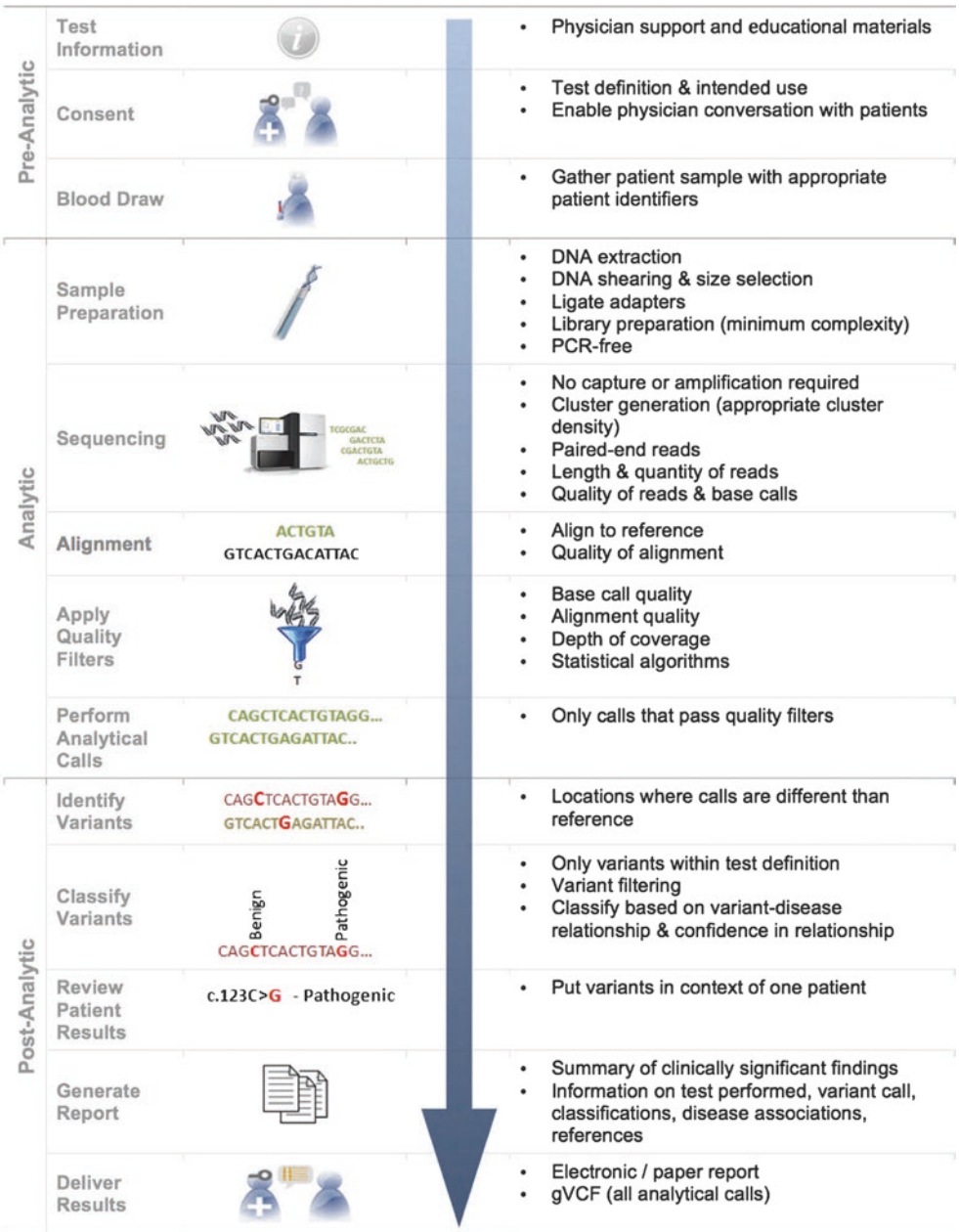


Fig. 18.1 Process and workflow for genome sequencing. This figure depicts the major steps in the processing of a genome through next-generation sequencing. The pre-analytic section illustrates the important steps in establishing the test and good communication between the ordering physician and the laboratory. The analytic section shows the processing of the physical sample in the laboratory and the calling of the data using bioinformatic processes. The post-analytic section depicts the steps involved in aggregating information about the results, interpreting those results, and generating a report that can be returned to the ordering physician. The sections in this chapter provide detailed descriptions of these steps

and Genomics) clinical laboratory standards for NGS [9], are particularly useful for considering the additional complexities that this new technology may introduce.

In the case of WGS, as it is commonly referred to, it is important to begin the initial assessment with a test definition and intended use statement to clarify the capabilities and expectations. The first clarification is that every base position in the genome will not be sequenced; depending on the application and the expense that the laboratory is willing to incur, about 97% of the mappable genome is currently captured with regularity. As this technology typically employs relatively short reads (less than 300 base fragments on average), it is not possible to map all fragments to the genome, particularly for regions with paralogs and other types of repetitive or duplication events. The second clarification is regarding the types of variants that can be detected. For example, all sequencing methodologies tend to be error-prone in regions with large nucleotide repeat expansions, such as the CGG repeat expansion associated with fragile X disease; WGS using NGS is no exception. Bioinformatic approaches have improved substantially in the last 5 years, leading to opportunities to analyze the same primary data using different programs in order to identify multiple types of variants. WGS is potentially able to detect many types of variants, including single nucleotide variants, copy num-

ber variants, small insertions, deletions, and translocation events. Additionally, methylation and expression can potentially be added to the analysis. However, NGS is not able to detect all of these different types of variants with the same levels of sensitivity and specificity. In particular, for clinical WGS, thresholds or statistical algorithms can be used to determine whether each variant call meets strict quality metrics that are used to ensure that when calls are made in a clinical context, they meet a minimum threshold of accuracy. Additionally, as detection of these different kinds of variants often requires a different bioinformatic analysis, the laboratory must have the infrastructure to perform multiple analyses and then integrate the results for downstream annotation and interpretation. This will be discussed further in the analytical portion of this chapter but is called out here to emphasize that clinical WGS requires additional rigor. Many clinician researchers in an academic setting will analyze these same data using relaxed quality thresholds in an effort to maximize sensitivity and opportunity for variant and gene discovery; however, this approach significantly increases the number of false positives, requiring additional follow-up with orthologous confirmatory testing before the results may be appropriate for aiding in medical management decisions. Enthusiastic clinicians may consider ordering this test for a patient, without realizing that WGS might require supplementary testing in order to prove useful and, in some cases, may not be the most appropriate test. Therefore, when offering clinical WGS, it is important to first define the primary goal of the test. Will it be used primarily for research and gene discovery, or is it strictly to aid in the diagnosis of rare genetic diseases? Secondly, it is critical that the development of the test and resources employed by the clinical laboratory match the primary goal.

Possibly the most powerful strength of WGS as a clinical test is that it is highly customizable; with the ability to detect most of the genetic variants for nearly all the genome, it is possible to evaluate and tailor the results for a variety of complex clinical presentations. For a critically ill child, it is an opportunity to assess nearly all possible causes of genetic disease simultaneously. This can save time and, in a small but significant percentage (~5%) of patients, may identify multiple genetic diagnoses. It also enables evaluation of genes that have been recently published and are not yet available using traditional single-gene or panel approaches. For patients already diagnosed with a heritable genetic disease using other types of testing (e.g., biochemical, cytogenetic, etc.), it may identify the gene or genes responsible and provide additional information for the family, including a better understanding of the prognosis or reproductive implications. In cases of cancer, both germline and somatic tissues can be sequenced to identify both heritable alterations and therapeutic targets to treat the cancer. It is also being used to screen prenatally for large cytogenetic events. Importantly, while all of these applications use WGS, to achieve optimal performance, development of assay components and analytical requirements will differ for each application.

A common misperception of genome (and exome) testing is that the entire genome (or exome) is reviewed and interpreted. This is often not true. Rather, interpretation is customized to focus on a subset of variants, usually based on some combination of the following factors: (1) rare variants with a predicted null or deleterious effect (i.e., nonsense, frameshift, splice site variants), (2) rare variants in a single gene consistent with the observed inheritance pattern disease in the patient and family (e.g., homozygous, compound heterozygous, de novo, segregates with the affected members in a family, etc.), and (3) rare and published variants in genes that have been previously associated with the described signs and symptoms of the patient.

Although the potential for increased clinical sensitivity is a strength of the test, particularly for those patients with complex or indistinct phenotypes, to realize this potential, better tools are needed to facilitate an open and transparent dialogue between the ordering clinician and the clinical laboratory. A clinical laboratory needs a description of the patient's symptoms, any suspicions that the clinician may have regarding a genetic cause, a relevant family history and the results of any testing that has been performed, including negative or inconclusive results. All of this information is necessary for a thorough and appropriate analysis of a patient's WGS to be performed. Creating appropriate mechanisms that enable clinicians to easily, accurately, and efficiently provide such information is one of the new challenges of offering genomic (both exome and genome) sequencing. In practice, many laboratories struggle between receiving too little or nonspecific information and receiving 200 pages of medical notes and records that the genetics staff must then sift through to identify the pertinent information. There is currently an effort to enable a more quantifiable, standardized approach to collecting phenotypic information using the Human Phenotype Ontology (HPO). This is a hierarchy of phenotypic terms that serve as a link between diseases and genes. Several clinical laboratories are currently experimenting with HPO terms and other customized approaches to improve both the sensitivity and specificity of the genome (or exome) results based on the patient's phenotypic information. Doing so will be critical for improving the efficiency of interpretation, reducing the cost, and increasing the utility of WGS clinical testing.

Likewise, clinicians need to be aware that not all WGS tests have been designed to answer the same questions. In this respect, it is beneficial for clinical labs to clarify the following questions before offering a clinical WGS test:

1. Is WGS to be used as a preliminary screen or a confirmatory test? Will it aid in the diagnosis of disease or assist prognostic or management decisions after a diagnosis has already been made?
2. Is it intended to address conditions caused by inherited or somatic genetic variants?
3. What types of clinically relevant variants (e.g., SNVs, CNVs, rearrangements) are most often identified in the population being tested? How well does the technology detect these different types of variants?
   (a) Will multiple analyses or methods be combined?
   (b) What DNA quality, sequencing depth, and bioinformatic analyses are required of the assay to ensure reliable and accurate detection of the clinically relevant variants?
4. What are the technical and coverage requirements for the diseases being assessed? Are there limitations of the test? How are coverage gaps communicated?
5. Who are the ordering physicians and what level of support will they need?
   (a) Are genetic counselors available to support questions from physicians?
   (b) What marketing materials, instructions, and definitions of terms will be needed? Will supplementary educational materials be needed?
   (c) Are there ordering tools that can help guide the efficient selection of the appropriate terms (or NPL types of tools to translate and suggest based on EHR/notes)?
   (d) Are medical geneticists or molecular pathologists who perform the test available to explain test results, limitations, and other important information?
6. What are the consent and information return policies?
   (a) Who owns or has access to results, and for how long?
   (b) Are results limited to the primary clinical report, or will physicians be able to request raw data and/or variants in all analyzed genes?
   (c) Is reanalysis of the raw data an option?
   (d) What are the timeline and the rationale for reanalysis?
      – Is there new variant or gene information?
      – Is there a different question to address at a later date?

When the clinical laboratory answers these questions, it rapidly becomes clear that the same whole-genome sequence could serve to support multiple different test definitions and might require different support staff and educational materials, as well as multiple processing and reporting policies, depending on the test definition. A thorough evaluation of the laboratory, the population it serves, and the abilities and needs of both parties are critical to defining how WGS is offered.

Today, the most common use of genome and exome sequencing is for the assessment of a rare disease with a suspected genetic etiology. Most often the signs and symptoms are nonspecific, and first-tier single-gene or panel testing was inconclusive [10–12], but increasingly WGS offers the fastest possible method for diagnosis and differential evaluation [13, 14]. Inherent in this approach is the expectation that the disease is caused by variants in a single gene (sometimes called monogenic or Mendelian conditions).

There is of course also great potential in using WGS for gene discovery. Searching the genome can reveal tantalizing variants with the correct inheritance pattern, in genes encoding proteins with molecular and functional studies consistent with the patient's phenotype. In some cases, other rare variants in the same gene may have been reported in one or two patients with a similar or overlapping phenotype (e.g., see http://www.matchmakerexchange.org). However, it is important to recognize when clinical testing crosses over to clinical research. The primary intention of clinical testing is not gene discovery; however, as with microarray testing, variants may be identified in genes for which the function is not yet established, only suspected based on limited data or perhaps completely unknown. In such cases, if there is a strong suspicion that variants may explain the patient's phenotype, additional testing will be required to establish clinical validity of this finding. Ideally, clinical laboratories should have plans for providing appropriate recommendations to the ordering physicians, such as additional laboratory tests that may further clarify the patient's phenotype or options for molecular research of the variants and genes.

Analytical Considerations: Analytical and Bioinformatic Validations and Quality Control

The analytic phase of the testing begins after the blood is drawn and involves all of the steps that enable the production of a sequencing result. For autosomal genes, a sequencing result includes all nucleotides covered that match the reference sequence at a predetermined quality, all differences from a reference sequence (variants) that were detected, and the zygosity of the variants (e.g., homozygous or heterozygous). For mitochondrial, tumor, or other somatic sequencing, in addition to all the variants detected in this population, the result will include the percentage of sequencing reads with the reported variant; thus somatic sequencing is a quantitative test and, as such, will have additional regulatory requirements associated with it. In an NGS test, the analytical steps to produce this product include DNA extraction, DNA shearing and size selection, adapter ligation, library preparation, cluster generation, sequencing, alignment, variant calling, and all of the quality metrics associated with the processes at each step.

The process of DNA extraction depends on the type of sample being received, which may differ between different types of WGS tests. Diagnostic testing for Mendelian conditions is typically performed using DNA extracted from peripheral blood. Whereas panel or whole-exome NGS tests may accept other types of samples (i.e., saliva), these can be problematic for whole-genome sequencing. Unlike panel or exome testing, whole-genome sequencing does not involve a bait and capture step in which specific human DNA targets are hybridized, isolated, and enriched. Instead, for WGS all DNA that was extracted from the sample will be used to make the library. While this has several advantages, including more even sequencing coverage across all regions of the genome and a faster more efficient library preparation, it also has one distinct disadvantage—the abundance and quality of the extracted DNA must be high. For example, if DNA is extracted from saliva, it means that all the DNA in the saliva (including the banana you had at lunch) will be sequenced as well. Although nonhuman DNA will not align to the human reference sequence and contaminate the results, it will result in less human DNA available for sequencing and alignment and affect the overall depth of coverage across the genome. Ultimately, this will be more expensive for the laboratory to attain sufficient depth of coverage for high-quality variant calling. Additionally, for nursing infants, there is a risk of maternal DNA contamination if buccal or saliva samples are used. As in other tests, an evaluation of the quality and quantity of the DNA should occur prior to testing and meet all previously determined and validated quality parameters.

With the exception of the targeted capture step, all sample preparation and sequencing steps (see Fig. 18.1) are the same for all NGS tests. Before offering a test clinically, the laboratory must validate the test for specific performance metrics established in the test definition. For example, the ability to detect single nucleotide variants (SNVs) must be validated, but this does not validate the ability to detect small insertion or deletion (indel) events or larger copy number variants (CNVs). Additionally, regions that are susceptible to variable sequencing results, such as high or low GC content, should be evaluated for base-calling quality and validated for inter-sample consistency and accuracy. Validations are intended to assess the analytical sensitivity and specificity, limits of detection, and regions of the genome where variants can be consistently and accurately reportable. During the process of validation, quality metrics and filters should be established that can then be used to assist with ongoing quality control (QC) assurance.

When considering an entire genome and the overwhelming number of data points that must be considered in that evaluation, multiple tiered validation approaches may be appropriate. One method of validation is to test the sequencing performance with a "truth set" of variants in a given DNA sample. Many samples are available that have been sequenced using orthogonal technologies and contain well-characterized and clinically valid sequence variants known to be pathogenic for specific diseases (repositories such as Coriell Institute for Medical Research, The Hospital for Sick Kids, etc.). Many of these samples include parent–child pedigrees, so in addition to confirming variant detection, filters and subtraction methodologies can be tested using these known relationships. While testing "pathogenic" variants is important, the validation should not be limited to these. Ideally, a variety of different variant types (SNVs, small and large indels) within different sequence contexts (homopolymers, tandem repeats, high and low GC content) should be evaluated for test performance in order to determine the limits of variant detection and regions of the genome that may be intractable to variant calling. The evaluation should account for background conflicts that can be attributed to de novo mutations in every generation (<100/genome). The number of conflicts observed that exceed this background rate is dependent on the choice of aligner and variant caller and the settings that have been used to align reads and make genotype calls. One can choose more stringent thresholds to reduce false-positive calls or more permissive thresholds to maximize sensitivity. When making decisions about these thresholds, it is important to consider the following questions:

1. Will confirmatory testing be performed for all or some subset of variants that fall within defined quality thresholds?
2. How will confirmatory testing and/or manual review of potentially false-positive variants affect the workflow, turnaround times, and cost of the clinical test?
3. What are the risks and medical implications associated with false negatives versus false positives?

Another approach to establishing quality thresholds involves very deep sequencing of WGS and subsequent subsampling or bootstrapping; in such a case, it is recommended that multiple samples are included in the analysis, which represent various regions of the genome, with differing levels of GC content and other sequence contexts, and a wide variety of variant types with a range of complexities. Re-sampling (bootstrapping) analyses can then be used to evaluate the coverage depth and quality filters that will yield reliably high-quality sequence across all regions and variants of interest. If done across multiple regions and using multiple samples, this experiment can be very useful in establishing the confidence in specific types of calls and in assessing how they correlate with quality metrics. Additionally, with this analysis, confidence levels can be established for different types of variant calls in different genomic regions (e.g., percent GC). It should be noted that while this may help establish high confidence thresholds for most types of variants, it will not resolve alignment problems due to segmental

duplication, homopolymers, or paralogues. Validation of the WGS test should be updated in the event of any process changes, regardless of whether it is a single step (e.g., sequencing chemistry) or an entire platform.

The quality of an NGS sequence relies on both the sequencing platform itself and the methods used to analyze the resulting data. For that reason, validations must be designed to establish both the sequencing and the pipeline used for analysis. Specific methods to evaluate the bioinformatic pipeline separate from the sequencing platform can be performed using datasets that are rapidly becoming available through efforts such as the National Institute of Standards and Technology (NIST) and Genetic Testing Reference Materials Coordination Program (Get-RM). Synthetically generated data can also be used to test and validate specific challenges to variant calling algorithms.

Transformation of signals produced during NGS into genetic calls of DNA bases involves a highly complex process that utilizes sophisticated bioinformatic analyses. Generally, there are three steps in the analysis—(1) preprocessing of reads, (2) alignment, and (3) variant calling. Preprocessing involves filtering out raw sequence data that do not meet certain quality criteria. The process of alignment involves mapping of reads to the reference human genome sequence, which may be obtained from the National Center for Biotechnology Information (NCBI), University of California Santa Cruz (UCSC) Genome Browser, or Ensembl. There are many tools that employ different algorithms to align reads; each offers trade-offs on speed and accuracy [15, 16]. Mapping is complicated by the fact that the reference genome is incomplete and because humans have some regions that may be individually variable. Because of this, approximately 5–10% of reads will fail to be aligned. Mapping quality is measured and the confidence score assigned with each read placement. One of the community-accepted standards to represent alignments is in Binary Alignment Map (BAM) file format [16, 17], which captures the abovementioned data, allows efficient compression, and enables random access of reads (when sorted) that align to a particular segment of the genome. Once the alignment procedure is complete, the BAM file serves as input to the next step in the bioinformatic pipeline—variant calling—where genetic variants are identified. Depending on the intended use of the test, a variety of variant calling tools, each one specializing in detecting small SNVs and indels or large genomic structural variants (SVs) and CNVs, might be employed. In some cases, several tools might be used in conjunction to identify as many types of variants as possible.

Variant calling algorithms are typically based on two main paradigms—the first one involves relying on base counting and allelic fraction to distinguish between a heterozygous and homozygous genotype call and the second involves probabilistic methods (Bayes' theorem) to calculate posterior probability given the observed read data and a genomic prior probability [18]. The latter method accounts for noise in the data and helps provide a measure of statistical uncertainty associated with each genotype call in the form of a score. The score is usually a representation of the confidence in the genotype call. Although many algorithms report on variant positions, it is important to consider that the reference genome may contain a non-wild-type allele and to monitor the quality of the positions called as homozygous to the reference; no calls and poor quality homozygous reference calls should be considered in the downstream interpretation effort. These algorithms are typically tuned to detect a heterozygous position in a diploid setting; that is, the expectation is to detect variants that are present at approximately 50% in a sample. However, mosaicism, heteroplasmy, and duplication/deletion events may occur, which then throw the allele balances such that a non-diploid event may be called as heterozygous or missed depending on the frequency of the minor allele in the reads that were sampled and the tuning of the algorithm. In cases where detection of these alternative allele frequencies is desired, additional validations should be performed to establish the level at which the frequency of the alternative allele can be detected and the depth of coverage at which the region must be sampled in order to achieve that level of sensitivity. Typically, serial dilutions are used to evaluate and set these types of thresholds.

During the validation process, a clinical laboratory that is implementing WGS should be aware of and test for potential artifacts in processing. For example, the reference genome is not necessarily wild type. Therefore, if a laboratory is only considering the variants that are called against the reference, such mutations may be missed in an individual who also carries this genotype. Assessment of reference allele frequency based on the 1000 Genomes Project data shows that there are approximately 63,000 positions in the genome where the reference genome carries an allele that is present in populations at less than 1% allele frequency. Additionally, for regions of the genome such as human leukocyte antigen (HLA) locus, there is not necessarily a "wild type" per se, and additional information such as phasing may be necessary to confidently evaluate the variants found. While similar challenges exist for many types of clinical tests, laboratories should be aware of and prepared to manage such issues.

One challenge to the implementation of WGS in a clinical laboratory is that the analytical validity may not be the same for all regions of the genome, nor is it for all types of variants that may be of interest. While this is also true of other types of genetic tests, the scale of genome sequencing makes effective communication of these types of challenges more difficult. The specific weaknesses and strengths of WGS must be considered when launching a test and then communicated effectively and evaluated, potentially on a case-by-case basis, for appropriateness given the needs of the test in that

specific situation. Regions of particular concern should be noted and either additional confirmatory testing performed or recommended in the report.

Post-analytical Considerations: Interpretation and Reporting

After sequence reads have been aligned, variants called, and quality filters applied, the post-analytic process can begin. Annotation, the process of identifying basic information regarding positions, such as transcript used, coding position in the gene, and consequence of variant (e.g., amino acid change, splice site, etc.), is sometimes considered an analytic phase activity and sometimes considered a post-analytical activity. Regardless, specifics that must be considered at this phase include references used for alignment, transcripts chosen for downstream reporting and interpretation, and which version of this information is being used. This is critical, as downstream searches for literature associated with the variant, frequencies of the variant, etc. are linked to this information, and depending on which reference/transcript is chosen, additional adjustments may be required to sync the information appropriately. This represents the division between the technical and professional components of the test. The technical components focus on analytical validity, whereas the professional component focuses on clinical validity. Regulatory standards address both analytical and clinical validities; however, additional guidelines and best practice standards are also issued by medical agencies and represent an additional set of professional guidelines. As with the previous phases, the type of testing being performed and its intended use greatly impacts how the post-analytic process is designed.

A genome is approximately 3.1 billion data points, and approximately 88 million variants have been identified across 1000 genomes sequenced (contains around 3–4 million variable positions, including on average 9600 amino acid changing positions and 73 premature termination positions (internal data, [19])). Given such a large amount of information, a thoughtful plan must exist for how to identify and evaluate the information that is most likely to be relevant to the particular individual and informative to the clinical questions being asked.

After high-quality variant calls have been made and the regions of the genome for which highly confident variants can be called have been defined, the clinical significance of the variants is assessed. This process can be divided into three steps:

1. Annotation, which includes naming and assigning a predicted effect (nonsense, missense, synonymous, etc.) to the variant
2. Interpretation, in which all available data about a variant is evaluated and synthesized into a clinical classification (pathogenic or benign)
3. Reporting, in which the variant classification, the evidence supporting the classification, and the clinical implications of the findings are communicated back to the ordering physician

Historically, the assessment of clinical validity or the strength of the relationship between a variant (or call) and a disease has been recommended but not required in genetic reporting. This has recently changed, and the College of American Pathologists (CAP) guidelines now address how clinical laboratories should support the assessment of the clinical implications of a call. For instance, in cases with a single gene, this typically consists of an expert or panel of experts within the laboratory who evaluate each variant based on peer-reviewed publications and other factual evidence and categorize them for inclusion in the report. This process has become significantly more sophisticated in recent years, as there are now several databases and online tools that can aid in the assessment of clinical implications of variants. Additionally, the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) has issued a set of guidelines for interpretation of sequence variants (ISV) that details the various types of information that should be gathered and how the types of information should be evaluated and weighed in classifying the clinical implications of the variant [20].

The first step in this process is to annotate each variant with curated information from a variety of databases. When implementing WGS in a laboratory, using automated tools, such as the Ensembl Variant Effect Predictor [21] or SnpEff [22], for annotation becomes necessary to support the abundance of variants that are detected and require interpretation. Recommendations for types of information that should be gathered and annotated can be found in official publications by CAP and ACMG; these include but are not necessarily limited to (1) the genome build (e.g., GRCh38); (2) variant location within the context of annotated genes and transcripts (e.g., promoter, intron, exon, UTRs); (3) variant type (e.g., single nucleotide, multi-nucleotide, indel) and predicted effect (e.g., splicing, missense, synonymous, nonsense, frameshift); (4) a concise and consistent notation, such as those recommended by the Human Genome Variation Society (HGVS) [23, 24]; and (5) conservation scores for the nucleotide and amino acid, if available. When implementing the annotation process, it is very important to assess and validate the annotation software suites that will be used and confirm that the variant is being searched correctly and the information gathered is being downloaded and displayed properly.

Having annotated the positions, interpretation of the variants for reporting can begin. Considering each individual genome can have more than three million variants, it is first beneficial to filter or exclude from further analysis all variants above a given minor allele frequency (MAF) in the general population. Currently, standard practice is to filter all variant calls that are present in the general population greater than 1–5%, although it is very important to use a threshold that is consistent with the intended use of the WGS test. For example, germline variants known to cause a rare Mendelian condition will be much less frequent in the general population than germline or somatic variants that modulate a patient's response to medications. Therefore, a WGS clinical test designed to aid in the diagnosis of rare Mendelian conditions can use a much lower MAF filter cutoff [25] than a test designed to identify all medically relevant variants in a presumably healthy population [26]. While setting the filter too high for a diagnostic test shouldn't greatly impact the expected yield, it will result in more manual interpretation of benign variants and greatly increase the cost of the test.

Even when excluding all variant calls above a relatively stringent MAF (e.g., greater than 1%), there will still be many thousands of rare variants for interpretation and clinical classification. Most of which will not be relevant to the patient's condition. The evaluation of evidence for clinical implications of variants is a critical process that is guided by both professional expertise [20, 27, 28] and a pipeline that can support such evaluations (Fig. 18.2). It is important to consider both clinical and biological characteristics of a variant. Clinical characteristics include whether the variant is enriched in individuals with a specific disease, while biologic characteristics include the variant type and its predicted or demonstrated effect on gene function.

When evaluating and weighing the enrichment of a variant in a patient population, it is important to consider the severity, age of onset, penetrance, and reported incidence of the associated disease. For example, if a disease has a prevalence of 1/100,000 and is autosomal recessive, then, using Hardy–Weinberg principles, any single variant with a frequency higher than 0.3% is unlikely to cause that disease, unless penetrance is known to be significantly reduced or the disease is greatly underdiagnosed due to relatively mild symptoms and/or late age of onset. Careful searches of the published literature and appropriate patient databases are important for identifying high-frequency alleles with increased risk for common diseases, pathogenic variants with reduced penetrance, and founder mutations, as well as aggregating the clinical evidence for classifying rare variants as pathogenic or benign. It is important to remember that these databases may or may not be updated regularly and may or may not be complete with regard to the actual publications that exist. Furthermore, many variants have been characterized in databases based on old information, and therefore, if a database reports a variant as pathogenic or uncertain significance, it is important that the clinical laboratory performs an updated and independent assessment to ensure that this information is still valid.

Although observations of a variant in affected individuals provide the most direct link between a variant and disease, for many genes, the pathogenicity of a novel variant can be inferred based on an abundance of evidence demonstrating that loss-of-function (LoF) is conclusively associated with disease. For example, a predicted null variant (i.e., nonsense, frameshift, gene deletion) in a well-studied gene with a proven LoF mechanism of disease (e.g., CFTR and cystic fibrosis or BRCA1 and hereditary breast and ovarian cancer) is very likely disease causing, even if the variant has not been previously reported in affected individuals. As a result, it is extremely beneficial, when analyzing rare WGS variants, to have a gene curation step prior to variant interpretation. This step should assess the relative strength of the gene–disease relationship (e.g., strong, suggested, unknown, negative) based on the published literature for germline variants in that gene [29, 30]. For the purpose of variant interpretation, it is extremely useful to also establish the molecular mechanism of disease if sufficient evidence exists [31]. Is the disease caused by loss of gene function or a gain of gene function? Are null-type variants (nonsense, frameshift, splice site) associated with disease, or are only missense variants reported? If mutations known to cause disease are exclusively gain-of-function, then a variant that creates a premature termination codon (PTC) is less likely to cause disease and should not be considered pathogenic a priori. This is a complex set of considerations and requires knowledge of clinical and technical genetics.

These data can be difficult to gather and, due to the evolving nature of clinical genetics, often inaccurate or incomplete at the time of curation. It is recommended that a minimal set of fields (e.g., molecular mechanism, inheritance pattern, penetrance, age of onset, and phenotypic severity) be stored in a database for future use, along with specifics about when and where the details were collected. This way the data can be reinvestigated at regular intervals or when it is suspected to be incorrect or outdated. Importantly, as the spectrum of phenotypes associated with a disease is rapidly growing, regular evaluation for any new information that could change the understanding of the gene-disease relationship should be part of the process.

Peer-reviewed literature is an important source and plays a central role in the curation of genes and interpretation of variants. Case reports in affected individuals and family studies for rare variants help to establish a relationship between these variants and disease, while large case-control studies may demonstrate an increased risk of disease for more common variants. In addition, in vitro or in vivo experiments testing the functional consequence of specific variants can support their

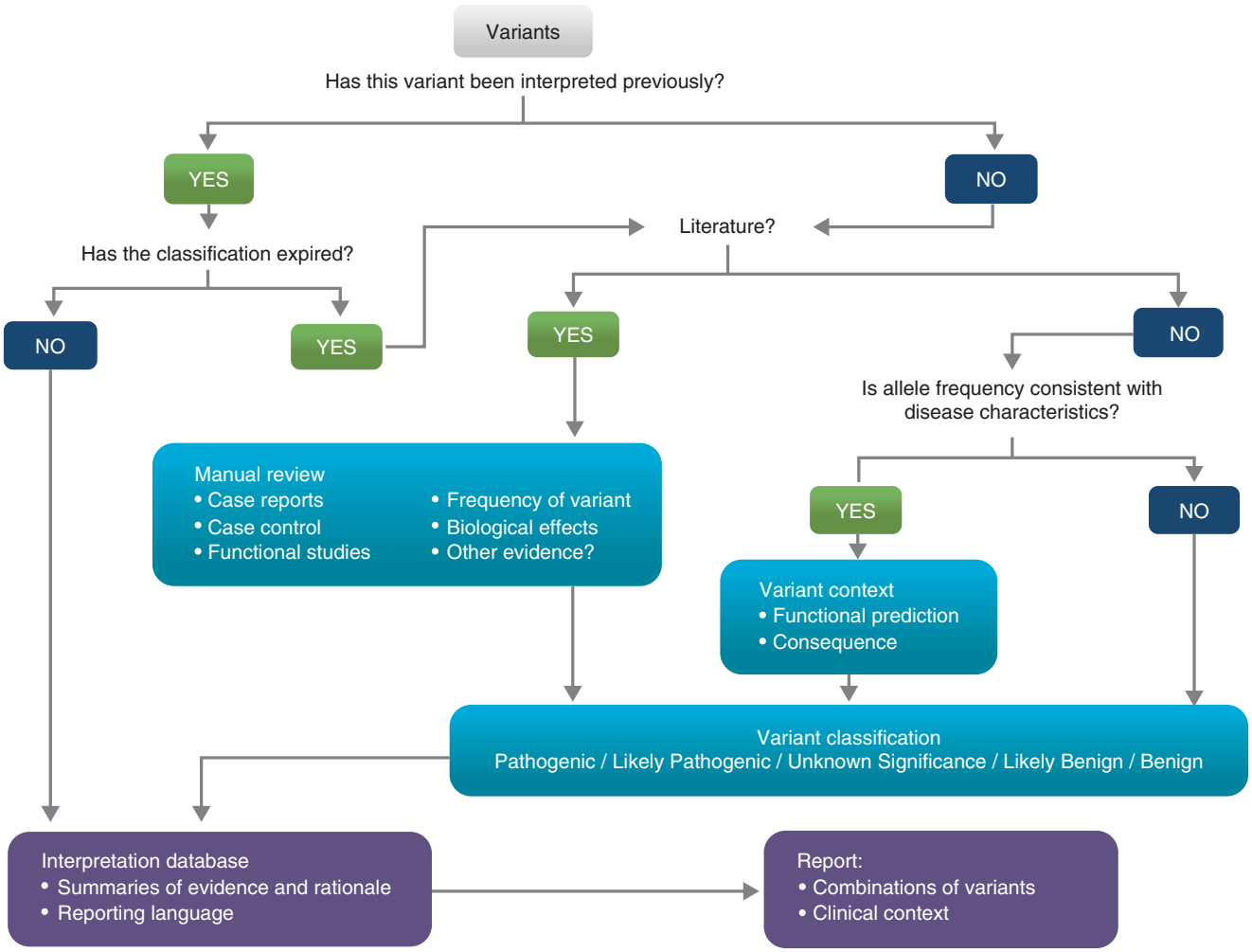


Fig. 18.2 Decision tree for the evaluation of clinical implications associated with sequencing calls. The process shown is the one that Invitae uses for the evaluation of evidence that links a particular allele to a clinical condition

role in disease, if the reported consequence matches the established mechanism of disease. However, very often an initial study will be published that reports a compelling association with disease, but further observations of a variant in unaffected individuals or experiments demonstrating little or no effect on the protein function conflict with the original assertion. It is therefore imperative to review the full corpus of published literature associated with each variant and critically evaluate the confidence and reliability of the data. Some observations and experiments are more robust than others and should be weighted more heavily [31–33]. Clinical laboratories often rely on well-trained, clinically specialized MDs, PhDs, and genetic counselors to read papers, review all the evidence, and incorporate these findings into a final classification and reporting language. Individuals must be able to read through a paper, evaluate the strength of evidence regardless of author's conclusions, and document this. This currently requires these professionals spending a significant amount of time sifting through that information. Every clinical laboratory faced with large numbers of variants to be assessed in WGS will be challenged to hire a qualified staff large enough to support such efforts. Altogether, this process is daunting task for WGS if automated approaches are not utilized. The extent to which natural language processing (NLP) and other algorithms can be used in the evaluation of variants is still highly debated; however, it is clear that software tools are invaluable for aggregating and collating the information. For example, thanks to well-curated variant databases like gnomAD [33], automated annotation tools like Ensembl's VEP [34] or SnpEff [22], and meta-predictors like REVL [35, 36] or VEST [22], the population frequency, variant type, and predicted consequence on gene function are currently available for nearly all variant calls. Once the appropriate weighting and rules for these evidence types have been carefully considered [20], software tools can be designed to aggregate the data and apply these evidence types consistently with only limited manual intervention. Considering most variants needing interpretation are rare, very few will have relevant publications supporting a

definitive clinical classification. As a result, NLP tools that accurately identify all variants for which published literature exists can greatly reduce the total number of variants that need to be reviewed by a trained professional. Even so, simply reading the papers is still a challenge and will become ever more burdensome as the publication rate increases in parallel with gene and variant discovery. For variants that have been published, NLP tools that highlight a variant in the paper and correlate it with patient information in tables and/or supplemental information has the potential to help readers focus on the most relevant information more quickly, making reading papers for interpretation more efficient and effective. For variants both published and unpublished, shared patient databases that link genetic variants to specific phenotypic information will greatly aid in the accuracy and efficiency of interpretation. Finally, machine learning algorithms that use the patient and population databases to identify regions of a gene for which sequence variation is enriched in patients or key residues in protein domains that are intolerant to normal variation will help with the interpretation of novel variants for which published observations do not yet exist. Development and maintenance of these databases and automated tools will benefit both clinical laboratories and patients and are likely to be an extremely interesting and active area of exploration in the future.

Designing the Post-analytic Process for Monogenic Conditions

Given the daunting number of variants detected within a genome, approaches must be developed to apply filters so that only variants of potential relevance are identified and evaluated. Both biological and clinical features can be used to help refine the search for genomic information that is of clinical relevance for the patient. In cases where parental samples are available, a geneticist or genetic counselor should begin with taking a family history to identify whether the current condition is most likely to be autosomal dominant (possibly with reduced penetrance), autosomal recessive or to have arisen de novo. If both parental samples can be sequenced along with the affected individual, then subtractions can be performed across the entire genome in order to evaluate variants that meet the biological hypothesis of the following conditions:

- Autosomal recessive, in which one would expect to find at least two variants within a single gene. Filters should be set up to detect the expectation of two variants in one gene, with one inherited from each parent or one variant inherited from a single parent along with a de novo variant. To perform this search, all three samples are sequenced, and the child's variants are filtered to match the expectation of two variants in a gene, one from each parent. This can significantly reduce the number of variants that must be considered. After this subset of gene/variants is identified, the genes and specific variants can be filtered further. For example, common variants with allele frequencies above 5% might be excluded from consideration; when making such decisions, patient ethnicity, prevalence of the condition in that ethnic group, penetrance, and modes of inheritance should be considered because sometimes common variants are pathogenic. Through the use of these types of filters, the resulting subset of variants should be of a tractable number that can be individually evaluated by qualified clinical laboratory staff.
- Autosomal dominant, in which one would expect to find only a single causative variant within a gene. This model is more difficult because there are significantly more possible variants to evaluate; however, if there is a family history (even with reduced penetrance), one can subtract variants from the unaffected side of the family and look for matches to the presumed carrier parent (who may or may not be affected). Again, additional filters to remove high-frequency variants can be applied, and the resulting variants can be considered.
- De novo, in which the causative variant arose within the proband. In this case, all variants inherited from both parents can be subtracted and only those variants that arose in the affected individual can be considered.

For all of the above methods, the process may also include the evaluation of the resulting variant set in the context of the clinical phenotype or a defined set of genes that are set out by the physician/medical geneticist. This could take the form of a filtering tool that enables the list of variants to those in genes known to be associated with the phenotype or simply as part of the context that the clinical laboratory staff uses during the evaluation process.

These types of approaches based on filtering by modes of inheritance and parental genomes are currently the most popular way of WGS testing. However, this does require the added expense of sequencing multiple genomes in order to identify potentially causative variant(s). Sometimes, the parental samples may not be available, or the additional cost may be prohibitive. Regardless of whether parental samples are available, additional filters should be set to identify regions of homozygosity that may indicate a disomy or copy number event. Such filters will also detect consanguinity, which indicates a higher likelihood of autosomal recessive and possibly multiple genetic conditions.

In addition to, or in the absence of, an assessment of genetic inheritance patterns between parents and child, a clinical phenotype approach can be used on its own. For this approach, one requires access to thorough clinical phenotype information, such as all presenting features or previous testing results, including negative results (e.g., "no increased

creatine kinase"). This information can then be used to search through phenotype-to-gene information that is available in various databases (e.g., Online Mendelian Inheritance in Man (OMIM) [37]), or accessible within phenotype software tools, to identify and rank order genes that might be involved with the symptoms affecting the proband. The Human Phenotype Ontology (HPO) and the Monarch Initiative have established useful mechanisms for linking phenotypes with cognate genes, and together they provide a basis for integrating phenotype information into variant interpretation. Specifically, the Monarch Initiative has developed "annotation sufficiency" (AS) scores to provide a measure of the depth of available phenotypic features for a given disorder, whereas the HPO-driven Phenomizer tool uses available phenotype information to produce confidence scores on the likelihood that the information represents a specific annotated genetic disorder. Natural language processing tools may also be employed to recognize synonyms of specific phenotypes and broaden the search. Then, all variants within that subset of genes that have a database association with the phenotypes can be included in the subsequent interpretation (with additional filters applied to remove variants that are too common to be likely involved with disease). This approach is highly sensitive but has several challenges associated with it:

1. Accurate and thorough phenotyping by clinicians is difficult and time-consuming. Transcribing this information into an order for testing is onerous, and often, the laboratory may receive incomplete information.
2. Both the tools being developed to match phenotypes with genes are incomplete, and the phenotypes described for different genes are variable in both completeness and accuracy.
3. The emphasis on sensitivity means that many genes that are not actually relevant to the patient's phenotype will be brought through for interpretation. This results in increased workload on the laboratory, increased risk for incidental findings, and brings in a level of subjectivity in how clinical laboratories choose to report on genes that have different levels of phenotype matches to patient's described symptoms.
4. Genes with varying degrees of clinical validity will be identified, and the clinical laboratory must perform a careful curation of the gene/disease relationship to establish whether this information is preliminary and experimental or meets criteria for being medically reportable and actionable. Some ordering physicians will be willing and able to perform the follow-up studies associated with receiving experimental or preliminary information, but others will not have that ability. Policies about which genes will be included in the clinical report, and the criteria used, should be available.

Each of these approaches is labor-intensive and requires a clinical laboratory staff trained in the evaluation of genetic disease, preferably formally trained and certified through the American Board of Medical Genetics (ABMG), the American Board of Pathology (ABP), or the American Board of Genetic Counseling (ABGC). In the case where the first assessment is found inconclusive, multiple different approaches might need to be performed. The clinical laboratory team performing the filtering and variant assessment should expect to spend several hours per genome evaluating the resulting variants, and this type of effort should be budgeted for in the planning for this type of testing.

Recent studies strongly suggest that for diagnostic odyssey cases, genome or exome sequencing approaches result in "1/3 the cost of diagnosis" [38], "30% higher diagnostic rate without increased costs" [39], with results as high as "68% of families diagnosed and 44% received a change in treatment" [40]. However, most of the higher claims include established pathogenic variants as well as variants of uncertain significance in genes that seem compelling but are actually candidates and require additional follow-up for diagnosis. The cost and time investment of the WGS test must be considered against the potential costs and consequences to the affected individuals of undiagnosed genetic disease.

Designing the Post-analytic Process for Oncology Applications

Another possible use for WGS is in the assessment of the molecular profile of tumors in patients who have already been diagnosed with cancer. This type of testing can be useful in determining candidate therapeutic treatments when standard of care approaches have been exhausted. In these cases, the tumor sample and a normal sample of DNA are procured from the patient. Variants found in the normal sample are subtracted from the tumor sample, so that only variants that have arisen somatically can be identified. In a somewhat unique manner, most laboratories that perform clinical oncological testing will have a tumor board associated with the laboratory that reviews the findings and contributes to the interpretation. With results from this type of testing, the clinical laboratory and the associated tumor board may be able to identify the most promising chemotherapeutic options based on the presentation of the molecular profile. Of particular interest for oncological applications are large chromosomal rearrangements, insertion or deletion events, and copy number variants that can be identified. Anecdotal reports of these approaches have been very encouraging [41, 42].

The analyses required for detection of tumor variants are significantly more complex than those described above for

Mendelian conditions. In the analytic phase, special consideration about the sample type should be given based on the type of cancer being tested. For example, blood samples in leukemic patients would likely be more representative of the tumor rather than the normal signal and the type of tissue most appropriate for the normal sample should be thoughtful considered. Beyond that, the analytic process for the normal sample is essentially the same as what would be done for the monogenic conditions (described above). Tumor samples, however, require special additional processing and handling. To begin with, the DNA isolated from a tumor may be from fresh, from fresh frozen, or, more commonly, from formalin-fixed paraffin-embedded (FFPE) tissue. The different tissues may require significantly different extraction techniques and evaluations of the quality of that DNA. Laboratories must evaluate their abilities to support each of these extraction techniques and subsequent evaluations of appropriate DNA quality and quantity. The downstream informatics processing of tumor samples also has some unique requirements. Tumor samples are often contaminated with some amount of normal cells. Quantifying this fraction is difficult and imprecise and has implications for downstream informatics processing that must be incorporated into the process. Additionally, NGS methods sequence individual molecules separately, and therefore, in a diploid situation, a heterozygote would be expected to have approximately half of the sequences showing one variant and half with the other. The algorithms that have been developed for NGS typically have been developed to optimize for this scenario, and general recommendations regarding the required number of independent sampling events are also usually made with this expectation. However, a tumor does not represent a diploid scenario. Therefore, one must establish at what frequency one wishes to detect somatic variants; this might be 20%, 5%, 1%, or less. Depending on what the laboratory decides, sequencing must be done to a depth that ensures likely detection of variants. The depth required to attain the required sensitivity can be estimated using a sampling statistic:

$$P(x,p,N)=\sum_{N-X}^{K=X}\frac{N!}{(X!)(N-X)!}p^{X}q^{(N-X)}$$

Empirical validation will be discussed in the validation section. However, in addition to different processing requirements, the bioinformatic algorithms used to detect variants may also need to be optimized, and additional or alternative algorithms may be needed. In some cases, different algorithms may be called for to detect different types of variants, for example, copy number or structural variants (chromosomal rearrangements). Laboratories planning to launch tumor–normal WGS analyses should be prepared to evaluate these needs and plan appropriately for implementation. This can be an arduous process, and a team may be needed to identify the requirements and evaluate the appropriate set of tools for implementation.

Cancer is not the only disease type that is associated with the occurrence of somatic variants; certain genetic conditions (often associated with hemihypertrophy or skin lesions and increased likelihood of developing cancer later in life) may also demonstrate these and be of interest for a clinical molecular lab. Additionally, in testing for mitochondrial diseases, it may be critical to enable detection of mitochondrial heteroplasmy. All of these applications involve the challenges described above for tumor scenarios and may require the same or similar planning and evaluations before implementation.

Designing the Post-analytic Process for Screening for Fetal Aneuploidies

WGS can also be used for various forms of screening tests. Screening involves identifying genetic variants with potential clinical implications, typically before there is any clinical presentation, and often that would be confirmed by additional testing before any medical action is taken. Currently, the most common and popular screen involving WGS is for aneuploidy in prenatal settings. Commonly called noninvasive prenatal screening or testing (NIPS or NIPT), this involves performing deep sequencing of either targeted regions or the whole genome in an effort to identify chromosomal regions that are present at non-diploid copy numbers. These kinds of screens have only been available in the last few years, but their sensitivity and specificity is greatly improved over serum screening paradigms and therefore is being rapidly adopted, particularly for high-risk pregnancies. These tests are performed from a maternal blood sample where the DNA for the testing is fetal DNA circulating in the maternal blood stream and thus considered noninvasive from the perspective of the fetus. Because this testing requires isolation and enhancement of the fetal DNA, specific planning should be given to additional techniques that might be necessary for implementation, such as for DNA isolation and quality evaluation to ensure that the appropriate quality and quantity of DNA are present to perform testing. This test also requires quantification of genomic regions that are present at non-diploid quantities and the subsequent analyses.

Designing the Post-analytic Process for Non-symptomatic Evaluations

Finally, WGS can also be used for more traditional screening of genetic variants for which individuals may be carriers or at risk. While this type of testing is currently more likely to

be performed using targeted panels, it is possible to employ WGS for this purpose. The post-analytic process for this type of testing is heavily dependent on the test definition provided in the pre-analytic phase. Typically, this would have identified a set of genes that would be included in the test, and this set would have established clinical utility of testing for a specified set of diseases. In the case of WGS, this can be a many-to-many relationship where there may be many genes tested that are providing information about predisposition or carrier status for one disease, but also any one gene could have multiple diseases clinically associated with it. The test definition would also define the regions within those genes that are included in the test (e.g., exonic regions, parts of intronic regions directly adjacent to the exons). Therefore, the set of variants requiring interpretation from the analytical stage would be filtered to those included in the established test definition. Most laboratories performing this type of testing restrict reporting to those variants assessed as clinically significant (e.g., pathogenic or likely pathogenic); however, ancillary documentation of all assessed variants and their classifications is included in some cases.

Report Organization

Once the clinical implications of a particular individual's variants have been decided, the information must be put into the clinical context for which the test was ordered. Like with exome testing, genome testing results are typically organized into the following categories:

1. Definitive or positive finding for some or all of the clinical features described. Keep in mind that recent reports suggest up to 5% of patients may receive more than one genetic diagnosis.
2. Findings that are inconclusive or have limited overlap with the clinical features. These findings may require additional evaluation or may be judged as not likely to be important by the ordering clinician. They include such findings as:
   (a) A variant of uncertain significance (VUS) for a dominant condition. In such cases as this, it may be helpful to provide information about additional types of testing that might resolve the clinical status of the variant, such as biochemical testing or an MRI.
   (b) A pathogenic, likely pathogenic or VUS in a recessive condition that otherwise matches the phenotype strongly. While this may simply be a carrier status, it may also indicate that an additional variant not detected by this test (such as methylation) could be involved, and expression testing may be appropriate. Additionally, it is important to keep in mind that carriers of a single pathogenic variant are being recognized as at increased risk for certain milder or later onset conditions (e.g., sickle cell anemia, cystic fibrosis). Thus, while not causing disease, a carrier status may be relevant for contributing to the severity or spectrum of symptoms in the patient.
   (c) A VUS in a gene that has some preliminary association with disease. These situations may be important for periodic reanalysis as new scientific and clinical information emerges or possible participation in research studies or programs such as the Matchmaker Exchange.
3. Incidental findings, which are findings that were identified for evaluation based on overlap with phenotype or inheritance pattern but, upon further evaluation, are judged to be not likely to explain the patient's primary presentation. Within this set of findings, individuals can have findings that have medical implications for the patient or family members or for which the patient is simply a carrier. Also within this category are conditions that are often purposely excluded from many analyses; these are the non-actionable, adult onset progressive neurological conditions such as Huntington's disease or early onset familial Alzheimer's disease.

Reports must be flexible enough to enable the benefit of a personalized survey of the genome but standardized enough to enable clear communication of results. A searchable electronic report might be the best solution; this could provide links to disease descriptions and additional evidence that practitioners could then have access to as needed. The goal is to provide a succinct answer to the major question of the moment but also to enable both the physician and patient to benefit from the additional information that may be present and of concern. Organization of the report so that the entirety of the information is available should individuals want it, but does not result in 30 pages of distraction is the challenge. Typically, the first page will address this by specifying:

1. The clinical indications for the testing and what kind of testing was performed
2. A succinct statement regarding whether a plausible genetic explanation for the phenotype(s) described was found and, if only for a subset of phenotypes, which ones
3. Any relevant but inconclusive results
4. Any medically actionable incidental findings
5. Statements regarding methods and limitations, with reference to where additional information may be accessed, including:
   (a) The standards that a laboratory uses in order to make calls and statements about the analytical sensitivity and specificity of the calls
   (b) How laboratories classify variants into the standard bins of pathogenic, likely pathogenic, variants of

uncertain significance (VUS), likely benign, benign, or others, as well as how they curate gene/disease relationships
(c) How much confidence a practitioner should have in that call
(d) What the weaknesses of the test are, and any recommendations regarding additional testing that could supplement these weaknesses

Appendices can be offered to provide information such as all variants detected, clinical classifications, genes and their coverage, etc. These appendices are typically only of interest for individuals pursuing additional research or if a reanalysis is being performed. Particularly when reviewing results for a patient consultation, this additional information is unwelcome and confusing; however, many individuals do want access to it.

Communication challenges between those ordering genetic testing and the clinical laboratory providing the testing are not unique to whole-genome sequencing, but the very name "whole genome" implies to many people an evaluation of the whole genome. As much of the whole genome is not yet biologically or clinically understood, there is often some confusion among those ordering or receiving these results about what was evaluated and whether additional evaluations might be needed in the future. This presents a significant risk, because the rate of learning in clinical genetics over the last decade is unprecedented and therefore reanalysis of the sequencing results may result in very different clinical reports over time. Recent anecdotal reports suggest that 20% of individuals who originally received an inconclusive result benefited from reanalysis within 5 years. Ensuring availability of clinical reanalysis of genomic data is an ongoing challenge for laboratories and clinicians. Patients move, age out, or otherwise change their physician care frequently so that even if the clinician is aware of the importance of reanalysis, the patient may not realize or easily have access to it. Patients must therefore be aware of the laboratories where the testing was performed and be informed of the possibility of reanalysis. Many clinical laboratories do not yet have clear programs in place for how often reanalysis of genomic data will be performed, whether it be initiated by the physician, the patient, or the laboratory and if there is a charge associated with the reanalysis. Since the cost of interpretation and reporting is substantial, but this reanalysis is not covered by insurance, it is not clear how scalable or sustainable this practice will be until additional policies are developed.

Communication tools might be readily located on clinical laboratory web sites, where quick, short podcast type communications specific to particular activities or questions might provide both doctors and genetic counselors with information that can significantly increase the power and confidence they have when using a test. Patient-specific pages that help patients navigate their questions may also be valuable.

An ongoing challenge will still be the large number of variants about which people are uncertain. While a large number of VUS is a point of concern, this is not new to the field. The International Standards for Cytogenomic Arrays Consortium (http://www.iscaconsortium.org) [43] has demonstrated approaches to dealing with the large number of novel and uncertain variants that are detected in individuals when genomic evaluations are standardly performed. In less than a decade, the cytogenetics community has made huge strides in understanding the nature and degree of variation at the cytogenetic level. Similar approaches could be used in the field of sequencing to better understand the nature of human genetic variation, which will aid significantly in improving and refining interpretation in the future. Additionally, genomic types of approaches to understanding patterns and distributions of genetic variation across the genome as well as specific gene/molecular implications may become useful for establishing the prior probabilities and help to resolve some of this VUS. Meanwhile, clinical laboratories can make every effort to communicate a priori that this is an anticipated outcome of these tests and help to prepare physicians and genetic counselors for managing the information.

Communication and Support

Once a test has been defined and the performance specifications and abilities established, it is critical to develop support materials. The laboratory should also be staffed with trained genetic support specialists. These specialists should be available to help physicians decide if WGS is the best test for the presenting situation and also to help plan for alternative or supplemental testing that might be necessary. It is of particular importance, but also particularly challenging, to communicate this when the very title "whole-genome sequencing" might imply all things to everyone. It is helpful to provide information through a web site that can help individuals evaluate what the test supports and what it does not.

Depending on the breadth of WGS services that a laboratory intends to offer, it may be helpful to develop an overview section that clarifies which tests offer what and are likely to be most appropriate. As information such as analytical validity, limitations of detection, and reportable regions need to be included in test definitions, and because these will be variable depending on the application of WGS, it is likely that it will be necessary to create multiple test definitions and descriptions. Including general educational materials will help physicians and patients navigate the options and choose most appropriately. Importantly, information should be readily available to help physicians understand the limits of detection, such as an ability to detect

variants present in the sample at, for example, 10% but not 5% in tumor samples, or the ability to detect deletions within certain size ranges. Laboratories should be prepared to monitor and track their capabilities to make calls of any type throughout the genome. As tests are ordered, laboratory staff evaluate the test requisitions and evaluate the laboratory's ability to support the request. If there are concerns about whether WGS is appropriate for the sample being ordered, the laboratory should contact the physician and discuss the options before the testing is initiated.

Genetic counseling is a best practice recommendation for genetic tests in which the results may have direct medical indications for immediate family members or in which the results might be predictive. WGS produces information that meets those criteria, not only for the specific indication of the testing but also for secondary findings. The ACMG has issued a series of recommendations for clinical genomic testing, counseling, and consent [44]. The ACMG has stated that genome or exome sequencing is appropriate in a series of circumstances that include strong reason to suspect a genetic etiology, symptoms associated with multiple genetic conditions for which simultaneous evaluation of multiple genes can be practical, inconclusive previous tests, and, in special cases, prenatal diagnosis. WGS is not advised at this time for prenatal or newborn screening. The recommendations specifically advise that the following elements be addressed in counseling and consent sessions: (1) pretest counseling including written documentation; (2) discussion of potential for incidental findings; (3) discussion of expected outcomes as well as incidental findings to be returned to physician; (4) potential benefits, risks, and limitations of testing and if there are alternatives; (5) distinction between clinical testing and research; (6) potential for results to be identifiable in databases; and (7) policies for updating information. It is also recommended that such testing only be performed on minors in cases where the testing can lead to diagnosis for conditions in which interventions might be possible and under institutional review board (IRB)-approved research. Additionally, the ACMG has recommended that everyone who has access to WGS, regardless of indications, should have results reported for a set of 56 conditions. These conditions represent highly penetrant genetic conditions for which there are potentially life-saving interventions available. Although these recommendations have been controversial, it is indicative of the medical community's rapid adoption and preparations to manage this information in regular clinical practice.

After WGS analysis and interpretation has been performed, additional communication with the ordering physician is likely to be necessary. While inconclusive test results are not uncommon for physicians, findings may require additional communication, particularly with regard to the management or further testing of VUS.

Infrastructure Considerations

After identifying what the WGS test will be used for, the clinical laboratory should consider the current infrastructure and any possible additional needs that would require additional build out. Depending on what resources and infrastructure a laboratory has, an assessment of necessary components includes the following:

- Facility
  - NGS sequencers are not usually very bulky, but they require space that is stable and climate controlled and has both power and Internet support. Specific requirements include uninterruptible power supply (UPS) and e-power setup, with heat, ventilation, and air conditioning (HVAC), temperature and humidity control at around 68–72 °F, and 70% relative humidity. Laboratories are required to practice space separation between pre- and post-amplification activities and ideally would have negative pressure control or a pressure-controlled hood on rooms that could have contamination. Additional safety precautions may also be necessary depending on specific requirements.
- Staff
  - NGS is considered to be high complexity testing and involves many steps. A well-trained staff is critical for this. Typically, a staff to support WGS will require people with expertise in high complexity molecular assays, genetics analyses, bioinformatics, genetic counseling, and medical genetics.
- Workflow process
  - WGS may be among the easier of the NGS assays to perform in that there are no capture or amplification steps (Fig. 18.1). Nonetheless, there are still several manual steps required and each of these can potentially introduce a contaminant or sample swap. In order to avoid such complications, a good workflow process and ideally a laboratory information management system (LIMS) to track and document a sample's process through the assay steps should be implemented. Assessment of steps in the process that can be error-prone is critical to designing a workflow in the laboratory that is robust, and consideration of appropriate controls, performance metrics, and tracking systems is prudent. In particular, positive sample controls are recommended because pre-analytical sample swapping is one of the most common errors introduced into clinical testing.
- Computing and bioinformatic infrastructure
  - A high-performance storage and computing cluster (a set of connected computers that work together as a single system) is necessary to perform whole-genome

sequence analyses in high volumes. These analyses can be performed on a computing cluster consisting of many multi-core computers. An evaluation of these needs should be based on predicted volumes and specific analytical requirements for the test(s) that will be supported. Additionally, a tracking system for recording quality metrics across and within each sequencing run, lane, and sample is extremely useful for catching runs that go poorly and not wasting time and money on failed runs. These types of tracking system can also enable users to identify when additional sequencing will be necessary. Finally, bioinformaticians who are skilled in these analyses are important members of the NGS clinical team.

- A data management system for storage of genomic information should be planned for before implementing WGS in the clinical laboratory. Various guidelines suggest that sequencing results that could be used in evaluation of hereditary conditions should be stored for multiple years [3, 4]. The recently released CAP NGS checklist requires that data be stored for a minimum of 2 years to enable reanalysis of NGS results. This is in addition to other requirements around storage of actual clinical deliverables. What will be stored and how it will be stored require thorough consideration.
- Many software tools are available to support the multiple steps involved in WGS analysis. An evaluation of which tools should be used based on the intended use of the test should be performed. Once the right set of tools is identified, users may need to create a workflow using custom scripts that enable the usage of several tools, keeping in mind that input and output abilities and requirements may be variable among these tools. The software tools used in the analytical calling and downstream analysis and classification of variants are among the most variable aspects of clinical WGS being performed today. It is critical that laboratories understand the caveats and limitations associated with any of the software tools being used in their data analysis pipeline.

- Security
  - It is likely that WGS will be considered impossible to make anonymous. Privacy concerns around how these data are stored, when and how they are updated, who should have access, and what should go into the medical record are currently not well addressed by policies. However, laboratories are thinking about how this is likely to change and what safeguards and options they will be able to offer the doctors and patients who are interested in ordering WGS.

Ongoing Quality Assessment and Control

After validations have been performed and quality filters and metrics established, mechanisms are developed to monitor ongoing performance during testing of clinical samples. The process of genome sequencing can be divided up into three stages: wet-lab processing, bioinformatic analysis, and interpretation and report generation. The wet-lab component encompasses DNA extraction, DNA shearing and size selection, ligation of oligonucleotide adaptors to create a size-selected library, and physical isolation of the library fragments during amplification and sequencing. Each step of the process should be considered for the implications of a failure or contamination event; accordingly, the quality monitoring should be designed to detect the most likely or significant possible failures. Specifically, DNA extraction, library preparation, cluster generation, and the sequencing run should be assayed for quality. There are many ways in which quality can be monitored, and these include establishing run metrics at various steps, performing quality assessment steps (such as quantitative PCR (qPCR), DNA quantification and purity measures, run metric measures). Robotics and automation are valuable additions that can be made to a protocol to minimize the possibility of human error. Future advances to further combine the sequencing laboratory steps with automation will increasingly assure a reduction in potential errors. Controls can also be useful in the assessment of run quality. External controls, such as lambda DNA fragments, can be spiked into samples to measure the success of the run. Alternatively, orthologous assays such as microarrays can be utilized to measure sequencing accuracy at a very high level by comparing the concordance of calls from a genomic level microarray to the sequencing calls.

Proficiency testing is one method that is used as part of ongoing quality assessment. The molecular pathology on-site inspections by the CAP occur every 2 years, but ongoing proficiency testing with both intra- and interlaboratory analysis improves testing procedures and helps to prevent errors (reviewed in [4]). As several clinical laboratories are currently offering genomic level sequencing, alternative proficiency testing programs are used to enable laboratories offering exome and genome sequencing to compare their calls. In a recent exchange between the Illumina Clinical Services Laboratory and the University of California, Los Angeles (UCLA) molecular pathology laboratory comparing two samples that had been run and reported in both laboratories, both laboratories made calls for 3,573,631 sites, of which 19,340 represented variants from the reference. Across all the calls made, 16 positions were called discordantly between the two laboratories. Investigation of such discordantly called sites, along with relative quality metrics from

each run and the types of variants these sites represented (e.g., high GC regions or repeat regions), will help participating laboratories improve quality.

Conclusions

The implementation of clinical WGS is not trivial, and the suggestions made in this chapter highlight the need for well-trained teams that bring diverse expertise to the clinical laboratory. One challenge that is often raised is the lack of experts available; this is a legitimate concern and for that reason community efforts for establishing guidelines, and promoting education and best practices are critically needed. Ongoing training and certification, active participation in societies and meetings, and regular review of recent guidelines and publications will be necessary particularly during the early phases when the learning curve will be steep and policies are likely to evolve. That said, this is also a great opportunity for clinical laboratorians to work closely with their medical practitioner colleagues, as well as with experts in diverse fields such as bioinformatics, population genetics, and information technology to create a new approach to evaluating, diagnosing, and managing genetic disease using entire genomes of information.

Glossary

Allele frequency	Proportion of a particular allele among all alleles for a gene
Disease prevalence	Proportion of a population to have a condition
Mendelian condition	A condition that is caused by variants within a single gene and that can be passed to offspring in an autosomal dominant or autosomal recessive pattern
Proband	Affected individual on whom testing is being performed

References

1. Genetic Testing Registry. https://www.ncbi.nlm.nih.gov/gtr/.
2. Clinical Laboratory Standards Institute (CLSI). Multiple standards publications relating to laboratory process and sequencing. http://www.clsi.org/standards/.
3. College of American Pathologists (CAP). Molecular pathology checklist; 2013. p. 1–67.
4. Chen B, Gagnon MC, Shahangian S, Anderson NL, Howerton DA, Boone DJ. Good laboratory practices for molecular genetic testing for heritable diseases and conditions: Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention. MMWR Recomm Rep. 2009;58(RR-6):1–37.
5. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30(11):1033–6.
6. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.
7. Group AMPWGAW. The Association for Molecular Pathology's approach to supporting a global agenda to embrace personalized genomic medicine. J Mol Diagn. 2011;13(3):249–51.
8. Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Greiner TC, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. J Mol Diagn. 2012;14(6):525–40.
9. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15(9):733–47.
10. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet. 2010;42(1):30–5.
11. Dinwiddie DL, Kingsmore SF, Caracciolo S, Rossi G, Moratto D, Mazza C, et al. Combined DOCK8 and CLEC7A mutations causing immunodeficiency in 3 brothers with diarrhea, eczema, and infections. J Allergy Clin Immunol. 2013;131(2):594–7.
12. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13(3):255–62.
13. Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med. 2012;4:154.
14. Bick D, Fraser PC, Gutzeit MF, Harris JM, Hambuch TM, Helbling DC, Jacob HJ, Kersten JN, Leuthner SR, May T, North PE, Prisco SZ, Schuler BA, Shimoyama M, Strong KA, Van Why SK, Veith R, Verbsky J, Weborg AM Jr, Wilk BM, Willoughby RE Jr, Worthey EA, Dimmock DP. Successful application of whole genome sequencing in a medical genetics clinic. J Pediatr Genet. 2017;6:61–76.
15. Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. Nat Methods. 2009;6(11 Suppl):S6–12.
16. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11:473–83.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
18. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12(6):443–51.
19. Auton A, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.
20. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24. https://doi.org/10.1038/gim.2015.30.
21. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122. https://doi.org/10.1186/s13059-016-0974-4.
22. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:

SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92. https://doi.org/10.4161/fly.19695.
23. Antonarakis SE, The Nomenclature Working Group. Recommendations for a nomenclature system for human gene mutations. Hum Mutat. 1998;11:1–3.
24. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat. 2000;15:7–12.
25. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073–81.
26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Aondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
27. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med. 2017;9:1–14. https://doi.org/10.1186/s13073-017-0403-7.
28. Vassy JL, Christensen KD, Schonman EF, et al. The impact of whole-genome sequencing on the primary care and outcomes of healthy adult patients. Ann Intern Med. 2017:1–12. https://doi.org/10.7326/M17-0188.
29. Maddalena A, Bale S, Das S, Grody W, Richards S, ACMG Laboratory Quality Assurance Committee. Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. Genet Med. 2005;7:571–83.
30. ALPCW Group. ACMG recommendations for standards for interpretation of sequence variations. Genet Med. 2000;2:302–3.
31. Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical genome resource. Am J Hum Genet. 2017;100(6):895–906. https://doi.org/10.1016/j.ajhg.2017.04.015.
32. Garcia J, Tahiliani J, Johnson NM, et al. Clinical genetic testing for the cardiomyopathies and arrhythmias: a systematic framework for establishing clinical validity and addressing genotypic and phenotypic heterogeneity. Front Cardiovasc Med. 2016;3(4):228–11. https://doi.org/10.3389/fcvm.2016.00020.
33. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med. 2017;19(10):1105–17. https://doi.org/10.1038/gim.2017.37.
34. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91. https://doi.org/10.1038/nature1905.
35. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877–85. https://doi.org/10.1016/j.ajhg.2016.08.016.
36. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. BMC Genomics. 2013;14(Suppl 3):S3. https://doi.org/10.1186/1471-2164-14-S3-S3.
37. Online Mendelian Inheritance in Man, OMIM®. Baltimore, MD: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University; 2013. http://omim.org/.
38. Walsh M, Bell KM, Chong B, Creed E, Brett GR, Pope K, Thorne NP, Sadedin S, Georgeson P, Phelan DG, Day T, Taylor JA, Sexton A, Lockhart PJ, Kiers L, Fahey M, Macciocca I, Gaff CL, Oshlack A, Yiu EM, James PA, Stark Z, Ryan MM, Melbourne Genomics Health Alliance. Diagnostic and cost utility of whole exome sequencing in peripheral neuropathy. Ann Clin Transl Neurol. 2017;4(5):318–25.
39. Vissers LE, van Nimwegen KJ, Schieving JH, Kamsteeg EJ, Kleefstra T, Yntema HG, Pfundt R, van der Wilt GJ, Krabbenborg L, Brunner HG, van der Burg S, Grutters J, Veltman JA, Willemsen MA. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. Genet Med. 2017;19(9):1055–63. https://doi.org/10.1038/gim.2017.1.
40. Tarailo-Graovac M, Wasserman WW, Van Karnebeek CD. Impact of next-generation sequencing on diagnosis and management of neurometabolic disorders: current advances and future perspectives. Expert Rev Mol Diagn. 2017;17(4):307–9.
41. Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. Genome Biol. 2010;11(8):R82.
42. Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. Nat Rev Genet. 2013;14(5):321–32.
43. Riggs ER, Wain KE, Riethmaier D, et al. Towards a Universal Clinical Genomics Database: The 2012 International Standards for Cytogenomic Arrays (ISCA) Consortium Meeting. Human mutation. 2013;34(6):915-919. https://doi.org/10.1002/humu.22306.
44. Kalia SS, Adelman K, Bale S, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKalvey KD, Ormond KE, Richards CS, Vlangros CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19:249–55.

Clinical Information Systems in the Era of Personalized Medicine

19

Jonathan Nowak and Lynn Bry

Introduction

Scientific and technical advances continue to further our understanding of how genetic alterations affect human health and the development of disease. Integrating genomic findings in the delivery of patient care represents an exciting and rapidly evolving area of medicine. The capacity to interpret and leverage this new source of information, however, and to do so in a broad and high-throughput manner, via clinical information systems, remains a key challenge.

In spite of the challenges, institutions and CLIA-certified laboratories should recognize that existing clinical systems, operating procedures, and standards to support interoperability across systems do provide important resources to enable genomic analyses in patient care. Beyond individual patient testing, testing strategies also need to encompass delivery of genomic content to healthcare providers, and the means to warehouse this information, both to assist in ongoing research and development (R&D) activities to support CLIA testing and to evaluate outcomes from the use of genomic data in patient diagnosis, prognosis, and management.

Clinical genomics brings many new concepts and requirements to CLIA laboratories and healthcare institutions. Some factors affect processes within the testing lab, whereas others require additional institutional input to solve. In this chapter, we focus on four common areas that influence effective use and development of clinical information systems to support the integration of genomic data in healthcare:

1. Developing clinical systems to support genomic testing
2. Genomic standards for clinical systems and data interoperability
3. Factors to consider within the CLIA laboratory that performs genomic testing
4. Factors that involve but, by necessity, extend beyond the CLIA laboratory, including data warehousing, integrated reporting across diagnostic specialties, decision support tools, and effective warehousing of genomic information

J. Nowak · L. Bry (✉)
Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
e-mail: lbry@bwh.harvard.edu

Clinical Systems Support of Genomic Testing

Developing Clinical Infrastructure to Support Genomic Testing

As yet, the lack of end-to-end solutions to support data handling across technical, bioinformatics, and interpretive workflows requires that laboratories and institutions undertake projects of substantive complexity to implement genomic testing for clinical purposes. As platforms and vendor solutions improve, the efforts and costs required should drop. However, given the current complexities inherent in implementing genomic testing, particularly at the level of multigene panels and exome sequencing, laboratories and institutions need to develop a cohesive plan that defines the testing to be undertaken and the resources needed to support it. Broadly, analyses should include a business plan, institutional initiatives to be supported, as well as clearly specified, clinically actionable contributions to patient care. Standard methods for project management and integration of information systems [1] can assist in developing a robust plan. At a high level, these methods commonly incorporate the steps discussed below.

Development of Use Cases for Clinical Genomic Testing

What are the cases for genomic testing? Define the reasons and evidence to support testing, including clinical utility and support of clinical trials or translational research programs. Evaluate what types of testing will be performed in terms of sample type, such as peripheral blood or paraffin-embedded tumor tissue; target analyte, typically DNA or RNA; and

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_19

assay scope, which may range from a highly targeted gene panel to whole-exome or whole-genome sequencing, as each has different needs in terms of information systems support within and external to the testing laboratory. Laboratories often do best to play to local and institutional strengths. To ensure that one-off processes are not developed in technical and IT plans, laboratories may wish to select at least two areas of focus that may relate to disease and type of testing (germline, somatic, infectious disease), but no more than three to four to ensure adequate focus and development of infrastructure that can handle testing within a reasonable time frame.

Requirements Gathering

Given the use cases, what resources and infrastructure are needed to support them, from the point of ordering genomic tests to reporting results back to the ordering physician and associated electronic health records (EHR)? In addition, it is important to consider population-scale needs for test interpretation and improvement, such as data warehousing to enable evaluation of the potential significance of new variants and for retrospective demonstration of outcomes based upon genomic results. Other factors to evaluate include reanalysis of genomic data with respect to clinical triggering events, such as future patient visits or the need to "push" novel and clinically actionable findings to clinicians as clinical evidence regarding the significance of particular variants evolves. Thorough gathering of requirements will touch not only upon the clinical information systems but includes an evaluation of operational, logistic, and other resources needed to support end-to-end processes.

Validation of Requirements

The validation of requirements is a necessary "sanity check" in the process in order to allow for the evaluation of requirements and their capacity to be supported with available institutional resources and budgets. It is not uncommon for CLIA laboratories and institutions to revise the initial plan and testing scope to align with what existing resources and funding can facilitate.

Gap Analysis

Given the requirements, what systems and/or resources exist, and which need to be developed? Evaluate costs and resources associated with each, including costs associated with the purchase of new systems, as well as upgrades to existing clinical information systems. After an initial gap analysis, reevaluate use cases and requirements and iterate as needed to develop a final plan that incorporates areas in which genomic testing can be accomplished in a manner that fits within institutional needs and available budgets and resources.

Functional Specifications

Functional specifications that include IT requirements in terms of software, hardware, and systems integrations need to be created (Fig. 19.1). The specifications will be part of the overall business, financial, and operational plan. The IT components should also include needs for supporting personnel including project managers, clinical systems analysts, database administrators, system administrators, bioinformaticians, and additional supporting computational staff and statisticians. Support of technical platforms, software, and hardware must also be incorporated. Included in these analyses should be an understanding of requirements for attaining a break-even point and determining the return on investment derived from support of clinical research programs within or across institutions.

Timeline and Plan

It is important to generate a timeline for development and integration of resources, showing key milestones to be met and dependencies across clinical, technical, and informatics needs.

After undertaking all these activities, laboratories and institutions may realize that starting with more complex forms of testing such as exome and/or genome analysis can be quite challenging, particularly if local expertise and infrastructure do not already exist. Modification of plans to focus on targeted areas, particularly where vendor kits and informatics solutions may be leveraged to provide a turnkey platform, can be helpful, along with a strategic goal of deploying more complex forms of testing in the future. In this manner, greater focus can be placed on ensuring that the needed IT systems and infrastructure are available to support the initial assay to be deployed while also maintaining a pathway to support future test development.

LIS Versus LIMS: Understanding the Setting in Which CLIA Genomic Testing Will Occur

Many institutions face the following challenges when implementing a plan for clinical genomic testing: (1) how to effectively leverage expertise and resources that may exist in a research core performing complex genetic testing but that is not certified as a CLIA laboratory [2] and (2) how to leverage expertise and resources from the CLIA laboratory for testing that contains many components that are novel and often foreign to many clinical laboratories. To this latter point, evaluations for infrastructure to support diagnostic testing will include understanding contributions from existing clinical laboratory information systems (LIS) versus non-CLIA laboratory information management systems (LIMS) that may be encountered in research environments.

Whereas clinical LIS that specialize in supporting complex genomic testing are now becoming commercially available [3, 4], these products remain external to the

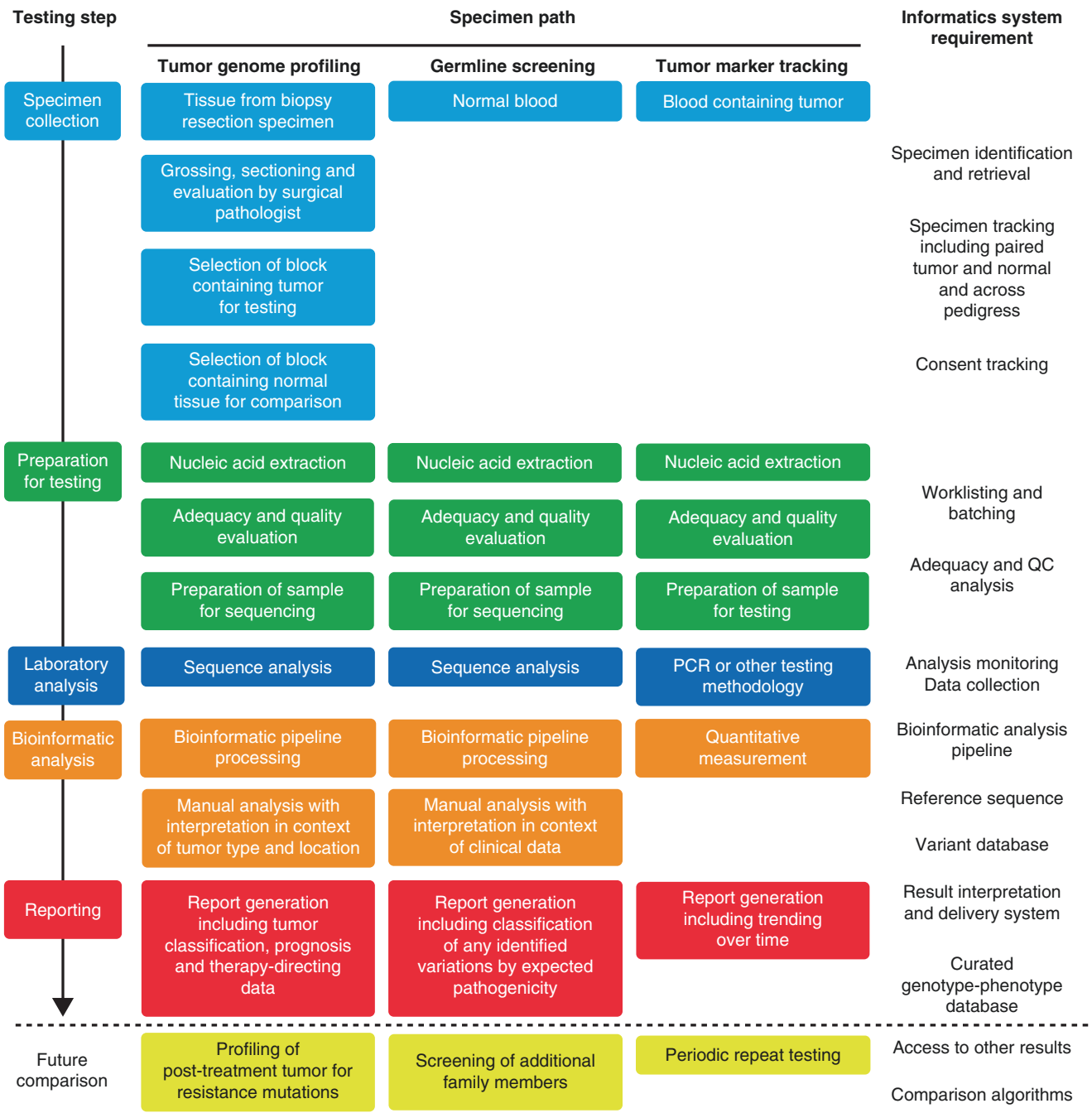


Fig. 19.1 Testing pathways and informatics system requirements for representative types of molecular testing

standard, vendor-based clinical LIS [5] that support high-volume testing in clinical laboratories and anatomic pathology services. In implementing programs for clinical genomic testing, CLIA laboratories thus need to assess whether to internally develop needed components or consider purchasing a dedicated, "best in breed" LIS to support genomic testing and then focusing on systems integration with the main LIS. Section 18.3 goes into more detail regarding specific areas to evaluate. As shown in Fig. 19.2, the interactions between the CLIA laboratory's LIS and the infrastructure supporting clinical genomic testing can take a variety of forms:

1. The clinical LIS remains the system of record for data structures and processes that support the essential business process for CLIA testing, such as those regarding clients ordering tests, patient information, insurers, sample and test dictionaries, and fee codes/schedules. Under this scenario, the genomics LIS may operate fairly independently of the clinical LIS but requires means to

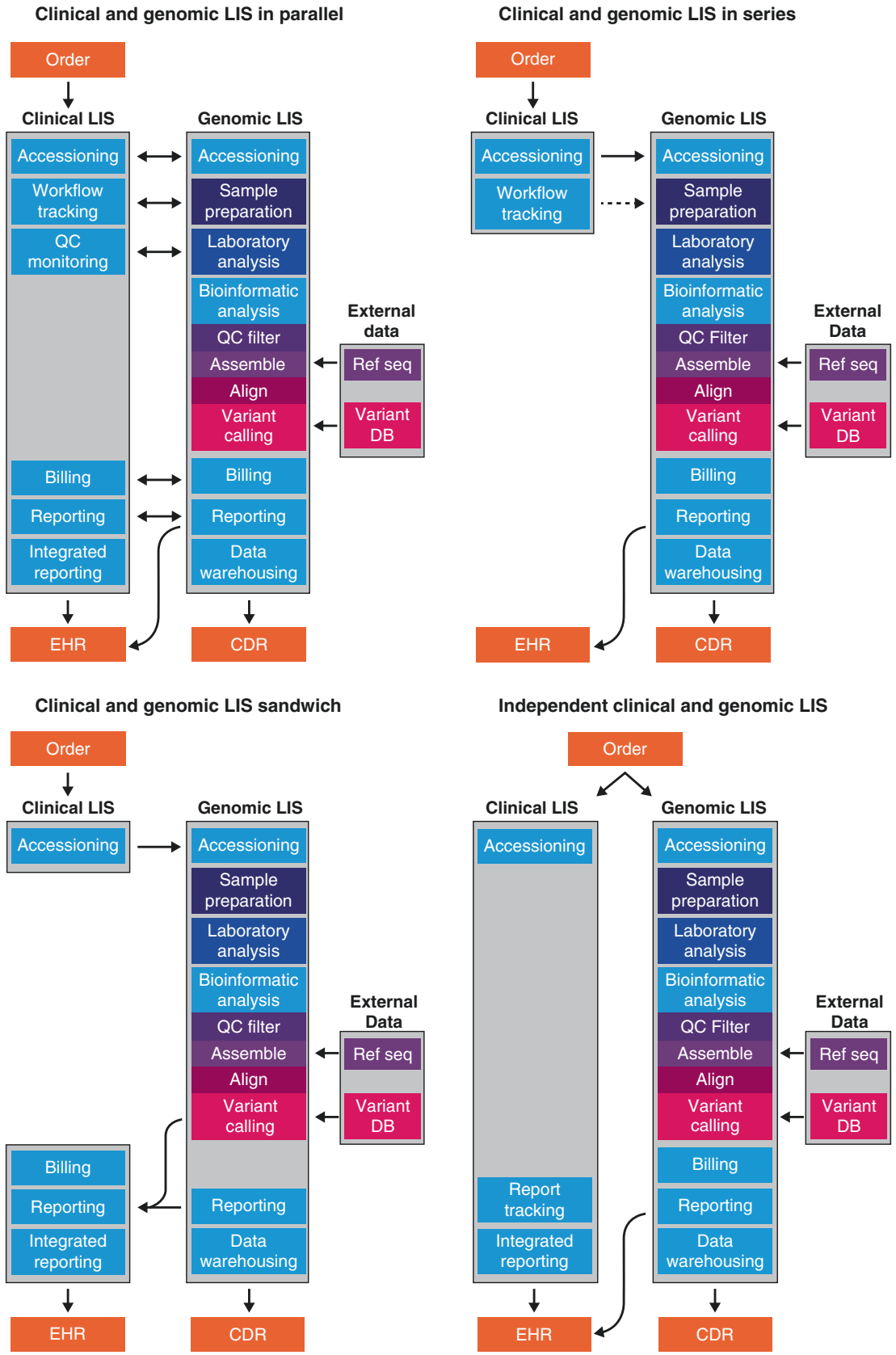


Fig. 19.2 Possible configurations for integrating the clinical LIS and the genomic LIS

retrieve and update data structures maintained in the clinical LIS. In this manner, both systems share common data structures and vocabularies, or ontologies, for ordering, testing, reporting, and billing.
2. The clinical LIS handles the up-front business process for test ordering and may handle additional steps including accessioning and work listing. Thereafter, orders are communicated to the genomics LIS, optimally via messaging standards such as those developed by Health Level 7 (HL7) (www.hl7.org) [6]. Receipt into the genomic LIS may require a separate accessioning process upon receipt of the patient order, but subsequent downstream steps are handled within the genomics LIMS, including communication of final results and reports.
3. The clinical LIS handles the initial and final end points of genomic testing, centered around order receipt and return of a final report to the ordering clinician and the EHR. Systems integration with the genomics LIS or LIMS defines operational and IT components needed to facilitate forwarding of needed sample and patient data for CLIA testing to occur and return of results from the genomic analyses, which may include steps from the call of variants to return of a structured report that will be forwarded to the client.
4. Both the clinical LIS and genomics LIS are completely separate, which can occur within a single institution, and is also the structure if leveraging genomic testing from an outside CLIA reference laboratory. In this situation, systems integration will focus on means to communicate orders and receive results from the testing lab.

Of note, laboratories and institutions facing a need to get LIMS and other non-CLIA resources to perform to CLIA specifications in support of clinical genomic testing can refer to the Next Generation Sequencing (NGS) section of the Molecular Pathology checklists [7] developed by the College of American Pathologists. These documents provide standards and quality parameters to be followed in validating clinical LIS and for implementing NGS in a CLIA environment.

Genomic Standards for Clinical Systems and Data Interoperability

While clinical genomics is still a maturing field, international efforts have developed standards to support data and systems interoperability. Though still new and evolving, the following resources provide means to send, receive, and warehouse genomic data.

Gene-Level Calls and Coordinates

Efforts by many groups including the Human Genome Variation Society (HGVS; [8]), NCBI (www.ncbi.nlm.nih.gov), EMBL (www.embl.de), and medical associations including the American College of Medical Genetics and Genomics (ACMG; www.acmg.net), College of American Pathologists (CAP; www.cap.org), and Association of Molecular Pathology (AMP; www.amp.org) have developed and supported use of common nomenclatures for describing gene variants. At their simplest, these systems typically specify the location of a variant using genomic reference coordinates and then describe the predicted impact on any protein that may be encoded by a gene at that location. Standardized nomenclatures also exist to support description of copy number alterations, haploid phasing of variants, and effects of genomic variants on alternative transcripts. These baseline formats can be leveraged in subsequent data structures, including the variant call format, and in HL7 messages that can communicate clinical data across systems. However, laboratories should be aware that discrepancies may exist when considering nomenclature systems that focus on cytogenetic versus genome sequence or transcript-based positions.

Although the HGVS recommendations cover the broad range of common genetic alterations, new applications require continual expansion of the nomenclature. For example, whereas the nomenclature for describing translocations detected by karyotypic or FISH analysis is well defined by the International System for Human Cytogenetic Nomenclature, there is not yet a broadly accepted way for reporting translocations detected by NGS. A second necessary component for standardized reporting is broadly agreed upon reference materials and databases of known variants. Although the sequence of the human genome was declared complete in 2003, analysis and annotation of the sequence are still ongoing, with a reference annotation only completed in 2012, and one that is routinely updated [9, 10]. It is essential to recognize that genomic variants and transcript variants are typically described in relation to a specific release or version of the human genome and reference transcript set. The same underlying nucleic acid change will often be mapped to different genomic and transcript locations in different versions of the datasets. Therefore, it is best practice to also record and transmit information about the source and version of the genomic reference sequence that is used for analysis.

Genomic File Formats

A variety of standard file formats are utilized during genomic testing, from the .fastq and .bam file formats used to store sequence data in early stages of bioinformatics analyses to the variant call format (VCF; [11]) that provides a commonly used format for the structure and reporting of variants identified against a reference genome. By storing only the variants identified against a reference, the VCF file greatly reduces the amount of information that needs to be stored or communicated. It has thus become a standard means for communicating variants, whether from targeted sequencing or exome- or genome-level analyses.

Used extensively within the 1000 Genomes Project, the VCF format includes metadata elements to store information regarding the gene sequence and specific identified variants. Variants are identified by their genomic coordinates relative to a defined reference genome. Version 4.0 of the format also includes quality information associated with the call of each variant and filtering information if an external system or algorithm assigned specific information regarding the state or quality of the variant. While the format does not provide defined structures or methods for documenting the pipelines used to perform analyses, this information may be captured in metadata fields or in the file header. As adoption of clinical genomic testing increases, we anticipate that the VCF format and underlying support structures may evolve to better support clinical testing. Additionally, other file formats that convey information about copy number and structural variants alterations may also become widely adopted, although formats for documenting these variant classes currently lag behind the VCF format in terms of maturity and widespread adoption.

Health Level 7

Health Level 7 (www.hl7.org) is a nonprofit organization that develops standards to support interoperability across healthcare systems. HL7's Clinical Genomics working group has devised standards for communicating pedigree data [12] and a structured genetic test report (GTR; [6]). Both projects contain detailed specifications and implementation guides that may be downloaded from their website. While both are still relatively new and continue to evolve, they provide an internationally developed standard to communicate complex genomic information across systems.

The pedigree model provides a data standard to capture and communicate family relationships for a given patient, including diseases and genetic risk factors. A working example of the model has been implemented for the "My Family Health Portrait" website managed by the US Surgeon General (https://familyhistory.hhs.gov/fhh-web/home.action). While broader adoption within commercial EHRs is being considered, this may require substantive alterations to data structures storing patient information, as well as addressing patient privacy and protection concerns under HIPAA. This latter concern largely relates to linking individuals within a medical database, if one or both parties have not explicitly given consent to do so [13, 14].

HL7's GTR supports reporting of sequence-based variants, cytogenetics, and gene expression studies. Message structures include standard components for communicating the ordering institution, clinician or practice, and patient demographic data. The "Test Details" section includes data structures to communicate reasons for testing (including diagnostic codified data), specimens sent for testing, as well as elements used by the testing laboratory in reporting variants or other findings, interpretations, and additional supporting information to accompany reports.

Version 2 of the GTR, released in January 2013, provides the capacity to link Logical Observation Identifiers Names and Codes (LOINC; [15]) to genetic tests and reports and has been piloted at various sites worldwide [16, 17].

Of note, the capacity for the Portable Document Format (PDF) to render healthcare data from embedded Extensible Markup Language (XML) [18] offers potential opportunity to store HL7 messages in these file formats. The healthcare PDF standard may thus provide a means for institutions to store messages in a format that can also generate a human-readable report.

As with any standard, HL7 alone does not provide the application layer needed to perform core functions once data are sent or received but provides an essential component for defining methods to communicate data across sites. In practice, most large CLIA laboratories and institutions have invested in teams and supporting IT infrastructure to implement and manage HL7 messages. As such, the availability of communication standards for genomic testing has the potential to reduce the time and effort required to otherwise develop and maintain de novo processes. These standards also generally provide an improved capacity to scale as clinical testing and associated volumes of data to be communicated increase.

Standard Data Sources and Content

Several global projects aim to develop standards and content for the clinical interpretation of genomic variants. Whereas projects such as the Online Mendelian Inheritance in Man database (OMIM; www.omim.org), the Catalogue of Somatic Mutations in Cancer (COSMIC; http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/), and the Genome Aggregation Database (gnomAD; http://gnomad.broadinstitute.org) arose from research activities, these databases often provide content to CLIA laboratories that evaluate the significance of genomic findings. Some initiatives, including ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), ClinGen (https://www.clinicalgenome.org/), and My Cancer Genome (http://mycancergenome.org), aim to provide additional curation as well as "CLIA-grade" tools through which laboratories may communicate variants identified as well as contribute supporting evidence regarding variant interpretation. As the field progresses, vendor solutions that aggregate or license defined content will also become more widely available.

Factors to Consider Within the CLIA Laboratory that Performs Genomic Testing

Various features of the standard LIS may be leveraged to support genomic testing. Box 19.1 highlights many of these components. However, genomic testing also typically requires a number of items that are new to the technical and IT staff within a CLIA laboratory. These latter areas, detailed in Box 19.2, are where infrastructure, resources, and personnel need to be developed to support the associated activities.

Box 19.1 LIS Features to Be Leveraged in Genomic Testing

- Data structure to support business processes for diagnostic testing
- Support for order entry and interfacing with order entry functions in an electronic health record (EHR)
- Sample accessioning and tracking
- Test ordering and worklisting
- Management of quality control (QC) and quality assurance (QA) processes
- Receipt of results, including interpretation
- Billing triggers to assist in billing for testing

Box 19.2 New Information System Requirements to Consider in Genomic Testing

- Decision support to guide clinicians in the ordering of genomic tests
- Capturing of additional data at the point of ordering genomic tests, such as additional patient consent, pedigree information, or other factors such as tumor cellularity
- Managing complex technical and quality control steps including library preparation, bar coding, and multiplexing of samples
- Linking to and managing bioinformatics pipelines, including version control, and monitoring of pipeline performance for individual patient cases and across sequencing runs
- Developing and managing CLIA-grade content for interpreting and reporting genomic results
- Integrated reporting of genomic information with other phenotypic analyses including histopathologic and/or clinical laboratory biomarkers
- Developing data storage resources for genomic information, both for ready retrieval of information when needed and to meet any medicolegal and associated state or local laws regarding the storage of clinical data
- Leveraging genomic data in clinical decision support
- Potential need to reevaluate genomic datasets on a periodic basis, relative to defined clinical triggers such as a patient visit or push of new information regarding variants that are medically actionable
- Data warehousing of genomic results to support evaluation of unknown variants and improve test panels

Clinical Systems Supporting Order Entry of Genomic Tests

Whereas orders for genomic tests share many aspects with orders routinely placed for other forms of patient testing, several properties merit special attention. A first consideration regards determining when testing is warranted. Particularly in the case of germline testing, the decision to test relies upon integrating data from the patient's medical history, clinical examination, and laboratory findings with pedigree information. Without a centralized mechanism for routinely entering and communicating these data in a structured manner, opportunities to make a genetic diagnosis may be missed. In the case of cancer testing for somatic variants, analyses may be conducted under a research protocol, or complex genomic analyses may only be considered after initial screening tests that use phenotypic markers or focused molecular diagnostic tests. These factors need to be communicated to the ordering clinician, and appropriate pre-existing information needs to be relayed back to the laboratory to direct testing, especially when multistep algorithms are in place.

In addition, the mechanism by which a given gene or genetic region can be tested can also influence how the test may be ordered. Unlike most clinical lab tests, in which the specific technique is a clear component of the test, genomic testing may require, as an example, sequencing multiple regions of the genome while being cognizant of the intrinsic limitations of the assay's technology, such as the inability to detect structural variations or alterations in copy number. While many of these processes may remain internal to the laboratory as testing for specific patient cases progresses, testing of certain genomic regions may also require that the ordering physician and supporting personnel be informed of such aspects at the time of ordering.

To fully address these issues, order entry systems are an essential part of the clinical workflow for genomic testing. Genetic test ordering benefits from order entry systems that provide the means to search and compare available assays and link into decision support tools to aid with the selection and ordering of appropriate tests that are supported by medical evidence [19–21]. Ideally, this process could even be automated. For example, patients exhibiting abnormal responses to pharmacologic therapy could be automatically flagged for evaluation of drug metabolism enzymes. Most importantly, such systems free clinicians from the burden of maintaining detailed knowledge about indications for both common and rare genetic tests, while providing ready access to resources that allow them to tailor possible testing to a patient's individual scenario [22, 23]. Notably, although such clinical decision support systems are not yet widely implemented, they are among the most requested EHR functions related to genomic medicine with the ultimate aim of improving personalized healthcare [21].

Specimen Identification and Tracking for Genetic Tests

Although most clinical LIS are well equipped for tracking a wide variety of specimen types, including those routinely used for genetic testing, several accommodations need to be made for genetic testing. One critical requirement is tracking of the patient materials used for testing, particularly in the case of cancer diagnosis where multiple samples may be sent for molecular and phenotypic analyses. For solid tissue specimens, testing is routinely performed on a subset of the available material, typically a portion of a single paraffin block. Testing may also include solid tissue and fluid samples, such as for B-cell clonality assays that could be performed on blood, cerebral spinal fluid (CSF), and tissue, where comparison of results across sites may be critical in guiding therapeutic decisions. Cell-free DNA (cfDNA), isolated from blood plasma, is a more recent specimen type to enter the clinical testing realm and is typically analyzed to make inferences about DNA being shed by solid tumors. Therefore, reports for all cases need to include an unambiguous statement about what material was used for testing and what the assay is designed to detect. For cancer-based testing, the adoption of automated, whole-slide imaging systems can facilitate the documentation of material used for testing by creating a permanent, high-resolution record of the exact material that was used, even if the material on the original slide is consumed to accomplish the testing. An additional, increasingly relevant requirement is the ability to track both germline DNA for a patient and somatic DNA, typically isolated from a tumor specimen. Bioinformatics analyses of somatic DNA often relies upon knowledge of germline DNA variants so that they can be subtracted from variants identified in somatic sequencing in order to generate a purely somatic dataset. However, germline DNA may also be analyzed independently, in order to identify inherited syndromes. While many LIS offer the ability to perform multiple tests based upon a single specimen, it is less common for LIS to provide native support for tracking multiple specimens, such as germline and somatic DNA, for joint analysis in a single assay.

In addition to tracking input material for testing, systems for tracking genetic test material need to have flexible and robust capabilities for handling samples that fail testing or are judged to be technically inadequate for analysis [19]. In each case, the laboratory's information system must be able to identify when cases have not passed quality control checks and divert them for appropriate handling. Such cases may be "failed" outright or may instead be triaged for other types of molecular testing that may have less demanding specimen requirements. A final consideration for material tracking is the archiving and storage of samples after testing. Although not all samples are necessarily retained by the laboratory, some forms of testing including chimerism and clonality analyses can rely critically upon the ability to retest previously analyzed samples. Furthermore, testing of paternity and of extended familial pedigrees often warrants storing tested samples beyond a specified time after reporting the final results. Management of such long-term storage is frequently not incorporated effectively into clinical LIS. Although numerous commercial software packages can handle many of these tasks individually, there is limited integration between these systems and LIS. However, as the volume and complexity of genetic testing grow, the capabilities of these programs should increase, and their ability to link to other LIS packages should strengthen.

LIS Tracking of Consents and Results Reporting

Several additional types of information unique to genetic testing often need to be managed within clinical laboratory systems [24]. Genetic testing may require additional consents beyond those obtained for routine clinical testing [25]. In addition to simply tracking patient consents, the LIS may also be called upon to track multiple types or levels of consent for a given test. As testing platforms based on NGS become more prevalent, results frequently include findings of unknown medical significance. Some tests may also identify incidental findings that are not directly related to the initial disease in question but which nevertheless may be medically informative for the patient. Different patients may have distinct preferences about being informed of such results. To address this possibility, recently proposed recommendations for informed consent prior to performing whole-genome sequencing have advocated a category-based model for disclosing different classes of findings [26], though the degree to which CLIA laboratories implement these levels depends upon local and institutional views of genomic testing and use of results. However, the LIS may have to track which results should be released to the patient based on information documented in the consent forms [26, 27]. In addition, as genetic data from a patient may be periodically reevaluated by more sophisticated algorithms drawing from updated knowledge bases, the amount and nature of new information may be quite different from that for which informed consent was initially obtained. Not only will consents need to be designed broadly enough to account for new information from periodic reevaluation, but the clinical infrastructure in CLIA laboratories may be called upon to alert clinicians that new results are available for their patient [26].

Proper interpretation of genetic tests may also require information about multiple individuals from the family pedigree to be associated with the individual being tested, particularly from the parents and siblings of a patient [28]. This information may include the approval to provide the results of testing to other family members. In these situations, multiple specimens from different individuals may need to be

linked within the information system so that they are tested and analyzed together before the release of a single report to the patient's EHR. The necessity of linking multiple patients to a single patient record is a rather unique requirement of genetic testing and is conceptually different from the standard one-to-one relationship between patients, specimens, and results that underlines traditional LIS design. In fact, commercial LIS do not readily handle receipt of supporting samples under the individual who provided them. Rather, in most CLIA laboratories, additional samples may be accessioned under the primary patient, with additional fields added to uniquely identify the individual and their relationship with the patient being tested.

Standardized Report Formats for Genetic Tests

Traditionally, the LIS reports result in a highly structured and standardized format, though free-text elements in a narrative format may exist. Although genetic results typically contain a mixture of structured data, such as the precise genomic location of an identified mutation, and unstructured data, such as a text-based interpretation of the results, this information is typically reported into the EHR in an unstructured format. However, as the amount and complexity of genetic tests increase, it will be necessary to adopt a standardized and structured template for reporting results [21, 29]. Structured data will not only facilitate comparing and transmitting results among provider systems but are also essential to enable downstream algorithms and tools to provide decision support to clinicians and patients [30]. Additionally, structured reporting also facilitates the warehousing of genomic information, to create knowledge bases for developing content as well as to enhance laboratory quality control programs that monitor new and previously encountered variants. As described above, several standards are being adopted to facilitate structured and standardized reporting. Efforts to identify and categorize normal variants and disease-causing alterations are still evolving rapidly [31]. Therefore, it is essential for molecular reports to include detailed information about the reference material used and genomic databases queried, as changes to these data sources could potentially alter test interpretation. It is also important to explicitly convey which genes and genomic regions have been interrogated by an assay. For example, depending upon the intended use, a multigene panel may target only selected exons of a gene, all exons of a gene, or all exons and some introns of a gene. Finally, it will be important for structured reports to indicate the method used for analysis, because multiple methods may be applicable when testing a given region of the genome [29]. Although Current Procedural Terminology (CPT) codes exist for many molecular diagnostics assays and although these have been recently updated to better reflect current testing practices, they frequently lag behind the introduction of new technologies for genetic testing. Additionally, as they are primarily designed for billing needs, they may not be able to capture the necessary details about how a test was performed, particularly as bioinformatics and computational analyses play an increasing role in the reporting of molecular results. Additional medical nomenclature systems such as the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC), and the Unified Medical Language System (UMLS) may be necessary to succinctly and unambiguously communicate testing and analysis methodologies.

Factors that Involve but, by Necessity, Extend Beyond the CLIA Laboratory, Including Data Warehousing, Integrated Reporting Across Diagnostic Specialties, Decision Support Tools, and Effective Warehousing of Genomic Information

Integrated Reporting with Other Anatomic Pathology and Clinical Laboratory Data

Although some molecular assays represent independent laboratory studies, many begin with pre-existing specimens that have simultaneously undergone nongenetic testing in anatomic or clinical pathology laboratories. As such, the clinical value of the molecular data only becomes apparent when interpreted in the context of the other nonmolecular laboratory data for the specimen. Beyond the realm of certain complex anatomic pathology cases, typically hematopathology and soft tissue pathology, the ordering clinician has traditionally borne the burden of integrating disparate and potentially asynchronously provided results (Fig. 19.3). However, as the breadth of molecular testing grows and its interdependence with other laboratory results increases, the capacity to integrate molecular and nonmolecular findings becomes increasingly important, and implementation will fall to the clinical diagnostic laboratory.

Cases involving simple and routine data integration have the potential to be handled within the LIS, particularly to integrate genomic findings with pertinent phenotypic markers that have also been performed in the clinical molecular diagnostic laboratory. For example, patients with HIV who receive highly active antiretroviral therapy (HAART) undergo routine monitoring of their viral load and CD4+ T-cell counts. They may also undergo periodic HIV genotyping, per defined changes in clinical status, to assess development of antiretroviral drug resistance in the underlying population of HIV virions. These frequently repeated and highly standardized results could be integrated within a standard report format in the LIS to better facilitate longitudinal assessment of response to therapy and continued management. This approach integrates the molecular data with

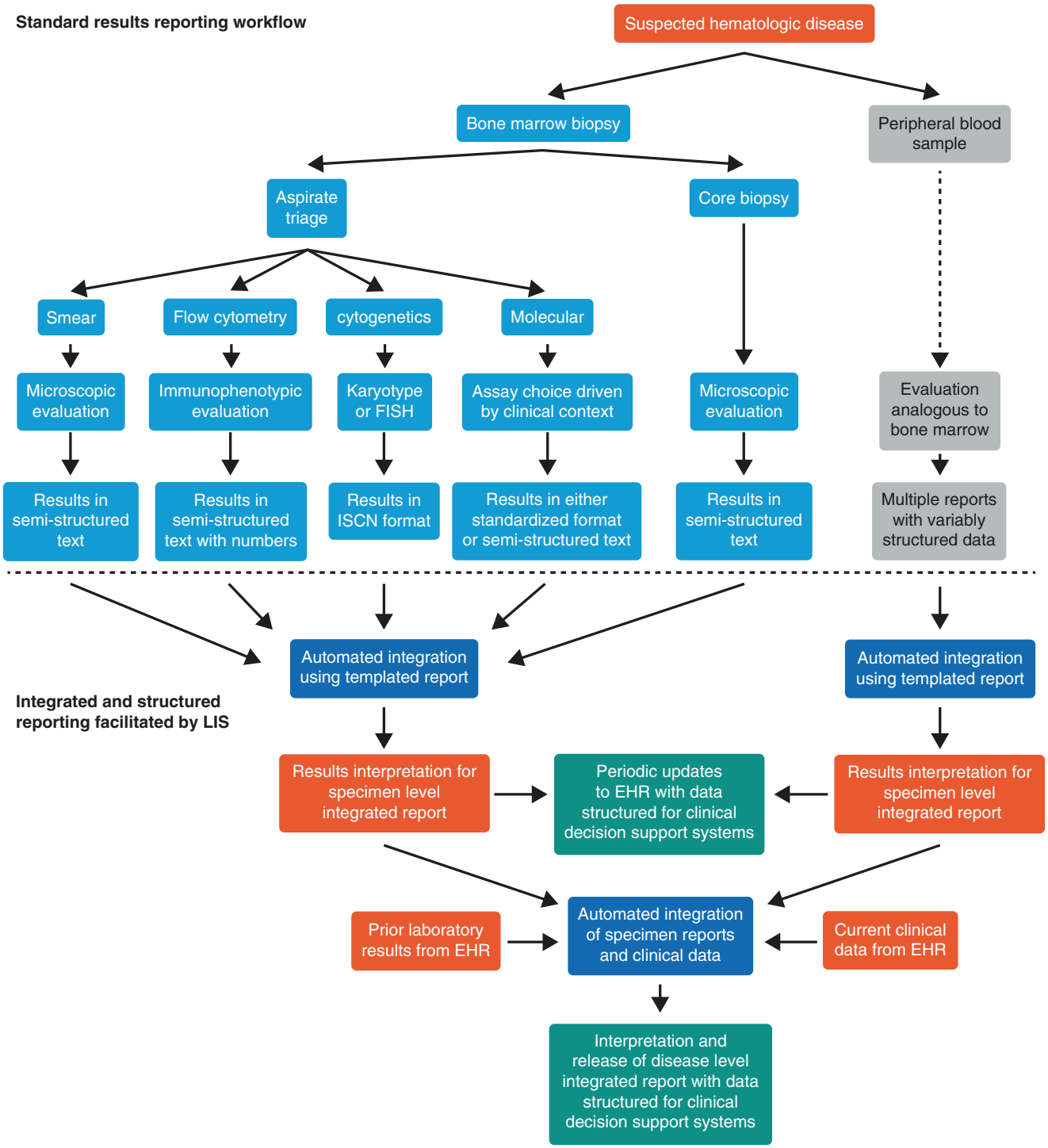


Fig. 19.3 The standard pathology reporting workflow and a proposed mechanism for implementing integrated reporting in the LIS. EHR electronic health records

important phenotypic markers, providing an improved context in which to assess the meaning and validity of findings and their integration in the overall clinical status of the patient. Such an approach is also highly relevant to the cancer realm, where blood-based monitoring for disease burden and recurrence is well-established for some disease types, such as *BCR-ABL* detection for monitoring chronic myelogenous leukemia, and where it is rapidly coming into play using cell-free DNA for a variety of solid tumor types.

As another example, a similar approach could be employed for monitoring glucose control in patients with type 2 diabetes by leveraging pharmacogenomic markers

predictive of drug responsiveness with phenotypic markers of glycemic control. Pharmacogenomic studies have recently identified polymorphisms in genes affected by sulfonylureas that can predict an individual's response to treatment with these drugs [32]. Beyond guiding initial therapy selection, the patient's genetically predicted sensitivity profile to different drug classes could be retrieved within the LIS and integrated with periodic glucose and hemoglobin A1c data to predict an expected response to sulfonylurea treatment as a comparison with the patient's actual response and as a measure of treatment compliance [33, 34]. Such an integrated report would be especially helpful in tracking the progression of type 2 diabetes and would have the capacity to assist in the selection of personalized therapy per the patient's underlying genetic background and current phenotypic presentation of the disease.

In more complex cases, molecular data will need to be structured and managed in the LIS in a manner that facilitates integration with other types of pathology information. Such will often be the case with somatic mutation analyses of tumors, where the dataset from diagnostic testing in pathology laboratories often includes histopathological assessment, phenotypic biomarkers, and results from multiple, potentially complex molecular tests. In the past, molecular results for tumor specimens have typically been reported in an isolated fashion either as addenda to already finalized surgical pathology reports or as completely separate reports released into the medical record [35]. However, two related trends in tumor biology are driving the need for integration of anatomic pathology data. First, molecular alterations are increasingly defining tumors and tumor subtypes, as well as aiding in the selection of therapies. For example, recent guidelines from the College of American Pathologists for reporting ancillary biomarker studies for lung and colorectal adenocarcinomas underscore the importance of molecular data in the standard characterization for these tumor types [36, 37]. Indeed, information such as the key driver mutation for a lung adenocarcinoma may be one of the single most important results that an oncologist wishes to obtain from a surgical pathology specimen, given the evidence supporting its use in selecting pharmacologic therapy. The recent approval of checkpoint blockade inhibitor pembrolizumab for the treatment of all advanced solid tumors with mismatch repair deficiency means that a large proportion of solid tumors now have an approved indication for molecular analysis [38]. When this information is generated, it will need to be clearly contained within an integrated report for the specimen in much the same way that data necessary for cancer staging are routinely included in surgical pathology reports.

A second aspect driving integrated reporting is the recognition that many types of molecular data cannot be interpreted meaningfully in the absence of additional pathology data. The significance of the same mutation in a given gene may vary widely depending upon the type of tumor in which it is detected [39]. This context dependence is extremely important for many NGS assays on tumors, for which the final interpretation is closely linked to the original tumor type. Simply reporting that a tumor has a mutation in *KRAS* provides limited information without the interpretative context of the associated tumor type. Finally, the sheer amount of data derived from evaluating many tumors with complex assays begins to exceed the point at which manual review and synthesis of findings can be supported in any scalable capacity. The pathologist therefore plays an essential role in providing medical direction and supervision regarding needed data integration and reporting (Fig. 19.3b).

Data Warehousing

In generating genomic results for the individual patient, testing laboratories and institutions should plan to warehouse the information aggregated across cases and populations tested [22]. Key benefits of warehousing include providing the means to mine the information when evaluating new variants, to identify their prevalence in certain populations and/or to assess clinical outcomes. Such retrospective datasets can also be leveraged in ongoing CLIA laboratory quality assurance (QA) and quality improvement (QI) activities. The warehoused information also provides an invaluable resource to support active research programs, including translational activities needed to assess new or unknown variants and to develop the medical evidence regarding their use in patient care [40–42].

Whereas many open-source and commercial applications have been developed to warehouse genomic information [42, 43], the evaluation of what system, or systems, to use needs to consider existing institutional resources and expertise, underlying funding, and continuing support for infrastructure maintenance.

In most healthcare institutions, pathology information systems contribute more than half of all data transactions into an EHR. Considering the volume of clinical laboratory and other high-throughput forms of testing, this dataset provides a rich source of phenotypic information and is commonly the most structured and codified in healthcare systems. In spite of the amount and richness of pathology data, the means to store and effectively warehouse genomic information within the EHR frequently requires resources outside of the pathology department or clinical laboratory, in part due to the fact that current commercial LIS are not optimized to generate and manage genomic data. Thus, at an institutional level, it is important that pathologists actively participate in the planning and development of resources to warehouse genomic information, including the tools used to leverage it for basic, translational, and clinical activities (Fig. 19.4).

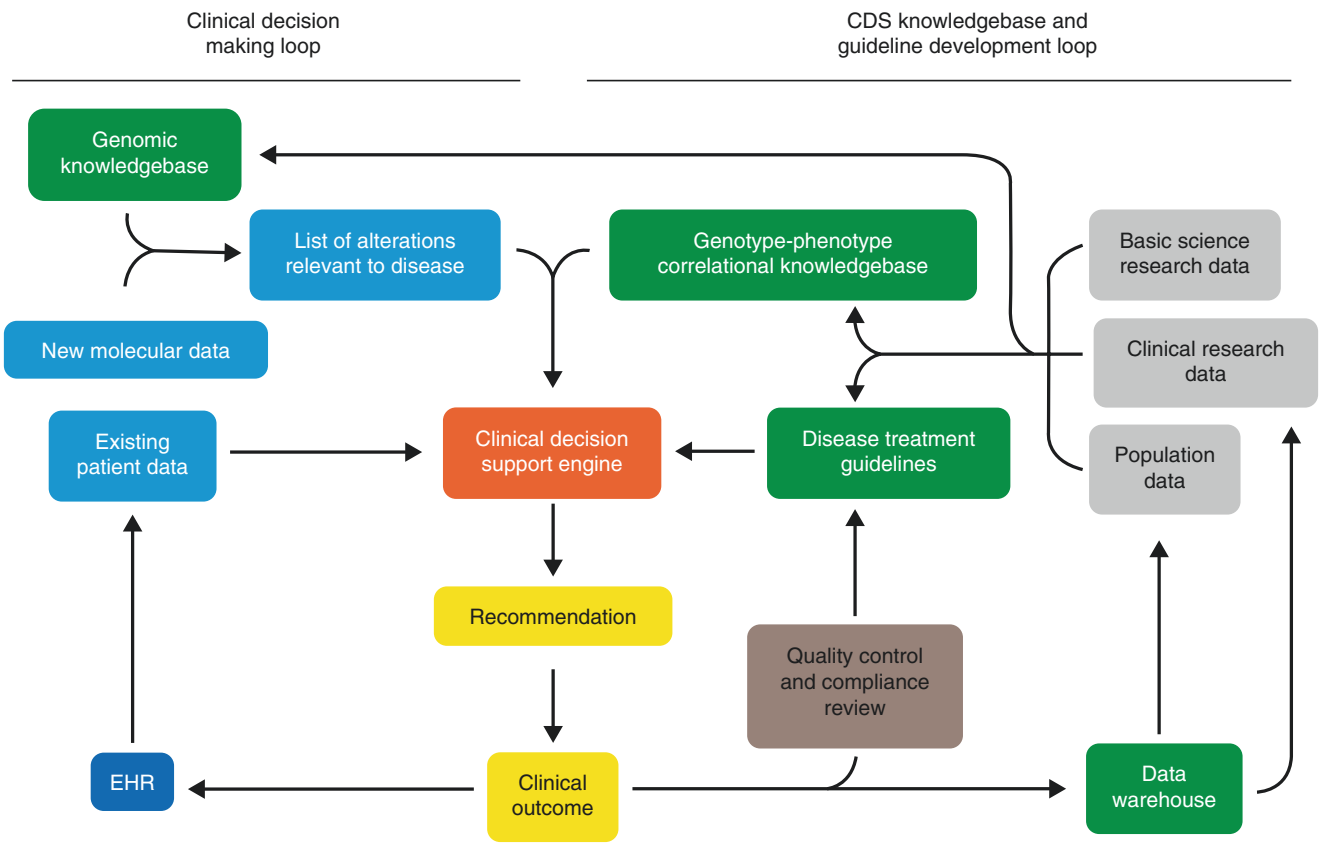


Fig. 19.4 Mechanism for incorporating research data with clinical outcomes to improve clinical decision support systems. CDS clinical decision support, EHR electronic health records

Decision Support

Until recently, most genetic test results were interpreted in a manner comparable to single analyte results. This approach was feasible given the non-multiplex nature of many early molecular assays. While this strategy can work for highly penetrant genetic variants with defined phenotypes and associated medical evidence supporting their use in clinical care, this method proves suboptimal when evaluating multigene interactions and the need to present complex information to clinicians [44]. To meet this need, clinical decision support systems (CDSS) are being adapted to incorporate genetic results. CDSS are uniquely suited to analyzing genomic medicine information because of the absolute amount of information generated, the highly structured nature of the genetic results, and the rapidity with which our knowledge and interpretation of genetic variants is increasing [20, 45]. These systems commonly leverage population-based knowledge bases to provide prevalences and prior information regarding genotype–phenotype relationships for a given disease, with sets of clinical rules or criteria to generate recommendations for clinical action (Fig. 19.5). These systems may be particularly helpful for enrolling patients in clinical trials, where inclusion criteria are increasingly being defined by molecular alterations and where trial availability may change frequently.

Notable examples where CDSS have proven useful in linking genomic information with clinical outcomes are in the areas of pharmacogenomics and cancer management [43]. In pharmacogenomics, many genomic variants have been identified that can predict the likelihood of overall drug effectiveness and the potential for adverse drug reactions. Such systems are now increasingly used to optimize dosage of drugs with narrow therapeutic ranges, minimizing adverse drug reactions as well as selecting optimal therapy based on the patient's genetic background [46–48]. Though early in their development, a number of CDSS systems have also been developed to aid in cancer risk reduction and cancer management [44, 45] using genetic information. Finally, whereas many of these currently available CDSS tools are widely applicable and show substantial benefit to patients, most operate from a relatively limited knowledge base and set of rules. More sophisticated systems are emerging that draw from broader and more deeply curated knowledge bases, to enable highly complex analyses and interpretations. For example, IBM's Watson Health system, developed in collaboration with Memorial Sloan Kettering Cancer Center, is designed to provide diagnostic and treatment recommendation for cancer patients by merging knowledge from clinical experts with molecular and genomic data, along with

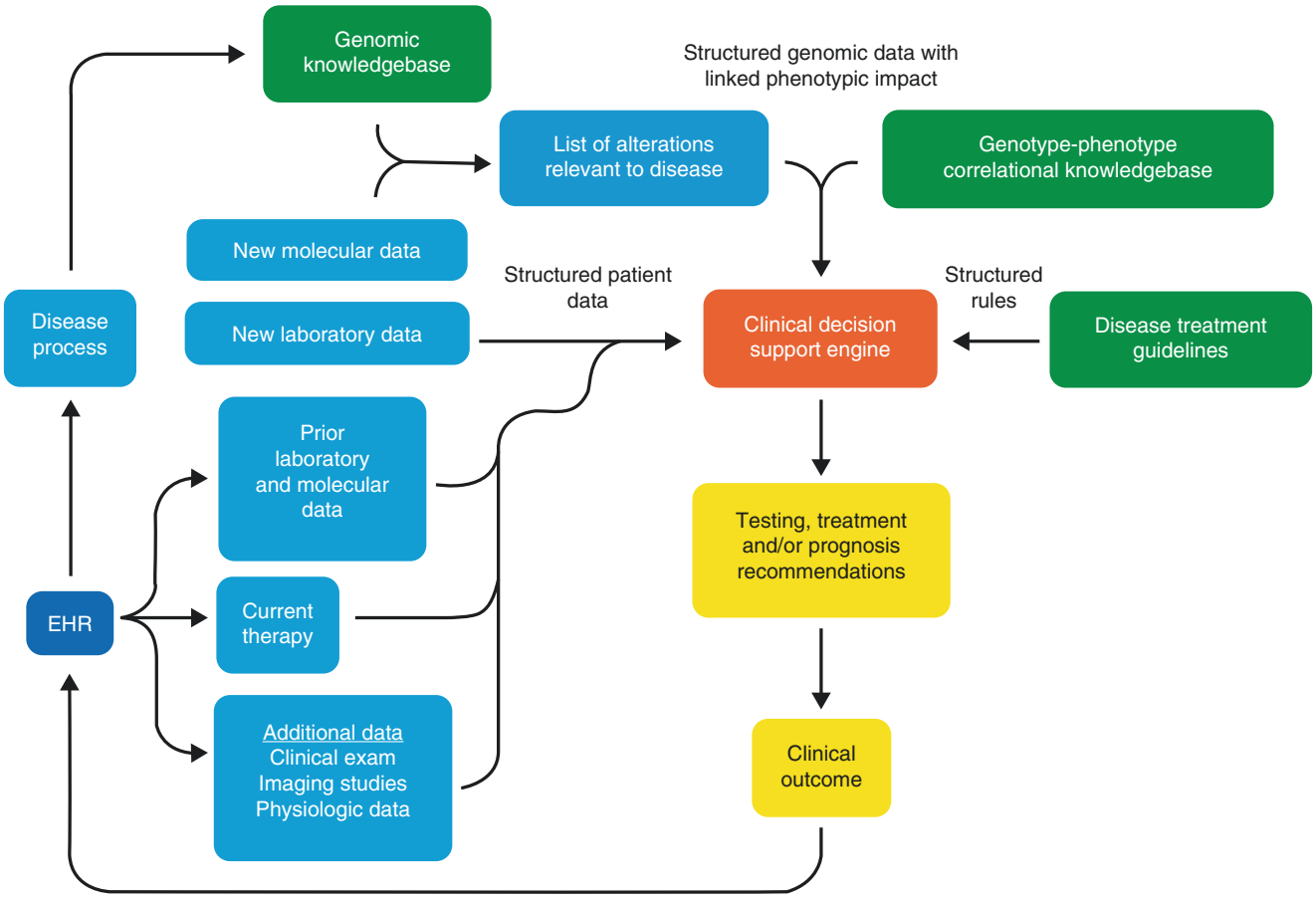


Fig. 19.5 Clinical decision support system function and required inputs in ongoing practice

outcomes from cancer case histories (https://www.ibm.com/watson/health/). While the system is designed to enhance the dissemination of practice-changing research to nonexperts, a process which can frequently take more than a decade, it will also function as a much broader platform for guiding decisions in oncology, providing clinicians with a continuously updated set of treatment recommendations that are shaped by genomic data streams and refinements in clinical guidelines. It is also important to be able to tailor CDSS to the specific treatment options and clinical trial availability at the institution treating the patient. Tools such as MatchMiner (https://matchminer.org/), a computational platform for real-time matching of cancer patients to precision medicine clinical trials using genomic data, can be tightly linked to the laboratory systems reporting genomic data and can serve as localized CDSS instances [49].

Regardless of the form that CDSS take, they share common informatics requirements when interacting with LIS. First, genomic information needs to be reported in a machine-readable format in a defined location, and not just released as human-viewable free-text results in the EHR [41]. Utilization of a system such as the HL7 Clinical Genomics messaging standard or Clinical Bioinformatic Ontology (CBO) will likely be necessary to unambiguously communicate genetic data between reporting systems and the decision support engine [30]. Analogously, additional nongenetic information will also need to be accessed by the decision support engine to provide context for the evaluation of the genetic result [22]. Whereas existing standards may be adequate to convey a subset of this information, it will also be necessary to use a controlled vocabulary to define patient phenotypic data so that they can be uniformly accessed and understood by the clinical decision engine. Programs such as the Electronic Medical Records and Genomics (eMERGE) Network and PhenX (Consensus Measures for Phenotypes and Exposures) have begun to standardize the collection and annotation of phenotypic information for use in genome-wide association studies but could also provide a phenotypic reporting system that would easily be adapted to CDSS use [50, 51]. In the somatic testing realm, the Oncotree nomenclature system (http://www.cbioportal.org/oncotree) has rapidly gained acceptance as a method for documenting tumor type and location in a manner that is compatible with genomic testing.

CDSS implementation also requires access to knowledge bases that document and link genotype and phenotype

relationships for the disease or medical process of interest. Although the knowledge bases utilized by CDSS thus far have typically been purpose-built, databases such as ClinVar/ClinGen and efforts such as the Clinical Pharmacogenomics Implementation Consortium (http://www.pharmgkb.org/page/cpic) may eventually evolve to act as integrated repositories containing structured genotype–phenotype data that can support automated decision engines. A separate but closely related resource also required for CDSS is a rule set for generating a recommendation based on the patient data and genotype–phenotype knowledge base. Although rule sets may be based upon accepted standards for treatment of different conditions, these standards will need to be translated and stored in machine-readable structures. Additionally, they will need to be customized and validated at a hospital level to ensure compatibility with established institutional workflows. Common frameworks for representing CDSS rule sets are not yet available, but collaborations such as Health eDecisions (www.healthedecisions.org) are being developed to provide a common syntax for CDSS rules. Finally, clinical decision support systems will need to be structured in a way that permits the information contained within them to be rapidly updated and validated as new genetic data accumulate and therapeutic options and prognostic data evolve. Even though the implementation of clinical decision support systems requires several new bioinformatics and computational tools, many of these resources may be reused, furthering the adoption of CDSS once standards are in place.

Conclusions

Genomic datasets present new challenges to clinical laboratories, pathology departments, and healthcare institutions, particularly in providing a wealth of data for which evidence is often lacking regarding their application to clinical care. Clinical LIS provide an essential set of systems to facilitate ordering, testing, and communication of medically relevant information yet also need to provide mechanisms by which findings of unknown significance can undergo future evaluation and be warehoused to aid in population-based analyses of findings. At an institutional level, major systems integration together with the development of new systems is frequently needed to enable clinical decision support that adequately utilizes genomic data. Broader adoption of electronic health records and incorporation of new technologies that leverage new computational models and means to store and transmit data will improve our capacity to harness this information, as well as handle anticipated large datasets from other forms of diagnostic testing. Regardless of the testing modality to be considered, standard methods for developing pathways to implement complex plans, with a focus on using robust standards when available, can assist with providing needed systems integration and can facilitate appropriate utilization of such resources.

References

1. Irani Z, Love P. Evaluating information systems. Burlington: Elsevier Ltd; 2008.
2. Centers for Medicare & Medicaid Services. Clinical laboratory improvement amendments. Retrieved from http://www.cms.gov; 2014.
3. GenoLogics—A LIMS for the Next-Gen Omics lab. Retrieved from http://www.genologics.com; 2014.
4. Sapio Sciences—The most configurable and flexible LIMS software available. Retrieved from http://www.sapiosciences.com; 2014.
5. Gale K. Laboratory and the art of enterprise integration. Healthc Financ Manage. 2009;63(10):36–8.
6. (HL7), H.L.S.I. Clinical genomics; 2013.
7. College of American Pathologists Molecular Pathology Checklist. Retrieved from http://www.cap.org; 2014.
8. (HGVS), H.G.V.S. Standards—definitions, symbols, nucleotides, codons, amino acids (v2.0); 2013.
9. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–45.
10. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760–74.
11. Genome Project. VCF (Variant Call Format) version 4.0; 2011.
12. (HL7), H.L.S.I. HL7 Version 3 Standard: clinical genomics; Pedigree, Release 1; 2013.
13. Lebo RV, Grody WW. Testing and reporting ACMG cystic fibrosis mutation panel results. Genet Test. 2007;11(1):11–31.
14. IRB Guidebook: Chapter V Biomedical and behavioral research: an overview. Retrieved from http://www.hhs.gov; 2014.
15. Regenstrief Institute. Logical Observation Identifiers Names and Codes (LOINC); 2013.
16. Bosca D, Marco L, Burriel V, Jaijo T, Millán JM, Levin A, Pastor O, Robles M, Maldonado JA. Genetic testing information standardization in HL7 CDA and ISO13606. Stud Health Technol Inform. 2013;192:338–42.
17. Chute CG, Kohane IS. Genomic medicine, health information technology, and patient care. JAMA. 2013;309(14):1467–8.
18. ASTM. Portable Document Format-Healthcare (PDF) a best practices guide. ASTM AIIMASTM—BP-01-08; 2013.
19. Hoffman MA. The genome-enabled electronic medical record. J Biomed Inform. 2007;40(1):44–6.
20. Garg AX, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005;293(10):1223–38.
21. Scheuner MT, et al. Are electronic health records ready for genomic medicine? Genet Med. 2009;11(7):510–7.
22. Downing GJ, et al. Information management to enable personalized medicine: stakeholder roles in building clinical decision support. BMC Med Inform Decis Mak. 2009;9:44.
23. Scheuner MT, Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. JAMA. 2008;299(11):1320–34.
24. Ronquillo JG. How the electronic health record will change the future of health care. Yale J Biol Med. 2012;85(3):379–86.
25. Lucassen A, Hall A. Consent and confidentiality in clinical genetic practice: guidance on genetic testing and sharing genetic information. Clin Med. 2012;12(1):5–6.

26. Ayuso C, et al. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. Eur J Hum Genet. 2013;21(10):1054–9.
27. Bunnik EM, et al. The new genetics and informed consent: differentiating choice to preserve autonomy. Bioethics. 2013;27(6):348–55.
28. Hinton RB Jr. The family history: reemergence of an established tool. Crit Care Nurs Clin North Am. 2008;20(2):149–58. v
29. Gulley ML, et al. Clinical laboratory reports in molecular pathology. Arch Pathol Lab Med. 2007;131(6):852–63.
30. Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. Hum Mutat. 2011;32(5):512–6.
31. Naidoo N, et al. Human genetics and genomics a decade after the release of the draft sequence of the human genome. Hum Genomics. 2011;5(6):577–622.
32. Aquilante CL. Sulfonylurea pharmacogenomics in type 2 diabetes: the influence of drug target and diabetes risk polymorphisms. Expert Rev Cardiovasc Ther. 2010;8(3):359–72.
33. Huang C, Florez JC. Pharmacogenetics in type 2 diabetes: potential implications for clinical practice. Genome Med. 2011;3(11):76.
34. Wilcox AR, Neri PM, Volk LA, Newmark LP, Clark EH, Babb LJ, Varugheese M, Aronson SJ, Rehm HL, Bates DW. A novel clinician interface to improve clinician access to up-to-date genetic results. J Am Med Inform Assoc. 2014;21(e1):e117–21. https://doi.org/10.1136/amiajnl-2013-001965.
35. Wilkins BS, Clark DM. Making the most of bone marrow trephine biopsy. Histopathology. 2009;55(6):631–40.
36. Cagle PT, et al. Template for reporting results of biomarker testing of specimens from patients with non-small cell carcinoma of the lung. Arch Pathol Lab Med. 2014;138(2):171–4.
37. Bartley AN, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the colon and rectum. Arch Pathol Lab Med. 2014;138(2):166–70.
38. Lemery S, Keegan P, Pazdur R. First FDA approval agnostic of cancer site–when a biomarker defines the indication. N Engl J Med. 2017;377(15):1409–12.
39. Richards CS, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med. 2008;10(4):294–300.
40. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. J Am Med Inform Assoc. 2012;19(2):181–5.
41. Masys DR, et al. Technical desiderata for the integration of genomic data into electronic health records. J Biomed Inform. 2012;45(3):419–22.
42. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R, Zeng Q, Dubey A, Gainer V, Mendis M, Glaser J, Kohane I. Instrumenting the health care enterprise for discovery research in the genomic era. Genome Res. 2009;19(9):1675–81.
43. Aronson SJ, Clark EH, Babb LJ, Baxter S, Farwell LM, Funke BH, Hernandez AL, Joshi VA, Lyon E, Parthum AR, Russell FJ, Varugheese M, Venman TC, Rehm HL. The GeneInsight suite: a platform to support laboratory and provider use of DNA-based genetic testing. Hum Mutat. 2011;32(5):532–6.
44. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. J Am Med Inform Assoc. 2013;20(2):388–400.
45. Belle A, Kon MA, Najarian K. Biomedical informatics for computer-aided decision support systems: a survey. Scientific World Journal. 2013;2013:769639.
46. Peterson JF, et al. Electronic health record design and implementation for pharmacogenomics: a local perspective. Genet Med. 2013;15(10):833–41.
47. Pulley JM, Denny JC, Peterson JF, Bernard GR, Vnencak-Jones CL, Ramirez AH, Delaney JT, Bowton E, Brothers K, Johnson K, Crawford DC, Schildcrout J, Masys DR, Dilks HH, Wilke RA, Clayton EW, Shultz E, Laposata M, McPherson J, Jirjis JN, Roden DM. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. Clin Pharmacol Ther. 2012;92(1):87–95.
48. Tural C, et al. Clinical utility of HIV-1 genotyping and expert advice: the Havana trial. AIDS. 2002;16(2):209–18.
49. Lindsay J, et al. MatchMiner: An open source computational platform for real-time matching of cancer patients to precision medicine clinical trials using genomic and clinical criteria. bioRxiv. 2017;199489 https://doi.org/10.1101/199489.
50. Pathak J, et al. Evaluating phenotypic data elements for genetics and epidemiological research: experiences from the eMERGE and PhenX network projects. AMIA Summits Transl Sci Proc. 2011;2011:41–5.
51. Pathak J, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE network experience. J Am Med Inform Assoc. 2011;18(4):376–86.

Reporting Clinical Genomic Assay Results and the Role of the Pathologist

20

Janina A. Longtine

Introduction

Over the past decade, we have seen a rapid rise in the number of clinically relevant molecular diagnostic assays accompanied by increasingly sophisticated technologies and complexity of generated data. This development has been driven in part by the discovery of the molecular underpinnings of disease. Relatively simple genotyping assays designed to detect a single allelic variant in one gene, such as *F5* c.1601G > A (p.Arg534Gln) or factor V Leiden, advanced to genotyping panels within one gene. The American College of Medical Genetics and Genomics/American Congress of Obstetricians and Gynecologists recommended a panel of 23 mutations for cystic fibrosis screening as an example of the latter [1]. Further knowledge led to assays involving multiple mutations in multiple genes, such as the pathogenic variants in sarcomere proteins associated with hypertrophic cardiomyopathy [2, 3] or the molecular stratification of lung adenocarcinoma used to predict response to targeted therapies [4, 5] as well as large cancer panels with hundreds of genes which interrogate base substitution, insertions and deletions (indels), copy number variation, and rearrangements [6, 7]. We have now entered an era of even greater complexity (and uncertainty) with the clinical application of exome or genome sequencing. Throughout this time, molecular pathologists and molecular geneticists have developed and implemented diagnostic assays following guidelines for quality assurance and test reporting issued by the College of American Pathologists (CAP), the American College of Medical Genetics and Genomics (ACMG), and the Association for Molecular Pathology (AMP). This chapter will review the challenges inherent in generating and delivering rational, informative genomic clinical reports and highlight emerging solutions.

J. A. Longtine (✉)
Departments of Pathology and Laboratory Medicine, Yale University School of Medicine, Pathology and Laboratory Medicine, Yale New Haven Hospital and Smilow Cancer Hospital, New Haven, CT, USA
e-mail: janina.longtine@yale.edu

Reporting of Single-Gene or Gene Panel Results

Genotyping or targeted sequencing assays are designed to interrogate single-nucleotide variants or small indels with *known* phenotype-genotype correlation. In these assays, the clinical report should follow recommended guidelines of reporting; include laboratory, patient, and sample identifiers, the results indicating that the tested mutation is detected or not using standardized gene nomenclature; and provide analytical and clinical interpretations with appropriate documentation from the medical literature [8–10].

Proceeding to single-gene sequencing or gene panel sequencing created the Pandora's box of variants of unknown significance (VUS), which are DNA variants that have not been reliably characterized as benign or pathogenic. Conventional genetic approaches, using segregation of the variant with disease in large family studies with affected individuals, are effective for assessing the significance of a VUS. This is particularly powerful for high-penetrance, rare variants. Unfortunately, this method is not applicable in the evaluation of most VUS. Some VUS have been characterized as pathogenic using a combination of clinical data and in vitro or animal model experiments that were conducted to prove biologic relevance, but this approach is difficult and not readily applied. For VUS in protein-coding exons, there are a number of computational (in silico) predictive programs that can assist in determining whether a variant is likely to be damaging to the protein structure or function by using bioinformatics tools to assess evolutionary conservation and the variant's effect on protein structure, such as Polymorphism Phenotyping v**2** (PolyPhen-2) [11], Sorting Intolerant from Tolerant (SIFT) [12], Protein Variation Effect Analyzer (PROVEAN) [13], and others [14]. The predictive power of these tools is quite variable and may not correlate with clinical disease in humans. There is also a risk of overinterpretation of pathogenicity due to limited understanding of contextual information, such as biologic modifiers [15]. In addition, it can be just as difficult to prove that a variant is

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_20

benign as it is to prove that it is pathogenic. Common variants in minority populations that have not yet been well-defined further confound interpretation because variants annotated as pathogenic with low minor allele frequency may be single-nucleotide variants (SNVs) in minority populations. This issue is being addressed through population databases such as the 1000 genomes project (http://www.1000genomes.org) and the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org, last accessed June 11, 2017). The ExAC database has variants identified during exome sequencing of 61,486 unrelated individuals of European, African, South Asian, East Asian, and Latino ancestry [16, 17].

For clinical reporting ACMG and AMP have jointly recommended five tiers for classifying sequence variants in patients with suspected inherited (Mendelian) disorders [14] (Table 20.1). To determine the appropriate tier, the variant is evaluated by 16 specific criteria based on the evidence observed for pathogenicity of that variant (very strong, strong, moderate, or supporting) and for 12 specific criteria for benign impact (stand-alone, strong, supporting). The criteria are then combined to determine the classification tier in Table 20.1. For example, a null variant in a gene where the loss of function is a known mechanism of disease (very strong) and has well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product (strong) would be classified as pathogenic. A similar "very strong" null variant located in a mutational hotspot of a well-established functional domain (moderate) would be classified as likely pathogenic. The guidelines recognize that the criteria may be more stringent than typically used in clinical laboratories but are likely to reduce the number of variants being classified as causative and therefore actionable without sufficient supporting evidence.

Several databases are available to assist in interpreting genomic variants (e.g., UniProtKB (www.uniprot.org, last accessed June 21, 2017), Ensembl (http://useast.ensembl.org/index.html, last accessed June 21, 2017), and UCSC Genome Browser (http://www.genome.ucsc.edu, last accessed June 21, 2017) as well as the National Center for Biotechnology Information (NCBI) resources such as ClinVar, dbSNP, and Variation Viewer) [18]. Many public databases are not of clinical grade, may contain errors, and should be utilized critically.

Table 20.1 ACMG/AMP classification of inherited sequence variants

Pathogenic
Likely pathogenic
Benign
Likely benign
Uncertain significance

Reporting of Whole-Exome and Whole-Genome Sequencing Results

With the decrease in cost of whole-exome (WES) and whole-genome (WGS) sequencing, there has been a push to move from multigene panels to these potentially more cost-effective approaches for clinical care. With gene or gene panel testing, the genes are selected because they are known to be implicated in a disease or influence therapeutic options. The test is ordered to address a specific clinical question. The comprehensive data generated by WES or WGS inherently includes a discovery process because only a subset of the data can be rationally associated with disease using current knowledge. Therefore, the challenge is to determine how much of the data will be reported in a clinical setting. There are several different approaches. One is that bioinformatics data analysis identifies all variants, but only those known to be associated with the disease in question are fully analyzed, interpreted, and reported. Alternatively, in addition to gene variants known to be relevant to the patient's disease, all gene variants known to be associated with human disease that are medically actionable and analytically verified are reported. For the latter, it is critical to set the bar high to minimize reporting of variants as pathogenic, which may later turn out to be benign. In 2013, the ACMG recommended a minimum list of known pathogenic or expected pathogenic variants in 56 genes that should be reported during clinical WES or WGS even when unrelated to the medical reason for the testing [19]. These secondary findings are related to monogenic disorders that have evidence of clinical utility. The policy was revised in 2015 to offer people undergoing germline WES/WGS the option to opt out of receiving the secondary findings [20]. In 2016, the recommendations were further revised to include 59 medically actionable genes [21]. Another approach is to implement systematic reanalysis of nondiagnostic clinical exome sequencing and/or to view the data as a resource that could be interrogated over the life of the patient as different medical needs and conditions develop [22, 23]. There is no clear mechanism for payment of reanalysis at this time. Due to the expense of data storage and the rapid change in technology, it may be more cost-effective and easier to repeat the testing from the start rather than store data.

It is important to understand the technical limitations of the utilized assay prior to generating a report. The next-generation short sequence reads are aligned to a reference genome, a best estimate of the gene sequence is determined (base calling), and variants are identified. The variants are filtered bioinformatically in order to generate a candidate list of pathogenic variants. To critically analyze the candidate list, one must be fully cognizant of how the data were filtered to know what type of variants may or may not have been detected. In addition, one must understand the sequencing

methodology (e.g., capture design) and quality metrics of the sequence reads (how well individual regions are sequenced) to generate an informative report. To ensure this, the geneticist and/or the pathologist must work closely with their bioinformatics team to understand the process and potential pitfalls. The patient (and clinician) must also understand the limits of the testing and its interpretation through the consenting process.

Reporting of Cancer Test Results

The classification of cancer has been transformed by the discovery of specific cytogenetic and molecular aberrations that identify biologic subgroups of neoplasms within those previously grouped and classified according to histologic type. Recurrent genetic translocations that define subgroups of acute myeloid leukemias (AML) in the WHO classification are prognostic and frequently predictive. They are also often associated with characteristic morphologic and immunophenotypic features [24]. For karyotype normal AMLs, mutations in *NPM1, FLT3*, and *CEBPA* provide important prognostic information to guide therapy. The advent of targeted therapy has further advanced cancer molecular diagnostics by identifying "driver" mutations inherent in the pathogenesis of the specific cancer types that are also sensitive to inhibitory therapy. The *BCR-ABL1* fusion gene encoded by t(9:22)(q34;q11.2) is the driver of chronic myelogenous leukemia. Its protein is the target of tyrosine kinase inhibitors and its chimeric mRNA is a sensitive tool for monitoring response to treatment and identifying drug resistance. The discovery of constitutively activating mutations in the *EGFR* gene in lung adenocarcinoma and the therapeutic efficacy of targeted small molecular inhibitors, such as gefitinib and erlotinib, heralded the importance of molecular diagnostics in solid tumor taxonomy [25–27]. Histologic classification is insufficient, and molecular testing is required to identify *EGFR*–mutant, responsive cancers. The molecular stratification of lung adenocarcinoma has continued to evolve, identifying multiple mutually exclusive driver mutations associated with different targeted treatments, such as *ALK*, *ROS1, BRAF, ERBB2, MET* mutations, *MET* amplifications, and *RET* rearrangements. In addition there are variants that predict a lack of response to targeted therapy, such as *KRAS* codons 12, 13, and 61 (National Comprehensive Cancer Network. Non-Small Cell Lung Cancer (Version 4.2017), https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf. Accessed June 22, 2017). As such, pathologists now must seamlessly integrate molecular and cytogenetic/FISH testing into routine care and incorporate the mutation profile lexicon into their diagnostic armamentarium.

Clinically relevant testing algorithms have been developed in molecular pathology laboratories to sequentially identify hotspot mutations, based on prevalence in different cancer types. As multiple hotspots in multiple genes became clinically relevant and as multiplex technologies evolved, many laboratories moved to multigene mutation profiling to identify "actionable" mutations in a timely and cost-effective manner [28–30]. This evolution raises several important points related to informative cancer mutation reports. Most of the actionable mutations are heterozygous, diluting the targeted mutant alleles (1:1) with non-mutated, wild-type alleles. In addition, clinical specimens are heterogenous with a mix of normal and tumor cells. Both factors contribute to the potential reduction of the mutant alleles within the cancer DNA specimen and require the development of sensitive tests with defined limits of detection and rigorous quality controls [31]. In addition, a skilled pathologist is needed to estimate the percent of cancer cells within the specimen to determine specimen adequacy. Both anatomic and clinical pathology training and expertise are helpful. The reporting pathologist must understand the sequencing technology, bioinformatics algorithms, and the assay limitations before issuing a negative report.

Some clinical laboratories are now moving to exome panels for cancer mutation profiling. A challenge that emerged with broad genotyping panels and which holds with exome panels or WES/WGS is the need for well-curated cancer mutation knowledge bases. For example, *EGFR* exon 19 deletions and the exon 21 p.Leu858Arg variant point mutation in lung adenocarcinomas confer sensitivity to tyrosine kinase inhibitors, whereas exon 20 insertions mutations are associated with primary resistance to these drugs (Costa, D. 2015. EGFR Exon 20 Insertion in Non-Small Cell Lung Cancer. *My Cancer Genome* https://www.mycancergenome.org/content/disease/lung-cancer/egfr/64/ (updated November 5); last accessed June 26, 2017). Approximately 10–15% of *EGFR*-mutant lung cancers have less common *EGFR* mutations that were not included in many clinical trials, making it difficult to predict response to targeted therapy [32]. *BRAF* mutation p.Val600Glu (commonly known as p.V600E) has different therapeutic implications for melanoma and colorectal carcinoma. *BRAF* p.Val600GluE mutations lead to constitutive activation of the MAPK signaling pathway. *BRAF* p.ValV600GluE positive metastatic melanomas have a dramatic response (60–80%) to the selective *BRAF* inhibitor, vemurafenib [33]. However, colon cancer with *BRAF* p.Val600Glu infrequently (<5%) responds to vemurafenib due to EGFR-mediated MAPK pathway reactivation, leading to vemurafenib resistance [34]. In addition, BRAF p.Val600Glu in colon cancer is associated with lack of response to anti-EFGR monoclonal antibody therapy (e.g., cetuximab, panitumumab) (NCCN. Colon Cancer Version 2.2017, https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf, last accessed July 7, 2017). It is a challenge for each laboratory and its staff to be fully informed of

the current literature as well as all available clinical trials. Furthermore, exome panels applied across all cancer types may reveal variants with limited clinical evidence of utility in a particular cancer subtype or reveal more common variants in an uncommon or unexpected tumor type. These necessitate time-consuming literature investigations that may yield only preclinical data or small studies or case reports that preclude a truly informative report.

There is a need for accurate, curated knowledge databases for pathologists and other laboratory professionals to generate informative reports so that physicians and their patients may better understand the clinical implications of a rendered report. Catalogue of Somatic Mutations in Cancer (COSMIC) has an extensive compilation of mutations found in cancer with expert manual curation of a subset (http://cancer.sanger.ac.uk/cosmic, last accessed June 26, 2017). My Cancer Genome (www.mycancergenome.org, last accessed June 26, 2017) is a freely available, curated online knowledge base for specific mutations in different cancer types indicating the frequency and clinical significance of each mutation with supporting literature references and information about related clinical trials. The Jackson Laboratory Clinical Knowledgebase (https://www.jax.org/clinical-genomics/clinical-offerings/ckb, last accessed June 26, 2017), Weill Cornell Medicine's Precision Medicine Knowledgebase (https://pmkb.weill.cornell.edu/, last accessed June 26, 2017), and MD Anderson Cancer Center Personalized Cancer Therapy Knowledge Base for Precision Oncology (https://pct.mdanderson.org, last accessed June 26, 2017) are other curated cancer gene-related knowledge bases.

The Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists have published consensus recommendations for the interpretation and reporting of sequence variants in cancer [35]. The proposed guideline groups somatic variants with therapeutic, prognostic, diagnostic, and preventative clinical impact into four levels (A–D) based on the strength of evidence. For example, a biomarker that predicts response or residence to a FDA-approved therapy for a specific tumor type or is included in professional guidelines would be level A, whereas a biomarker that predicts response or resistance based on well-powered studies with consensus from experts in the field would be level B. These are then categorized into tiers (Table 20.2). The manuscript provides clear guidance on ways to assess the clinical significance of a variant as well and the pertinent publically available resources. The guideline does not recommend reporting tier IV (likely benign) variants.

Table 20.2 AMP/ASCO/CAP categories of somatic sequence variants

Tier I: Variants of strong clinical significance (level A or B)
Tier II: Variants of potential clinical significance (level C or D)
Tier III: Variants of unknown clinical significance
Tier IV: Benign or likely benign variants

Communicating Through Reports

Figure 20.1 is an example of a report for next-generation sequencing for germline variants associated with autism spectrum disorder using an exome panel of 30 genes. The report illustrates key features to be included to enable clear communication of results. Specifically, the report incorporates:

1. Contact information for the laboratory including a link to the laboratory website.
2. Three patient identifiers, indication for testing, the test performed, specimen type, dates of specimen collection, and receipt and date of report.
3. Abnormal results using Human Genome Variation Society (HGVS) nomenclature and ACMG recommended interpretive categories.
4. A listing of the genes tested as well as any test limitations, such as excluded exons due to high GC content (as seen in *SHANK3*, *ARX*).
5. The version of the reference genome used (e.g., hg19 (NCBI build 37)).
6. Result interpretation including an explanation of supporting evidence and clinical implications as well as the recommendation for genetic counseling.
7. Limitations of the technology (e.g., regarding detection of large deletions/duplications, repeat expansions, and structural genomic variation).
8. A statement that all pathogenic mutations were confirmed by an alternative methodology (e.g., Sanger sequencing).
9. A FDA disclaimer.
10. References supporting the test panel and the interpretation.

Some laboratories may choose to include a more detailed description of the methods utilized in the report.

Although similar features should be incorporated into cancer reports, the variant classification for somatic mutations as compared to germline variants is different. Germline variants are evaluated for the pathogenicity of the variant for a specific disease or phenotype. Somatic variants are evaluated for their impact on clinical care encompassing therapeutic, diagnostic, prognostic, or predictive biomarkers using evidence-based categorization of the

YALE-NEW HAVEN HOSPITAL
DEPARTMENT OF
YALE SCHOOL OF MEDICINE

PATHOLOGY

20 York Street, EP 2-631
New Haven, CT 06504
CT Licenses CL-0084; NY License PF18163
Phone: (203) 785-2788
Fax: (203) 785-7146

YNHH Tumor Profiling Laboratory
789 Howard Ave, CB650 CLIA ID # 07D0098656 Phone: (203) 688-5582
New Haven, CT 06504 Fax: (203) 688-5588

Tumor DNA Sequencing Report

Patient:
MR #: 1110000
DOB/Age/Sex: 12/30/1951 (Age: 65) M
*Visit #:*123456789 (Referral)
Submitting Physician: Jane Smith, M.D.

Accession #: **MP17-XX**
Taken: 1/19/2017
Accessioned: 2/13/2017 16:11 YNHH
Adm-Disch Date: 1/19/2017-01/23/17
Reported: 02/24/2017

Results

CANCER MUTATION HOTSPOT (50 GENE) SEQUENCING PANEL

Specimen information: Colonic adenocarcinoma, omental nodule, excision (YNHH S17-XXXX, Part 1, Block 2 (50% estimated malignant cells).

DNA VARIANT DETECTED	ALLELIC FRACTION
KRAS c.38G>T (p.Gly13Asp)	22%

INTERPRETATION:
KRAS, a member of the Ras family of small GTPases that mediate signal transduction downstream of growth factor receptors, plays a critical role in cell proliferation, survival and differentiation. *KRAS* variants are found in 30-40% of colorectal adenocarcinomas. According to NCCN guidelines, *KRAS* activating mutations predict a lack of response to therapy with antibodies targeted to the epidermal growth factor receptor (e.g. cetuximab, panitumumab). [NCCN Guidelines Version 2.2017 Colon Cancer] The variant allelic fraction is consistent with heterozygosity within the malignant cells in this sample, and therefore this tumor is considered positive for this mutation.

ADDITIONAL DNA VARIANTS DETECTED	
SMAD c.1245_1248delCAGA (p.Asp415GlufsTer20)	21%
TP53 c.916 C>T(p.Arg306Ter)	20%

To date, the published evidence relating these additional genetic changes to cancer biology or treatment response may be insufficient to use this information for decisions regarding patient management outside the context of clinical research.

Appendix - Genes analyzed
This test only analyzes targeted regions ("hotspots") of the exons known to be frequently mutated in cancer.

ABL1	*BRAF*	*EGFR*	*FGFR1*	*GNAQ*	*IDH2*	*KRAS*	*NPM1*	*PTPN11*	*SMO*
AKT1	*CDH1*	*ERBB2*	*FGFR2*	*GNAS*	*JAK2*	*MET*	*NRAS*	*RB1*	*SRC*
ALK	*CDKN2A*	*ERBB4*	*FGFR3*	*HNF1A*	*JAK3*	*MLH1*	*PDGFRA*	*RET*	*STK11*
APC	*CSF1R*	*EZH2*	*FLT3*	*HRAS*	*KDR*	*MPL*	*PIK3CA*	*SMAD4*	*TP53*
ATM	*CTNNB1*	*FBXW7*	*GNA11*	*IDH1*	*KIT*	*NOTCH1*	*PTEN*	*SMARCB1*	*VHL*

Fig. 20.1 Example of 50-gene hotspot panel next-generation sequencing report on a metastatic colonic adenocarcinoma. (Courtesy of Yale New Haven Hospital)

Methodology

The clinical tumor sample was enriched for malignant cells by manual microdissection of formalin-fixed, paraffin-embedded (FFPE) tissue sections. DNA was extracted using a commercially available kit (Qiagen, Inc.) and quantitated on a Qubit 2.0 fluorimeter. The isolated DNA was amplified using the Ion AmpliSeq™ Cancer Hotspot Panel v2 multiplex PCR primer set. These primers amplify 207 amplicons covering exonic regions of 50 cancer-related genes in which mutations have been reported in various types of cancer. In aggregate, ~30 kilobases of DNA sequence were amplified for subsequent nucleotide sequence analysis, which was performed on either an Ion Torrent PGM™ or an S5™ XL next generation sequencer (Thermo Fisher Scientific). The raw data generated were pre-processed within Torrent Suite, with the alignment of sequencing reads performed by the Torrent Mapping (TMAP) algorithm. Variants identified by the Ion Reporter (IR), MuTect2 and Strelka variant callers were passed through Ensembl Variant Effect Predictor (VEP) to derive annotations from multiple genomic databases. Chromosomal positions of variants refer to the human reference sequence assembly released by the Genome Reference Consortium in February 2009 (GRCh37). Unless otherwise indicated, the relative abundances (allelic fractions) of variants in tumor DNA were calculated on the basis of at least 100 sequencing reads spanning the region of DNA containing the variation. Variants corresponding to common germline polymorphisms (minor allele frequency >0.001) are not reported. The reported allelic fractions should be regarded as approximate and considered within the context of the estimated concentration of malignant cells in the material analyzed. The analytic sensitivity of this test for single nucleotide variants (SNVs) is estimated to be 5%. This assay examines tumor tissue only and does not examine normal (non-tumor) tissue. The test was designed to find gene mutations within tumors (somatic variants). It was not designed to find germline (or hereditary) mutations.

This test was developed with class I analytic specific reagents (ASRs) and its performance characteristics were determined by the Tumor Profiling Laboratory of Yale-New Haven Hospital, as required by the regulations of the Clinical Laboratory Improvement Amendments of 1988 (CLIA-1988). Pursuant to these regulations, the ASRs used in this test have been established and verified for accuracy and precision. This test has not been cleared or approved by the U.S. Food and Drug Administration (FDA). The FDA has determined that such clearance or approval is not necessary for clinical application. This test is used for clinical purposes and should not be regarded as investigational or for research. Testing was performed by the Yale Tumor Profiling Laboratory, 789 Howard Avenue, CB650, New Haven, CT 06504. Tel: 203-688-5582. This laboratory is certified under the requirements of CLIA-1988 as qualified to perform high complexity clinical testing. Additional information about the tests performed by the laboratory is available upon request. CLIA ID # 07D0098656; State of Connecticut License # HP-0211.

Fig. 20.1 (continued)

variants [35]. Figure 20.2 is an example of a report on a "tumor-only" 50-gene next-generation sequencing cancer hotspot panel performed on an advanced-stage colonic adenocarcinoma. The primary finding is a *KRAS* codon 13 variant at an allelic fraction of 22% in a specimen with 50% estimated malignant cells (i.e., a heterozygous variant). According to the AMP/ASCO/CAP guidelines, the variant is tier I with level A therapeutic significance as it predicts resistance to anti-epidermal growth factor receptor monoclonal antibodies as noted in the NCCN professional guidelines (https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf, last accessed June 30, 2017). The TP53 terminating variant is present in ~200 tissue samples in COSMIC, listed as pathogenic in ClinVar, and is predicted to eliminate the P53 tetramerization domain. The active conformation of p53 when it binds DNA is tetrameric [36]. The *SMAD4* frameshift variant is pathogenic in ClinVar associated in the germline with juvenile polyposis syndrome. It is not in COSMIC, but 3 terminating variants in the same region are included. SMADs are transcription factors that transduce TGF-β ligand signaling to the nucleus. Somatic mutations in SMAD4 frequently occur in the C-terminal Mad homology 2 (MH2) region (residues 323–552) like this one, which results in premature termination of the 553 amino acid protein. Mutations in the MH2 domain interfere with Smad4 homo-oligomers and Smad4/Smad2 hetero-oligomerization resulting in the disruption of TGF-β signaling. *SMAD* variants may have prognostic significance in colorectal carcinoma (tier II) [37].

To assist in filtering germline SNPs in the analysis of larger gene panels or WES/WGS of cancer tissue, patient's germline DNA is analyzed simultaneously to properly identify the unique cancer somatic mutations and not overcall germline SNPs as somatic variants. This approach also identifies germline variants that may be associated with cancer predisposition syndromes that can impact the care of the patient and identify family members for screening. Clinical laboratories need to develop policies about identifying and reporting clinically important germline variants and have institutional policies for consenting procedures to report germline vari-

sema4
a Mount Sinai venture

1428 Madison Ave, Atran Bldg
Rm2-25, New York, NY 10029-6574
Tel: 800-298-6470 Fax: 212-241-0139
www.sema4genomics.com
CLIA # 33D2097541

MOLECULAR GENETICS

Patient Name:
Date of Birth: **1/1/2015**
Reference #: **P123456**
Indication: **Developmental delay**
Test Type: **Autism NGS Sequencing Panel**
Specimen Type: **Blood**
Lab #: **17123456AU**
Date Collected: **1/12/2017**
Date Received: **1/12/2017**

Final Report: **2/8/2017**

Referring:
Great Doctor, M.D.
Mount Sinai Hospital
One Gustave L. Levy Place
New York, NY 10029
Fax: 212-111-1111

RESULTS AND INTERPRETATION

Result: A *de novo*, heterozygous (one copy), likely pathogenic variant c.5726_5727delCT, p.S1909* was detected in the *NSD1* gene.

GENE*	RESULTS
NRXN1 NM_001135659.1, NM_004801.4, NM_138735.2	No clinically significant variants detected
***NSD1* NM_022455.4**, NM_172349.2	**Variant detected: c.5726_5727delCT, p.S1909***
AHI1 NM_001134830.1, NM_001134831.1, NM_001134832.1, NM_017651.4	No clinically significant variants detected
CNTNAP2 NM_014141.5	No clinically significant variants detected
TSC1 NM_000368.4, NM_001162426.1, NM_001162427.1	No clinically significant variants detected
PTEN NM_000314.4 *plus* a portion of the 5' UTR (chr10:89623220-89623484)	No clinically significant variants detected
SHANK2 NM_012309.3	No clinically significant variants detected
DHCR7 NM_001163817.1, NM_001360.2	No clinically significant variants detected
CACNA1C only exon 8 (chr12:2613597-2613710 and chr12:2614003-2614116)	No clinically significant variants detected
UBE3A NM_000462.3, NM_130838.1, NM_130839.2	No clinically significant variants detected
TSC2 NM_000548.3, NM_001077183.1, NM_001114382.1	No clinically significant variants detected
SHANK3 NM_033517.1 *except* for two exons plus splice sites (chr22:51113470-51113684 and chr22:51135986-51136148)**	No clinically significant variants detected
NLGN4X NM_020742.2, NM_181332.1	No clinically significant variants detected
AP1S2 NM_003916.3	No clinically significant variants detected
CDKL5 NM_001037343.1, NM_003159.2	No clinically significant variants detected
PTCHD1 NM_173495.2	No clinically significant variants detected
ARX NM_139058.2 *except* for a portion of exon 2 (chrX:25031504-25031920)***	No clinically significant variants detected
IL1RAPL1 NM_014271.3	No clinically significant variants detected
OTC NM_000531.5	No clinically significant variants detected
KDM5C NM_001146702.1, NM_004187.3	No clinically significant variants detected
OPHN1 NM_002547.2	No clinically significant variants detected
PCDH19 NM_001105243.1, NM_020766.2, NM_001184880.1	No clinically significant variants detected
UPF3B NM_023010.3, NM_080632.2	No clinically significant variants detected
GRIA3 NM_000828.4, NM_007325.4	No clinically significant variants detected
GPC3 NM_001164617.1, NM_001164618.1, NM_001164619.1, NM_004484.3	No clinically significant variants detected
SLC9A6 NM_001042537.1, NM_001177651.1, NM_006359.2	No clinically significant variants detected
FMR1 NM_001185075.1, NM_001185076.1, NM_001185081.1, NM_001185082.1, NM_002024.5	No clinically significant variants detected
SLC6A8 NM_001142805.1, NM_001142806.1, NM_005629.3	No clinically significant variants detected
MECP2 NM_001110792.1, NM_004992.3 *plus* a portion of the 3' UTR (chrX:153295704- 153295748)	No clinically significant variants detected
RAB39B NM_171998.2	No clinically significant variants detected

*All coding DNA sequence of the genes corresponding to the transcripts listed plus the flanking 5 base pair splice sites are sequenced relative to the hg19 assembly. Exceptions are noted.
** Only a single variant has been reported as pathogenic in either of these two excluded exons in "The Human Gene Mutation Database" (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php, December 2012, Gauthier *et al.*). Please note that this region is excluded due to inconsistency in sequencing result quality due to high GC content.

Fig. 20.2 Example of a next-generation sequencing report for germline mutations associated with autism spectrum disorder using an exome panel of 30 genes. (Courtesy of sema4 laboratory, a Mount Sinai Health System venture)

***Twelve percent of all different *ARX* sequence mutations (20 bp or less) reported in HGMD (December 2012) are located within this region. The region also contains the poly Alanine tracts affected by recurrent mutations (Nawara et al. and Stromme et al.). Therefore this test will not detect poly Alanine tract mutations in this gene as well as other mutations in this region. Please note that this area is excluded due to inconsistency in sequencing result quality due to high GC content.

Interpretation: Next generation sequencing of a panel of 30 genes that have been associated with Autism Spectrum Disorders (ASDs) was performed on DNA extracted from the peripheral blood specimen of this patient. A heterozygous (one copy) variant was detected in the *NSD1* gene (NM_022455.4): c.5726_5727delCT, p.S1909* (hg19, chr5:176707669-176707670). It results in a stop codon at position 1909 in exon 19 of the gene and is predicted to cause loss of normal protein either through protein truncation or nonsense-mediated mRNA delay. To our knowledge, it is not a previously reported mutation in the *NSD1* gene. It is not listed in the ExAC population database. Analysis of parental DNA samples indicated that the *NSD1* variant c.5726_5727delCT, p.S1909* was not present in the parental blood samples thus it is a *de novo* change in the patient. A missense change at the same position (p.S1909P) in the *NSD1* gene has been reported in one patient with Sotos syndrome (Pohjola 2012). Based on this evidence, this variant is considered to be a likely pathogenic variant. Mutations in *NSD1* cause Sotos syndrome 1(MIM# 117550), an autosomal dominant overgrowth condition characterized by a typical facial appearance, learning disabilities, and, in some patients, ASD and additional congenital anomalies.

Genetic counseling and correlation with the clinical phenotype of this patient are recommended. Please note that parental DNA samples were analyzed solely for the presence of the above described variants.

This technology may not detect all small insertion/deletions and is not diagnostic for large duplications/deletions, repeat expansions, and structural genomic variation. This test will only detect variants within the exons and the intron-exon boundaries of the target genes as listed in the report table. Variants outside these regions will not be detected. These regions include, but are not limited to, UTRs, promoters, and deep intronic areas. In addition, a mutation(s) in a gene not included on the panel could be present in this patient. Although each of the genes on the panel is a rare cause of ASD, this panel is expected to detect 5-10% of mutations present in ASD patients. The sensitivity of this panel is estimated at 99% for single base substitutions and 97% at the level of a few base-pairs. All potentially pathogenic variants were subjected to Sanger sequencing for confirmation of the result. Any benign polymorphisms identified during this analysis were not reported.

Comments: Please note this test was developed and its performance characteristics were determined by The Mount Sinai Genomics, Inc and were considered acceptable for patient testing. It has not been cleared or approved by the FDA. The FDA has determined that such clearance or approval is not necessary.

This type of mutation analysis generally provides highly accurate genotype information for point mutations and single nucleotide polymorphisms. Despite this level of accuracy, it should be kept in mind that there are many potential sources of diagnostic error, including misidentification of samples, polymorphisms, or other rare genetic variants that interfere with analysis. In addition, families should understand the limitations of the testing and that rare diagnostic errors may occur for the reasons described.

References:

1. BS Abrahams and DH Geschwind. Advances in autism genetics: on the threshold of a new neurobiology. Nature. 9(5):341-355, 2008.
2. Addington AM et al. A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders. Mol Psychiatry. 16(3):238-9, 2011.
3. Adegbola A et al. A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). Am J Med Genet A. 146A(4):505-11, 2008.
4. Archer HL et al. CDKL5 mutations cause infantile spasms, early onset seizures, and severe mental retardation in female patients. J Med Genet. 43(9):729-34, 2006.
5. Bakkaloglu B et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet. 82(1):165-73, 2008.
6. Berkel S, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet. 42(6):489-91, 2010.
7. Borck G et al. Clinical, cellular, and neuropathological consequences of AP1S2 mutations: further delineation of a recognizable X-linked mental retardation syndrome. Hum Mutat. 29(7):966-74,2008.
8. Buxbaum JD et al. Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly. Am J Genet B Neuropsychiatr Genet. 144B(4):484-91, 2007.
9. Chabrol B et al. Delineation of the clinical phenotype associated with OPHN1 mutations based on the clinical and neuropsychological evaluation of three families. Am J Med Genet A. 138(4):314-7, 2005.
10. Cohen D et al. Specific genetic disorders and autism: clinical contribution towards their identification. J Autism Dev Disord. 35(1):103-6, 2005.
11. Dibbens LM et al. X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. Nat Genet. 40(6):776-81, 2008.
12. Durand CM et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet. 39(1)25-7, 2007.
13. Gauthier J et al. Novel de novo SHANK3 mutation in autistic patients. Am J Med Genet B Neuropsychiatr Genet. 150B(3):421-4, 2009.
14. Giannandrea M et al. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. Am J Hum Genet. 86(2):185-95, 2010.
15. Guilmatre A et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch Gen Psychiatry. 66(9):957-56, 2009.
16. Herman GE et al. Increasing knowledge of PTEN germline mutations: Two additional patients with autism and macrocephaly. Am J Med Genet A. 143(6):589-93, 2007.
17. Jamal SM et al. Novel de novo PCDH19 mutations in three unrelated females with epilepsy female restricted mental retardation syndrome. Am J Med Genet A. 152A(10):2475-81, 2010.
18. Kielinen M, Rantala H, Timonen E, Linna SL, Moilanen I. Associated medical disorders and disabilities in children with autistic disorder: a population-based study.

Fig. 20.2 (continued)

Autism. 8:49-60, 2004.
19. Kim HG et al. Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet. 82(1):199-207, 2008.
20. Laumonnier F et al. Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. Mol Psychiatry. 15(7):767-776, 2010.
21. Laumonnier, F et al. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. Am J Hum Genet. 74(3):552-7, 2004.
22. Liao P and Soong TW. CaV1.2 channelopathies: from arrhythmias to autism, bipolar disorder, and immunodeficiency. Pflugers Arch. 460(2):353-9, 2010.
23. Marco EJ and Skuse DH. Autism-lessons from the X chromosome. Soc Cogn Affect Neurosci. 1(3):183-93, 2006.
24. Marshall CR et al. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet. 82(2):477-88, 2008.
25. Metzker, ML. Sequencing technologies - the next generation. Nature Reviews, Genetics. 11(1): 31-46, 2010.
26. Miles JH, Hillman RE. Value of a clinical morphology examination in autism. Am J Med Genet. 91:245-253, 2000.
27. Morrow JD, Whitman BY, and Accardo PJ. Autistic disorder in Sotos syndrome: a case report. Eur J Pediatr. 149(8):567-9, 1990.
28. Nawara M et al. The ARX mutations: a frequent cause of X-linked mental retardation. Am J Med Genet A. 140(7):727-32, 2006.
29. Neri G et al. Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. Am J Med Genet A. 79(4):279-83, 1998.
30. Noor A et al. Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci Transl Med. 2(49):49-68, 2010.
31. Ozonoff S, Williams BJ, Gale S, Miller JN. Autism and autistic behavior in Joubert syndrome. J Child Neurol. 14(10):636-41, 1999.
32. Pinto D et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 466(7304):368-72, 2010.
33. Piton A et al. Mutations in the calcium-related gene IL1RAPL1 are associated with autism. Hum Mol Genet. 17(24):3965-74,2008.
34. Piton A et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol Psychiatry. 16(8):867-80, 2011.
35. Schroer RJ et al. Natural history of Christianson syndrome. Am J Med Genet A. 152A(11):2775-83, 2010.
36. Serrano M et al. Neuropsychiatric manifestations in late-onset urea cycle disorder patients. J Child Neurol. 25(3):352-8, 2010.
37. Straus KA eta al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med. 354(13):1370-7, 2006.
38. Stromme P et al. Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet. 30(4):441-5, 2002.
39. Takahashi Y et al. A loss-of-function mutation in the SLC9A6 gene causes X-linked mental retardation resembling Angelman syndrome. Am J Med Genet B Neuropsychiatr Genet. 156B(7)799-807, 2011.
40. Tarpey PS et al. Mutations in the gene encoding the Sigma 2 subunit of the adaptor protein 1 complex, AP1S2, cause X-linked mental retardation. Am J Hum Genet. 79(6):1119-24,2006.
41. Tzschach A et al. Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. Hum Mutat. 27(4):389, 2006.
42. Vissers LE et al. A de novo paradigm for mental retardation. Nat Genet. 42(12):1109-12, 2010.
43. Weaving LS et al. Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet. 75(6):1079-93, 2004.
44. Wu Y et al. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. Proc Natl Acad Sci USA. 104(46)18163-8, 2007.
45. Yan J et al. Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. Mol Psychiatry. 10(4):329-32, 2005.
46. Zahir FR et al. A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. J Med Genet. 45(4):239-43, 2008.
47. Zanni G et al. Oligophrenin 1 mutations frequently cause X-linked mental retardation with cerebellar hypoplasia. Neurology. 65(9):1364-9, 2005.
48. Pohjola P et al. Translation of a research-based genetic test on a rare syndrome into clinical service testing, with Sotos syndrome as an example. Genet Test Mol Biomarkers. 16(10):1188-94, 2012.

This case has been reviewed and electronically signed by xxx xxx, Ph.D., FACMG, Assistant Laboratory Director
Laboratory Medical Consultant: xxx xxx, M.D., Ph.D.

Fig. 20.2 (continued)

ants. Figure 20.3 is an example of a cancer panel test report that includes paired tumor and normal samples and documents a germline variant associated with an autosomal dominant familial cancer predisposition syndrome [38]. Our pathology, oncology, and genetics departments have a policy and procedure for referring patients with suspected germline cancer predisposition variants for genetic counseling.

The static reports illustrated in this chapter are only examples and necessarily have limitations. Static reports only reflect the knowledge available at the time the report was generated and do not provide an easy mechanism to update and notify the clinician and patient as new information becomes available. There is also no payment structure in place for laboratories to develop and execute this responsibility. In addition, as more and more information becomes available through next-generation sequencing, both the clinician and the patient will need more time to "digest" the laboratory results. It is possible that a more dynamic model incorporating interactive electronic reporting with one-on-one counseling will evolve.

YALE-NEW HAVEN HOSPITAL
DEPARTMENT OF
YALE SCHOOL OF MEDICINE

PATHOLOGY

20 York Street, EP 2-631
New Haven, CT 06504
CT Licenses CL-0084; NY License PF18163
Phone: (203) 785-2788
Fax: (203) 785-7146

YNHH Tumor Profiling Laboratory
789 Howard Ave, CB650
New Haven, CT 06504
CLIA ID # 07D0098656
Phone: (203) 688-5582
Fax: (203) 688-5588

Oncomine Gene Panel Report

Patient: **James Smith**
MR #: 1220000
DOB/Age/Sex: 10/11/1969 (Age: 47) M
*Visit #:*123456799 (Referral)
Submitting Physician: Mary Brown, M.D.

Accession #: **MP17-XY**
Taken: 2/22/2017
Accessioned: 2/22/2017 14:16
Adm-Disch Date: 2/22/17
Reported: 03/15/2017

Results

Targeted NGS Cancer Gene Panel (Oncomine Assay)

Tumor specimen analyzed: Outside Hospital, Town, CT, S16-XXXX, Part 1, Block 1
Site / Organ: Liver mass, biopsy
Histology: Metastatic spindle cell melanoma (history of ocular melanoma)
Percent malignant cells (after manual microdissection): 90%
Normal (germline) control specimen: buccal swab

This assay examines tumor DNA for mutations and/or amplifications in 134 cancer-related genes. This assay also examines tumor RNA for the presence of 271 fusion transcripts involving 24 oncogenic driver genes.

No somatic single nucleotide or multi-nucleotide variants were detected in the tumor.

Pathogenic variant detected in both the tumor and normal (germline) tissue

Variant	Allelic Fraction (Tumor)	Allelic Fraction (Normal)
***BAP1* p.Phe171Ter**	85%	48%

Variants are reported as their predicted amino acid changes. Allelic fraction indicates the percentage of DNA sequencing reads containing the variant and has not been corrected for the estimated fraction of tumor cells in the sample or for gene copy number within tumor cells. Double asterisks (**) indicate variants that may be amenable to targeted therapy; see Interpretation.

No gene amplifications (copy number ≥5) were detected in the tumor.

No gene fusions were detected in the tumor

Interpretation

Gene name: ***BAP1*** (BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase))
Variant (protein): p.Phe181Ter
Variant (coding DNA): c.542_543delTT (ENST00000460680)
Predicted effects: **Damaging; loss of tumor suppressor function**
Previously reported: No
Additional notes: The tumor suppressor *BAP1* encodes a deubiquitinase that has been found associated with multiprotein complexes that regulate key cellular processes such as the cell cycle, cellular differentiation, cell death, gluconeogenesis, and the DNA damage response. Germline mutations in this gene may be associated with an autosomal dominant familial cancer predisposition syndrome with increased risk of cancers including malignant mesothelioma, uveal melanoma, cutaneous melanoma,

Fig. 20.3 Example of a report from a next-generation sequencing multiple biomarker assay including single-nucleotide variants, indels, copy number variation, and gene fusions on a metastatic ocular melanoma and a paired normal specimen (buccal swab). (Courtesy of Yale New Haven Hospital)

lung carcinoma, meningioma and renal cell carcinoma. [Carbone M. et al. BAP1 and Cancer. *Nat Rev Cancer*. 2013;13(3):153-59]. In addition, somatic mutations in *BAP1* have been detected in sporadic tumors of multiple types. [Murali R. et al. Tumors associated with BAP1 mutations. *Pathology*. 2013;45:116-26]. The *BAP1* variant detected in this tumor and germline is predicted to truncate the BAP1 protein at amino acid 181 (of 729), thereby eliminating both the nuclear localization signal and the site of interaction with the BRCA1 protein negatively impacting its tumor suppressor function. The variant allele fraction (85%) in the tumor DNA is consistent with loss of heterozygosity at the *BAP1* locus. Given that the pathogenic *BAP1* variant was detected in DNA from the patient's non-tumor tissue, genetic counseling is strongly recommended for this patient.

Appendix

Genes analyzed in this test

ABL1[14]	*BIRC3*[3]	*CSNK2A1*[3]	*FGFR1*[134]	*IFITM1*[1]	*MCL1*[3]	*NFE2L2*[1]	*PPP2R1A*[1]	*SOX2*[3]
ACVRL1[3]	*BRAF*[14]	*CTNNB1*[1]	*FGFR2*[134]	*IFITM3*[1]	*MDM2*[3]	*NKX2-1*[3]	*PTCH1*[2]	*SPOP*[1]
AKT1[13]	*BRCA1*[2]	*DCUN1D1*[3]	*FGFR3*[134]	*IGF1R*[3]	*MDM4*[3]	*NKX2-8*[3]	*PTEN*[2]	*SRC*[1]
AKT3[4]	*BRCA2*[2]	*DDR2*[1]	*FGFR4*[3]	*IL6*[3]	*MED12*[1]	*NOTCH1*[2]	*PTPN11*[1]	*STAT3*[1]
ALK[14]	*BTK*[1]	*DNMT3A*[1]	*FLT3*[13]	*JAK1*[1]	*MET*[134]	*NPM1*[1]	*RAC1*[1]	*STK11*[2]
APC[2]	*CBL*[1]	*EGFR*[134]	*FOXL2*[1]	*JAK2*[1]	*MLH1*[1]	*NRAS*[1]	*RAF1*[14]	*TERT*[3]
APEX1[3]	*CCND1*[3]	*ERBB2*[134]	*GAS6*[3]	*JAK3*[1]	*MPL*[1]	*NTRK1*[4]	*RB1*[2]	*TET2*[2]
AR[13]	*CCNE1*[3]	*ERBB3*[1]	*GATA2*[1]	*KDR*[1]	*MSH2*[2]	*NTRK2*[4]	*RET*[14]	*TIAF1*[3]
ARAF[1]	*CD274*[3]	*ERBB4*[1]	*GATA3*[2]	*KIT*[13]	*MTOR*[1]	*NTRK3*[4]	*RHEB*[1]	*TP53*[2]
ATM[2]	*CD44*[3]	*ERG*[4]	*GNA11*[1]	*KNSTRN*[1]	*MYC*[3]	*PAX5*[1]	*RHOA*[1]	*TSC1*[2]
ATP11B[3]	*CDH1*[2]	*ESR1*[1]	*GNAQ*[1]	*KRAS*[13]	*MYCL*[3]	*PDCD1LG2*[3]	*ROS1*[4]	*TSC2*[2]
AXL[4]	*CDK4*[134]	*ETV1*[4]	*GNAS*[1]	*MAGOH*[1]	*MYCN*[3]	*PDGFRA*[134]	*RPS6KB1*[3]	*U2AF1*[1]
BAP1[2]	*CDK6*[3]	*ETV4*[4]	*HNF1A*[1]	*MAP2K1*[1]	*MYD88*[1]	*PIK3CA*[13]	*SF3B1*[1]	*VHL*[2]
BCL2L1[3]	*CDKN2A*[2]	*ETV5*[4]	*HRAS*[1]	*MAP2K2*[1]	*MYO18A*[3]	*PIK3R1*[2]	*SMAD4*[2]	*WT1*[2]
BCL9[3]	*CHEK2*[1]	*EZH2*[1]	*IDH1*[1]	*MAPK1*[1]	*NF1*[2]	*PNP*[3]	*SMARCB1*[2]	*XPO1*[1]
BIRC2[3]	*CSF1R*[1]	*FBXW7*[2]	*IDH2*[1]	*MAX*[1]	*NF2*[2]	*PPARG*[34]	*SMO*[1]	*ZNF217*[3]

[1]Genes in which only mutational hotspot regions are analyzed ("hotspot genes").
[2]Genes in which the full coding sequence is analyzed (tumor suppressor genes)
[3]Genes assessed for amplification.
[4]Genes assessed for fusion ("fusion driver genes").

Methodology

The clinical tumor sample was enriched for malignant cells by manual microdissection of formalin-fixed, paraffin-embedded (FFPE) tissue sections. Germline control DNA from the same patient was obtained either from FFPE non-tumor tissue, from the patient's blood, or from a buccal swab. DNA and RNA were isolated from the patient's tumor, and DNA was isolated from the germline control sample. DNA and RNA extractions were done using a commercially available kit (Qiagen, Inc.) and nucleic acids were quantitated on a Qubit 2.0 fluorimeter. The isolated DNA was amplified using a modification of the Ion AmpliSeq™ Oncomine Comprehensive Assay multiplex PCR primer sets. These primers amplify about 2,500 amplicons within 134 genes (see table above). In aggregate, about 0.24 megabase of unique sequence from each DNA sample (tumor and control) was amplified and bar-coded. Amplicon sequencing was performed on either an Ion Torrent PGM™ or an Ion S5™ XL next generation sequencer (ThermoFisher Scientific, Inc.). The raw data generated were pre-processed within Torrent Suite, with the alignment of sequencing reads performed by the Torrent Mapping (TMAP) algorithm. Variants identified by the Ion Reporter (IR), MuTect2 and Strelka variant callers were passed through Ensembl Variant Effect Predictor (VEP) to derive annotations from multiple genomic databases. The IR copy number variation (CNV) detection algorithm was used to detect gene amplifications by comparing variations in read depth between the tumor and normal control across the target regions of the assay. If a suitable normal (germline) control was unavailable at the time of testing, a bioinformatically pooled normal appropriate for the patient's gender was selected. Chromosomal positions of variants refer to the human reference sequence assembly released by the Genome Reference Consortium in February 2009 (GRCh37). Variants corresponding to common polymorphisms (minor allele frequency >0.001) and found in both the tumor DNA and the germline control DNA are not reported. Unless otherwise indicated, the reported allelic fractions (i.e., abundance, relative to non-mutant sequence) of variants in tumor DNA were calculated on the basis of at least 50 sequencing reads spanning the region containing the variation. These reported variant allelic fractions should be regarded as approximate and considered within the context of the estimated concentration of malignant cells in the material analyzed. The analytic sensitivity (lower limit of detection) of this test for single nucleotide variants (SNVs) and insertions/deletions (indels) is estimated to be 5%.

Note on Interpretation and Reporting (Disclaimer): Somatic variants are listed in order of decreasing abundance within tumor DNA, without regard to clinical significance or actionability. These variants were detected at allelic fractions of ≥20% sequencing reads obtained from tumor DNA. The gene amplifications listed were identified as chromosomal regions with evidence suggestive of somatic copy number gain in DNA isolated from the tumor. These gene amplifications were detected at estimated gene copy number levels of ≥5. The described effects of the reported mutations on proteins and pathways, as well as their possible impact on the sensitivity or resistance of a tumor to targeted therapies, are based on published studies (clinical and/or pre-clinical) with

Fig. 20.3 (continued)

varying levels of evidence and with varying degrees of relevance to the patient whose sample is analyzed here. The predicted effects of amino acid substitutions on protein structure/function are computed using the SIFT and Polyphen bioinformatic tools. SIFT predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between the alternate amino acids. Polyphen predicts the effect of an amino acid substitution on the structure and function of a protein using sequence homology, Pfam annotations, 3D structures from PDB where available, and a number of other databases and tools. Any information provided about therapies may not be comprehensive and is for informational purposes only. This report does not offer treatment recommendations, but rather is intended to provide the treating physician with information that may be considered in conjunction with all other relevant data and in this patient's particular clinical context.

This test was developed with class I analytic specific reagents (ASRs) and its performance characteristics were determined by the Tumor Profiling Laboratory of Yale-New Haven Hospital, as required by the regulations of the Clinical Laboratory Improvement Amendments of 1988 (CLIA-1988). Pursuant to these regulations, the ASRs used in this test have been established and verified for accuracy and precision. This test has not been cleared or approved by the U.S. Food and Drug Administration (FDA). The FDA has determined that such clearance or approval is not necessary for clinical application. This test is used for clinical purposes and should not be regarded as investigational or for research. Testing was performed by the Yale Tumor Profiling Laboratory, 789 Howard Avenue, CB650, New Haven, CT 06504. Tel: 203-688-5582. This laboratory is certified under the requirements of CLIA-1988 as qualified to perform high complexity clinical testing. Additional information about the tests performed by the laboratory is available upon request. CLIA ID # 07D0098656; State of Connecticut License # HP-0211.

Fig. 20.3 (continued)

The Role of the Pathologist

These are exciting and unsettled times as pathology navigates how best to incorporate genomics into clinical practice. Pathologists have always been at the interface of collecting, analyzing, interpreting, and integrating data to effectively communicate clinically relevant results to clinicians and their patients. Pathologists are the direct link between medical data and clinical care. Next-generation sequencing is a disruptive technology for the practice of pathology. A comprehensive test with big data sets will replace a series or panel of targeted tests yet still requires quality assurance, analysis, interpretation, and effective communication. Historically, pathology has impacted clinical care through the adoption of new technologies that reveal relevant clinical correlates, from the light microscope to the electron microscope through to immunohistochemistry, auto-analyzers, flow cytometry, polymerase chain reaction, mass spectroscopy, and more. Next-generation sequencing should not change this primary role. Now is the time for us to embrace genomics and firmly engage so that we properly guide and lead its incorporation into clinical care. As a corollary, if technology and promise are the drivers, we must be responsible, cautious, and ethical in reporting results in the context of each patient.

Pathologists play a key role in cancer diagnosis using morphology integrated with ancillary techniques (protein expression, cytogenetics, FISH, molecular diagnostics). This role continues with next-generation sequencing. Pathologists understand the importance of the pre-analytical variables of tissue collection and processing on nucleic acid extraction and sequencing accuracy and can optimize this workflow process. Pathologists' morphologic skills are essential for selecting the best area of tumor tissue to ensure specimen adequacy and sufficient percentage of tumor cells to detect mutant alleles in heterogeneous samples mixed with benign tissue elements. To fully communicate next-generation sequencing results, pathologists need to not only master the evolving molecular classification of cancer but also need to know related biologic pathways, targets for drug therapy, changes related to drug resistance, and prognostic biomarkers.

There are several ways in which pathologists can overcome challenges and barriers to readily implementing next-generation sequencing reporting [39]:

1. Integrate bioinformaticists and programmers into the reporting team. Next-generation sequencing data analysis algorithms need programming expertise together with specialized servers to handle and store all of the data. We need standardized bioinformatics pipelines for refinement in base calling and annotation of identified variants. The reporting physicians must work alongside the bioinformaticists to understand the analysis pipeline and be cognizant of potential pitfalls. The development of smaller, lower throughput sequencing instruments may provide a better entry for the pathology team to develop confidence and expertise in the details of the pipeline. Clinical decision support systems and knowledge databases also need to be incorporated to generate informative reports.
2. Create interdisciplinary teams of pathologists, geneticists, oncologists, and translational researchers to build consensus on patient-specific variant interpretation and reporting.
3. Advocate for certified clinical grade annotated variant databases and knowledge bases.
4. Develop strategies for reflex next-generation sequencing tests on cancer specimens to meet professional guidelines as there is variability in oncologists ordering practices.

5. Develop and utilize structured training programs and curricula to provide pathologists and pathology trainees with genomic literacy and skills in interpreting and reporting next-generation sequencing data.

Conclusions

Medicine has entered the genomic era in which vast numbers of genetic variants will be incorporated into the criteria for the diagnosis, prognosis, and treatment of disease. Successful implementation requires truly informative and accessible clinical reports to guide physicians and their patients. Pathologists should be at the forefront of this transformation.

References

1. Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M, Palomaki GE, Popovich BW, Pratt VM, Rohlfs EM, Strom CM, Richards CS, Witt DR, Grody WW. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. Genet Med. 2004;6(5):387–91.
2. Teekakirikul P, Kelly MA, Rehm HL, Lakdawala NK, Funke BH. Inherited cardiomyopathies: Molecular genetics and clinical genetic testing in the postgenomic era. J Mol Diagn. 2013;15(2):158–70.
3. Burke MA, Cook SA, Seidman JG, Seidman CE. Clinical and Genetic Insights into the Genetics of Cardiomyopathy. J Am Coll Cardiol. 2016;68(25):2871–86.
4. Cardarella S, Ortiz TM, Joshi VA, Butaney M, Jackman DM, Kwiatkowski DJ, Yeap BY, Janne PA, Lindeman NI, Johnson BA. The introduction of systemic genomic testing for patients with non-small-cell lung cancer. J Thorac Oncol. 2012;7(12):1767–74.
5. Sequist LV, Heist RS, Shaw AT, Fidias P, Rosovsky R, Temel JS, Lennes IT, Digumarthy S, Waltman BA, Bast E, Tammireddy S, Morrissey L, Muzikansky A, Goldberg SB, Gainor J, Channick CL, Wain JC, Gaissert H, Donahue DM, Muniappan A, Wright C, Willers H, Mathisen DJ, Ellisen LW, Mino-Kenudson M, Lanuti M, Borger DR, Iafrate AJ, Engelman JA, Dias-Santagata D. Implementing multiplexed genotyping of non-small cell lung cancers into routine clinical practice. Ann Oncol. 2011;22(12):2616–24.
6. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, Cheng DT, Chandramohan R, Mohanty AS, Ptashkin RN, Jayakumaran G, Prasad M, Syed MH, Rema AB, Liu ZY, Nafa K, Borsu L, Sadowska J, Casanova J, Bacares R, Kiecka IJ, Razumova A, Son JB, Stewart L, Baldi T, Mullaney KA, al-Ahmadie H, Vakiani E, Abeshouse AA, Penson AV, Jonsson P, Camacho N, Chang MT, Won HH, Gross BE, Kundra R, Heins Z, Chen HW, Phillips S, Zhang H, Wang J, Ochoa A, Wills J, Eubank M, Thomas SB, Gardos SM, Reales DN, Galle J, Durany R, Cambria R, Abida W, Cercek A, Feldman DR, Grounder MM, Hakimi AA, Harding JJ, Iyer G, Janjigian YY, Jordan EJ, Kelly CM, Lowery MA, LGT M, Omuro AM, Raj N, Razavi P, Shoushtari AN, Shukla N, Soumerai TE, Varghese AM, Yaeger R, Coleman J, Bochner B, Riely GJ, Saltz LB, Scher HI, Sabbatini PJ, Robson ME, Klimstra DS, Taylor BS, Baselga J, Schultz N, Hyman DM, Arcila ME, Sofit DB, Ladanyi M, Berger MF. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23(6):703–13.
7. Hartmaier RJ, Albacker LA, Chmielecki J, Bailey M, He J, Goldberg ME, Ramkissoon S, Suh J, Elvin JA, Chiacchia S, Frampton GM, Ross JS, Miller V, Stephens PJ, Lipson D. High-throughput genomic profiling of adult solid tumors reveals novel insights into cancer pathogenesis. Cancer Res. 2017;77(9):2464–75.
8. Gulley ML, Braziel RM, Halling HC, Hsi ED, Kant JA, Nikiforova MN, Nowak JA, Ogino S, Oliveira A, Polesky HF, Silverman L, Tubbs RR, Van Deerlin VM, Vance GH, Versalovic J, Molecular Pathology Resource Committee, College of American Pathologists. Clinical laboratory reports in molecular pathology. Arch Pathol Lab Med. 2007;131(6):852–63.
9. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 2015;43(Database issue):D1079–85.
10. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE. Tascher PEM on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO). HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat. 2016;37(6):564–9.
11. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248–9.
12. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452–257.
13. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.
14. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody W, Hedge M, Lyon E, Spector E, Voelkerding K, Rehm H, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.
15. Drmanac R. The advent of personal genomic sequencing. Genet Med. 2011;13(3):188–90.
16. Lek M, Karczewski KJ, Minikei EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birmbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gautheir L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG. Exome Aggregation Consortium. Nature. 2016;536:285–91.
17. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med. 2017;9(1):13.
18. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleci Acids Res. 2017;45(D1):D12–7.
19. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Rehm HL, Watson MS,

Williams MS, Biesecker LG, American College of Medical and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565–74.
20. ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet Med. 2015;17(1):68–9.
21. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SK, Klein TE, Korf BR, McKelvey KD, Ormand KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19(2):249–55.
22. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. Genet Med. 2017;19(2):209–14.
23. Biesecker LG. Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. Genet Med. 2012;14(4):393–8.
24. Arber DA, Brunning RD, Le Beau MM, Falini B, Vardiman JW, Porwit A, Thiele J, Bloomfield CD. Acute myeloid leukaemia with recurrent genetic abnormalities. In: Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman J, editors. WHO classification of tumours of Haematopoietic and Lymphoid tissues. Lyon: International Agency for Research on Cancer; 2008.
25. Lynch TJ, Bell DS, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129–39.
26. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500.
27. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A. 2004;101(36):13306–11.
28. Dias-Santagata D, Akhavanfard S, David SS, Vernovsky K, Kuhlmann G, Boisvert SL, Stubbs H, McDermott U, Settleman J, Kwak EL, Clark JW, Isakoff SJ, Sequist LV, Engelman JA, Lynch TJ, Haber DA, Louis DN, Ellisen LW, Borger DR, Iafrate AJ. Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. EMBO Mol Med. 2010;2(5):146–58.
29. MacConaill LE, Campbell CD, Kehoe SM, Bass AJ, Hatton C, Niu L, Davis M, Yao K, Hanna M, Mondal C, Luongo L, Emery CM, Baker AC, Philips J, Goff DJ, Fiorentino M, Rubin MA, Polyak K, Chan J, Wang Y, Fletcher JA, Santagata S, Corso G, Roviello F, Shivdasani R, Kieran MW, Ligon KL, Stiles CD, Hahn WC, Meyerson ML, Garraway LA. Profiling critical cancer gene mutations in clinical tumor samples. PLoS One. 2009;4(11):e7887.
30. Pao W, Kris MG, Iafrate AJ, Ladanyi M, Jänne PA, Wistuba II, Miake-Lye R, Herbst RS, Carbone DP, Johnson BE, Lynch TJ. Integration of molecular profiling into the lung cancer clinic. Clin Cancer Res. 2009;15(17):5317–22.
31. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin I, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN. Guidelines for validation of next-generation sequencing-based oncology panels. A joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn. 2017;19(3):341–65.
32. Oxnard GR, Jänne P. Power in numbers: Meta-analysis to identify inhibitor-sensitive tumor genotypes. Clin Cancer Res. 2013;19(7):1634–6. https://doi.org/10.1158/1078-0432.CCR-13-0169.
33. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, O'Dwyer PJ, Lee RJ, Grippo JF, Nolop K, Chapman PB. Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med. 2010;363(9):809–19.
34. Sundar R, Hong DS, Kopetz S, Yap TA. Targeting *BRAF*-mutant colorectal cancer: Progress in combination strategies. Cancer Discov. 2017;7(6):558–60.
35. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, Tsimberidou AM, Vnencak-Jones CL, Wolff DJ, Younes A, Nikiforova MN. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and the College of American Pathologists. J Mol Diagn. 2017;19(1):4–23.
36. Chene P. The role of tetramerization in p53 function. Oncogene. 2001;20(21):2611–7.
37. Mehrvarz Sarshekeh A, Advani S, Overman MJ, Manyam G, Kee BK, Fogelman DR, Dasari A, Raghav K, Vilar E, Manuel S, Shureiqi I, Wolff RA, Patel KP, Luthra R, Shaw K, Eng C, Maru DM, Routbort MJ, Meric-Bernstam F, Kopetz S. Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. PLoS One. 2017;12(3):e0173345.
38. Pilarski R, Rai K, Cebulla C, Abdel-Rahman. BAP1 tumor predisposition syndrome, 2016 Oct 13. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle: University of Washington, Seattle; 1993–2017. https://www.ncbi.nlm.nih.gov/books/NBK390611/.
39. Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Grenier TC, Grody WW, Hambuch T, Kalman L, Kant JA, Klein RD, Leonard DG, Lubin IM, Mao R, Nagan N, Pratt VM, Sobel ME, Voelkerding KV, Gibson JS. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. J Mol Diagn. 2012;14(6):525–40.

Part IV

Genomic Applications in Oncology

Genomic Applications in Hematologic Oncology

21

Kevin E. Fisher, Linsheng Zhang, and Charles E. Hill

Introduction to Single-Gene Assays

Single-gene assays used in the diagnosis and monitoring of hematologic malignancies can be broadly divided into DNA-based and RNA-based assays. Fundamentally, each of these types of assays uses similar analytical steps to extract molecular information from a patient sample. These steps include a standard biochemical nucleic acid extraction, template amplification by the polymerase chain reaction (PCR) or PCR equivalent, and detection of the sequence of interest. DNA-based assays commonly encountered in molecular hematologic oncology that highlight different molecular approaches to DNA analysis include B-lymphocyte and T-lymphocyte receptor gene rearrangement assays and exon-specific Janus kinase 2 (*JAK2*) and FMS-like tyrosine kinase 3 (*FLT3*) mutation analyses. Although multigene panels are becoming commonplace, single-gene tests for numerous targets may be performed (e.g., *CALR*, *NPM1*, *IDH1/2*, etc.).

DNA-Based Single-Gene Assays

B-Cell Immunoglobulin Gene Rearrangement

Both B and T lymphocytes (B cells and T cells) generate specific immune responses to diverse antigenic stimuli via a series of highly regulated somatic recombination events [1]. B cells utilize three genes that encode secreted immunoglobulins: *IGH@* located on chromosome 14q32 and either *IGκ* located on chromosome 2p12 or *IGλ* located on chromosome 22q11. The *IGH@* gene contains approximately 87 variable (V_H), 30 diversity (D_H), and 6 joining (J_H) segments that randomly recombine in a process called VDJ recombination [2]. The constant (C_H) regions are then joined with the VDJ complexes via RNA splicing, and the C_H segment determines the antibody class (e.g., IgG, IgA, and others) and allows for class switching. Once VDJ-C recombination occurs, B cells encounter antigen within the germinal center of lymph nodes. This antigen presentation triggers somatic hypermutation within the complementary-determining regions (CDR) of the V region and produces a virtually inexhaustible array of genetic diversity. This process allows for selection of a B-cell clone that codes for an immunoglobulin (Ig) with high specificity for the presented antigen. IgG molecules contain two identical heavy chain and two identical kappa or lambda light chains. *IGH* gene rearrangements occur first, followed by kappa, and then lambda rearrangements. The *IGκ* and *IGλ* light-chain loci lack D regions and therefore undergo VJ recombination only.

Molecular diagnostics typically utilizes *IGH@* and *IGκ* gene rearrangements to identify clonal B-cell populations. In the *IGH@* gene, CDRs are flanked by 15–30 amino acid framework regions (FRs) that rarely undergo mutation and thus are amenable to PCR primer targeting [3]. The three major FRs present within the V region (FR1, FR2, and FR3) serve as the 5′ forward primer-binding sites for amplification of the intervening DNA, and all reactions share the same 3′ reverse primer-binding site (Fig. 21.1). If a clonal proliferation has occurred, the gene rearrangement present in the clone will be overrepresented in the total population of rearrangements, allowing the identification of the cellular proliferation as clonal.

Assessing for B-cell clonality is useful in the diagnostic process for a variety of lymphomas and leukemias and in providing possible targets for monitoring disease. The three FR primers cover approximately 95% of all possible *IGH@* rearrangements, and assessing *IGκ* analysis is particularly useful for the diagnosis of marginal zone and follicular and Hodgkin lymphoma in paraffin-embedded tissue [4, 5]. Most laboratories use capillary electrophoresis and fluorescence to detect the amplicons; each of the forward primers is conjugated with a different color fluorescent dye, and the amplicons

K. E. Fisher
Pathology and Immunology, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA

L. Zhang · C. E. Hill (✉)
Pathology and Laboratory Medicine, Emory University Hospital, Atlanta, GA, USA
e-mail: cehill@emory.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_21

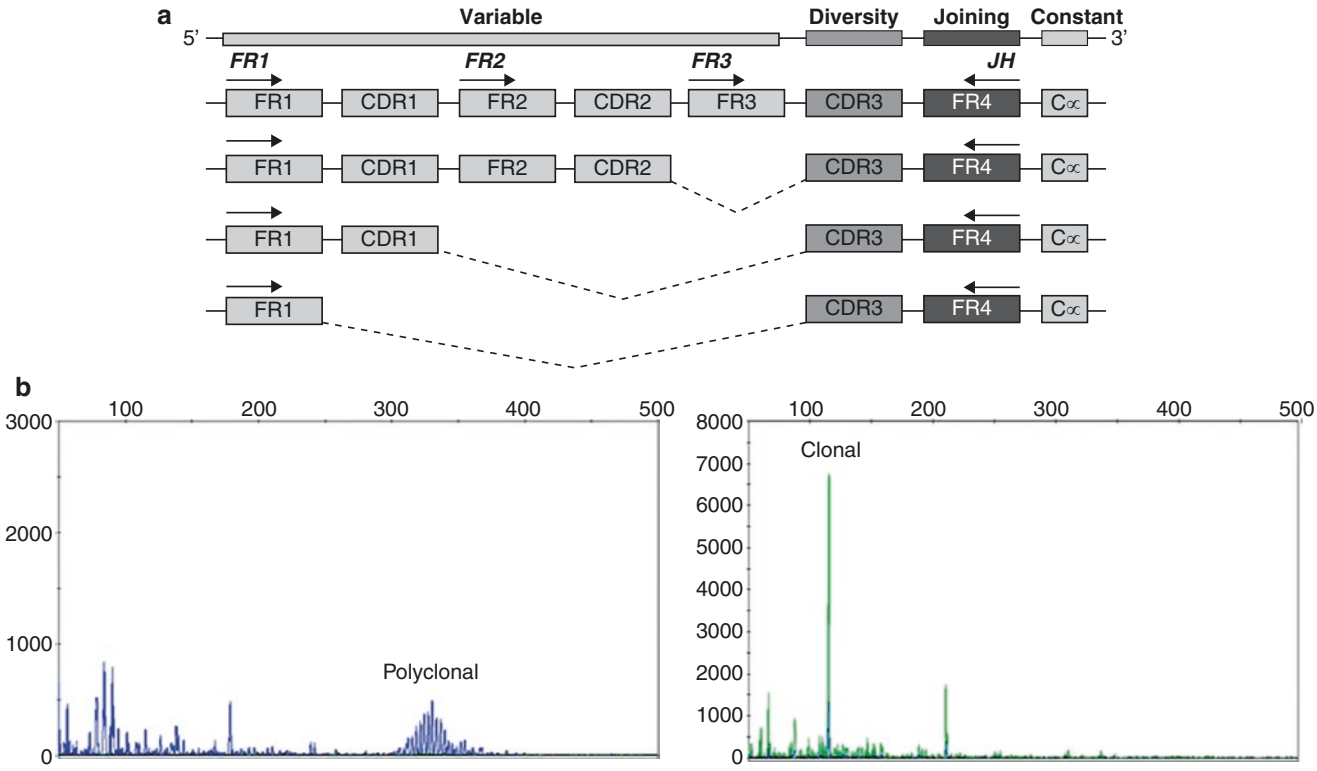


Fig. 21.1 *IGH* gene rearrangements. (**a**) Schematic representation of the *IGH* gene rearrangements. There are four relatively conserved framework regions (FR1, FR2, FR3, and FR4) and complementarity-determining regions (CDR1, CDR2, and CDR3). Detection of VDJ rearrangements is achieved using an FR1, FR2, or FR3 primer (forward) and a JH primer (reverse) located at the FR4 region. (**b**) Capillary electrophoresis tracing of PCR products. The FR1-JH combination of primers generates PCR products ranging from 290 to 360 base pairs (bp), the FR2-JH generates up to 280 bp, and FR3-JH generates the shortest PCR products (70–170 bp) and provides the most reliable amplification. The *left panel* shows a polyclonal background using primers targeting FR1. The *right panel* shows a clonal population of cells (peak) with a rearrangement detected by the FR3 primer set. (From Nikiforova et al. [3], with permission from Archives Pathology and Laboratory Medicine. Copyright 2007 College of American Pathologists)

of different sizes (FR1 290–360 bp, FR2 235–295 bp, FR3 69–129 bp) are separated by their charge/mass ratio (Fig. 21.1).

T-Cell Receptor Gene Rearrangement

As with immunoglobulins, multiple gene segments rearrange during T-cell development to encode the T-cell receptor (TCR). TCR-alpha (*TCRα*) and TCR-gamma (*TCRγ*) genes rearrange the V, J, and C loci, whereas TCR-beta (*TCRβ*) and TCR-delta (*TCRδ*) genes rearrange the V, D, J, and C loci. Each T cell possesses a gene that codes for a single TCR subunit gene that is unique in both sequence and length. Each subunit heterodimerizes to form the final TCR, either TCR-αβ or TCR-γδ. However, when assessing for T-cell clonality, *TCRγ* is used because it is rearranged at an early stage of T-cell development, and unlike *TCRδ*, it is not deleted in TCR-αβ cells. Thus, clonal rearrangements of *TCRγ* can be detected in both TCR-αβ and TCR-γδ clonal proliferations [6]. Clonal T-cell rearrangements are seen in greater than 90% of T-cell leukemias [e.g., pre-T-cell acute lymphoblastic leukemia (pre-T-ALL), T-cell prolymphocytic leukemia, and T-cell large granular lymphocytic leukemia] and 50–75% of T-cell lymphomas (e.g., peripheral T-cell lymphoma, mycosis fungoides, and anaplastic large-cell lymphoma) [7].

Similar to B-cell clonality testing, PCR primers target conserved regions in the V and J exons that flank variable regions within the *TCR*γ locus [8]. Capillary electrophoresis and fluorescence are used to detect the fluorescently tagged amplicons. There is not one shared 3′ reverse primer but rather two primers that target the J exon; thus amplicons of interest are present between 55–85 bp, 155–185 bp, 200–235 bp, and 235–270 bp. There are no consensus criteria for interpretation of these results, and each laboratory is responsible for test interpretation. Most laboratories correlate findings with morphologic and other laboratory data.

Diagnostic Challenges in Clonality Testing

A low quantity of B or T cells in a sample can lead to "false-positive" results or pseudoclonality. When there are very few

B or T cells in a sample, there is a limited repertoire to amplify. If one of these is slightly more abundant than others or if there is a minimal difference in amplification efficiency, this rearrangement may be preferentially amplified and therefore be of significantly higher peak amplitude compared to the background. This "peak" can be erroneously interpreted as a clonal proliferation.

It is possible for gene rearrangements to occur that yield products outside of the typical range of amplicon sizes. Usually, the product size will be just outside of the predicted ranges. It is important to remember that changes such as somatic hypermutation or other mutational processes can lead to this scenario. Therefore, it is possible to have a "false-negative" result due to the rearrangement producing a clonal product that is outside of the predicted size range. When amplification products fit these characteristics, it is important to interpret the results with caution. Sequence analysis can help distinguish a spurious peak from a clonal rearrangement.

Even when the presence of a monoclonal B- or T-cell proliferation is unequivocal, one must remember that the presence of clonality does not confirm malignancy. The results of molecular clonality assays ultimately must *always* be interpreted in the context of the clinical, morphologic, and immunophenotypic data of the clinicopathologic entity in question. Many benign dermatologic, inflammatory, and infectious disorders may demonstrate clonality in B- or T-cell rearrangement assays [9]. Furthermore, the presence of B-cell clonality or T-cell clonality does not necessarily imply a B-cell or a T-cell malignancy, respectively. For example, approximately 60% of pre-B acute lymphoblastic leukemias and 10% of acute myeloid leukemias (AMLs) can harbor T-cell gene rearrangements, and *IGH@* rearrangements can be seen in 25–30% of angioimmunoblastic T-cell lymphomas as a result of expanded EBV-positive B cells [7]. Therefore, the presence of a B-cell or T-cell rearrangement should be viewed independently as evidence of clonality, rather than malignancy or lineage specificity.

JAK2 Mutation Analysis

Whereas B-cell and T-cell gene rearrangement assays investigate discrete regions of multiple genetic loci, some single-gene assays target single-point mutations that are diagnostically or prognostically relevant. Janus kinase 2 (JAK2) is a tyrosine kinase that mediates signaling downstream of cytokine receptors, such as erythropoietin, thrombopoietin, and granulocyte colony-stimulating factor. JAK-mediated phosphorylation of signal transducers of activated transcription (STAT) proteins mediates target gene expression in the nucleus [10]. Gain-of-function mutations in the JAK/STAT pathway are principally involved in the development of myeloproliferative disorders.

One such mutation is a single somatic G > T nucleotide change in exon 14 of *JAK2* at position 1849. This mutation codes for a valine to phenylalanine conversion at codon 617 in the JAK2 pseudokinase domain and is seen in more than 95% of patients with polycythemia vera (PV) and in approximately 50% of patients with essential thrombocythemia or primary myelofibrosis, 20% of patients with refractory anemia with ring sideroblasts and thrombocytosis, and 5% of patients with AML or myelodysplastic syndrome (MDS). When *JAK2* mutational testing is extended to exons 12 and 13, virtually all PV patients will have *JAK2* mutations, thus fulfilling one of the major 2008 WHO diagnostic criteria for PV [7, 11].

A widely utilized testing method for *JAK2* p.V617F mutations employs real-time quantitative PCR (RQ-PCR) using a sequence-specific [mutant versus wild type (WT)] forward primer tagged with a specific probe for detection [12]. This mutant-specific or WT primer allows for single-base-pair discrimination because transcription elongation occurs at a very low rate with a 3′ mismatched base pair (Fig. 21.2). When detection of a product occurs, it is quantified by comparison to a WT and mutant standard curve, and calculation of the percent of mutant and WT allele is performed. The results are reported both qualitatively (detected/not detected) and quantitatively (as a percentage of mutated allele and percentage of WT allele). Qualitative reporting is useful for initial diagnosis, while quantitative reporting allows for assessment of allele burden and "zygosity," which yields additional prognostic information [13], and for monitoring molecular response to therapy [14].

In all RQ-PCR assays, the diagnostic limit of detection must be clearly defined. Eventually, a base pair/primer mismatch will proceed with amplification. Due to this "escape amplification," cycle thresholds are needed to mitigate false-positive amplifications. Cycle thresholds are determined in each laboratory and are analogous to the "limit of blank" in the chemistry laboratory that can be used to distinguish true positives from false positives. Any amplification detected after this threshold is considered to be a result of nonspecific primer annealing and amplification (Fig. 21.2).

A diagnostic dilemma occurs when the mutated allele is present at low detectable levels (e.g., 0.2% mutant allele) and falls below the defined accepted cutoff level for a positive result (e.g., 1.0% mutant allele). In these problematic cases, the assay could be rerun with the same sample. If the result still is equivocal, the result should be discussed with the clinician, and additional sample could be requested, if clinically indicated.

Because the spectrum of mutations associated with chronic myeloproliferative neoplasms has grown to include several genes, there is interest in sequential or multigene testing to include *JAK2*, *CALR*, and/or *MPL* [15].

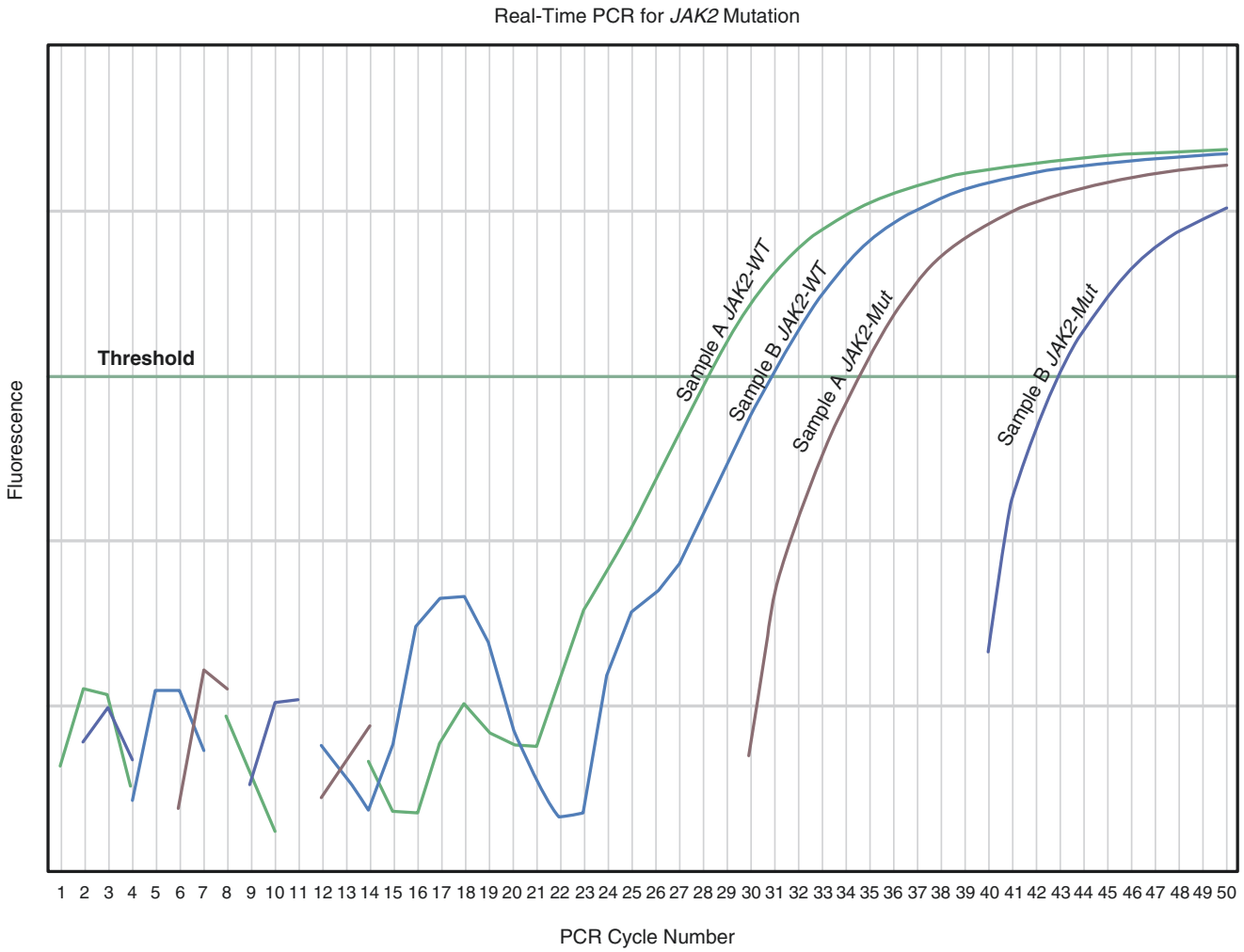


Fig. 21.2 Example of real-time PCR testing for the *JAK2* mutation from two patient samples. Sample A demonstrates a small population of neoplastic cells that harbor a mutated *JAK2* allele. The *green line* labeled "Sample A *JAK-WT*" shows the amplification of the abundant wild-type allele (designated by the amplification at an earlier cycle number), while the *brown line* labeled "Sample A *JAK-mut*" shows the amplification of the less abundant mutated allele indicating the presence of the *JAK2* p.V617F mutation. Sample B is an example of non-specific "escape" amplification. The *blue line* labeled "Sample B *JAK-WT*" is again showing amplification of the abundant wild-type allele. The *purple line* labeled "Sample B *JAK-mut*" crosses the *y*-axis threshold for positive fluorescence intensity at around cycle 43. This is beyond the designated cutoff for the assay which in this case is predetermined by the laboratory as cycle 41. This result would be reported as "negative for the *JAK2* p.V617F mutation"

FLT3 Mutation Analysis

FLT3 is a receptor tyrosine kinase that is normally expressed on hematopoietic stem cells and is lost as hematopoietic cells differentiate. *FLT3* mutations are the most common somatic alterations in AML and occur in approximately 25% of patients. Two main *FLT3* mutations result in constitutive activation of FLT3 signaling: internal tandem duplication (ITD) mutations in exons 14 and 15 and point mutations in exon 20 that alter the aspartic acid at codon 835 (known as D835). Studies have shown that *FLT3*-ITD mutations portend a poor prognosis, but the prognostic significance of D835 point mutants is less well understood [15].

The methodology to assess for *FLT3*-ITD mutations is fairly straightforward. PCR can be performed with fluorescently tagged forward and reverse primers targeting exons 14 and 15. In this assay, a result consists of a PCR product that is 330 base pairs (bp) in length. Due to the presence of inserted triplet repeats, ITD mutants yield PCR products that are longer than WT peaks – 333 bp or greater in size (Fig. 21.3).

Testing for the D835 mutation can be performed by another method that assesses the presence of point mutations: restriction enzyme digestion. Restriction enzymes are bacterially derived enzymes that cleave double-stranded DNA at specific palindromic sequences. In our assay example, a portion of the *FLT3*-TK2 domain is amplified with a forward PCR primer that is fluorescently tagged at the 5′ end and with a reverse unlabeled PCR primer. This amplification

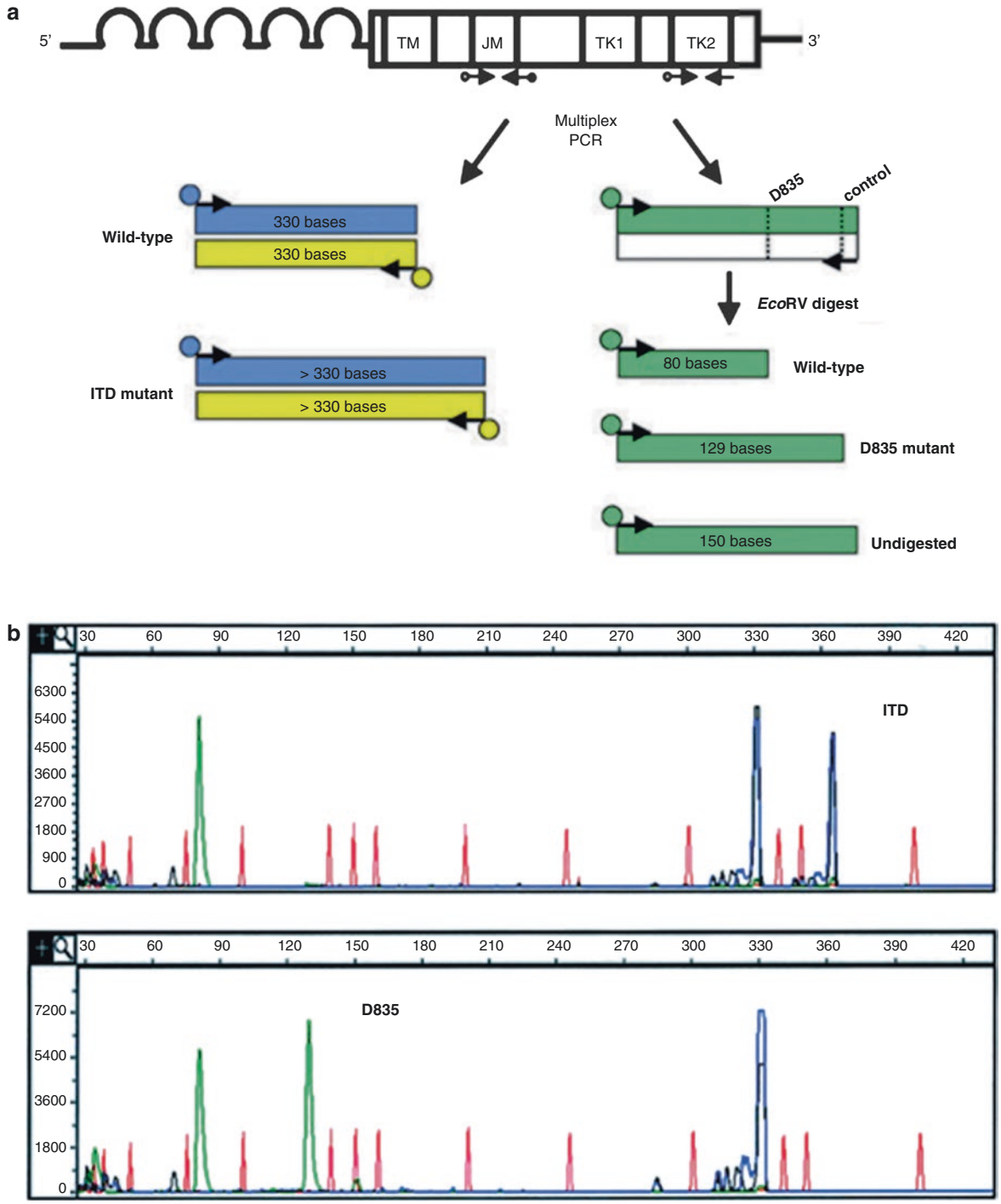


Fig. 21.3 A *FLT3* assay. (**a**) Diagram of the *FLT3* assay design. The *FLT3* gene consists of five extracellular immunoglobulin-like domains, a transmembrane domain (TM), a JM domain, and an interrupted kinase domain (TK1 and TK2). PCR primers flanking the JM domain (forward primer labeled with *blue*, reverse primer labeled with *yellow*) and primers specific for the TK2 domain (forward labeled with TET *green*, reverse unlabeled) are multiplexed into a single PCR reaction. After amplification, the PCR products are digested with the restriction enzyme *Eco*RV. The *dotted lines* in the TK2 PCR product represent the *Eco*RV cut sites, with the recognition sequence (GATATC). The JM portion of the PCR yields a wild-type PCR product of 330 bases labeled with both blue and yellow dyes. *FLT3* internal tandem duplication (ITD) mutations result in PCR products that are longer than wild type (>330 bp) and also labeled with both *blue* and *yellow*. After digestion, the D835 portion of the assay yields wild-type products sizing at 80 bases that are labeled in *green*. D835 mutant *green-labeled* products size at 129 bases, and undigested *green-labeled* products size at 150 bases. (**b**) Examples of results of the *FLT3* assay. Capillary electrophoresis pherograms: the *x* axis represents size of the PCR products in bases, and the *y* axis represents relative fluorescence intensity. *Red peaks* represent the internal size standard. *Green* PCR product peaks result from the D835 portion of the assay. *Blue* and *black peaks* result from the ITD portion of the assay. *Top pherogram*: Example of a *FLT3*-ITD mutant result. *Bottom pherogram*: Example of a D835 mutant result. (From Murphy et al. [16], Fig. 1, Fig. 2, with permission of Elsevier)

produces a PCR product of 150 bp. The 150 bp PCR product is then subjected to an *Eco*RV restriction digestion. The restriction enzyme recognizes and cleaves a specific palindromic sequence (5'-GATATC-3′/3'-CTATAG-5′) that is normally present in the WT *FLT3* exon that codes for aspartic acid at codon 835. When a mutation occurs, the specific palindromic sequence is altered, and *Eco*RV no longer recognizes it as a cleavage site. Thus, *Eco*RV digestion of the WT D835 PCR product results in an 80 bp fragment that can be detected by capillary electrophoresis. Since D835 mutations eliminate the *Eco*RV wild-type digestion site, the product is not cleaved, and there is an intact 130 bp fluorescently labeled fragment (Fig. 21.3). Twenty bp are lost because there is an *Eco*RV restriction site in the reverse primer. Because the reverse primer is not tagged with a fluorescent label, the 20 bp product remains undetected. This serves as a useful "digestion control" because the presence of abundant 150 bp amplicon would indicate that the restriction digest failed, or was incomplete [16].

RNA-Based Single-Gene Assays

A fundamental difference between DNA-based and RNA-based assays is the need to incorporate a reverse transcriptase step in RNA-based assays in order to convert RNA templates into DNA, amplifiable by conventional PCR methods. The prototypical RNA-based assay from a methodological standpoint is *BCR-ABL1* transcript testing. The presence of *BCR-ABL1* and its underlying chromosomal translocation (the Philadelphia chromosome [Ph]) can be detected in numerous ways, but the assay discussed here is for molecular monitoring of the mRNA fusion gene product.

Detection of *BCR-ABL1* mRNA Transcript

BCR-ABL1 fusion genes occur when there is a balanced translocation between a portion of the *BCR* gene on chromosome 22 with the *ABL1* gene on chromosome 9 [t(9;22)(q34;q11.2)]. Depending on which *BCR* and *ABL1* exons fuse with one another, mRNA transcripts of varying sizes are transcribed. Three major mRNA transcripts are described: m-*bcr* encodes a 190 kDa fusion protein (p190), M-*bcr* encodes a 210 kDa fusion protein (p210), and μ-*bcr* encodes a 230 kDa fusion protein (p230) (Fig. 21.4a). The p210 transcript/protein is predominately involved in chronic myelogenous leukemia (CML) [17], the p190 transcript/protein is most frequently associated with *BCR-ABL1*-positive ALL, and patients with the p230 transcript/protein often demonstrate CML with prominent neutrophilic maturation and/or conspicuous thrombocytosis [7]. All BCR-ABL1 fusion proteins are permissive for oncogenesis.

CML is a myeloproliferative neoplasm that is consistently associated with *BCR-ABL1* translocations, and the detection of the translocation is required for the diagnosis. Because normal cells do not harbor t(9;22)(q34;q11) translocations, it is implicit that they should not express fusion transcripts. In CML specifically, monitoring of transcript level during therapy provides important prognostic information. The rationale for accurate quantitative molecular testing in CML arose from the international randomized study of interferon and STI571 (IRIS) trial [18]. This trial demonstrated the superiority of imatinib (a tyrosine kinase inhibitor now used as first-line treatment for CML) over cytarabine and interferon-alpha and concluded that patients who demonstrated complete cytogenetic remission had a better prognosis than those who did not. It was determined later that patients with a reduction in *BCR-ABL1* transcript levels of at least three $\log_{10}$ by 12 months on imatinib therapy had a negligible risk for disease progression during the subsequent 12 months [19].

Quantitation of *BCR-ABL1* transcripts requires reverse-transcriptase PCR in combination with RQ-PCR to provide simultaneous detection and quantitation of *BCR-ABL1* fusion transcripts. Nucleic acid is extracted from leukocytes, and the DNA is degraded with DNAse. The remaining RNA is transcribed into cDNA using a reverse-transcriptase step. The cDNA is then subjected to RQ-PCR. Forward primers targeting *BCR* exons e13, e14, and e1 and a reverse primer targeting *ABL1* exon a2 are used. Primers targeting *ABL1* are used for a housekeeping gene endogenous control (Fig. 21.4a). The real-time detection of the amplicons occurs with fluorescent probes that target the *BCR-ABL1* transcript. With increasing amplification of the *BCR-ABL1* transcript, the amount of reporter dye excitation increases exponentially, and this is proportional to the amount of transcript present in the sample (Fig. 21.4b).

Initially, one of the major drawbacks to quantitative *BCR-ABL1* testing was the lack of a consensus reference standard. Therefore, quantitative results determined in one laboratory could not be reliably reproduced in a separate laboratory. In 2010, four fixed *BCR-ABL1*-control gene values were established as the first World Health Organization International Standards (WHO IS) for quantitation of *BCR-ABL1* [20], and these are now commercially available. The four reference standards are run in conjunction with patient samples, and a calculation of a correction parameter (CP) value between the patient values of a laboratory (in copies/mL) and the WHO IS is performed. This design allows that the final value is reported as an IS% ratio that can be compared to values obtained in other laboratories.

Whereas quantitative monitoring of *BCR-ABL1* has become the standard of care, there are other factors in the care of these patients that can best be served by more com-

prehensive molecular analyses. The advent of new pharmaceutical treatments has led to a need to test for *ABL1* kinase domain mutations to assess drug resistance. In addition, there are other chromosomal abnormalities that provide prognostic or therapeutic information. These issues are best addressed by more comprehensive molecular analyses.

Chromosome Assays

Chromosome assays have been utilized in hematologic oncology for almost four decades since the inception of Giemsa staining of replicating chromosomes (commonly known as G-banding). The visualization of chromosomes has allowed scientists to detect and decipher aberrations in

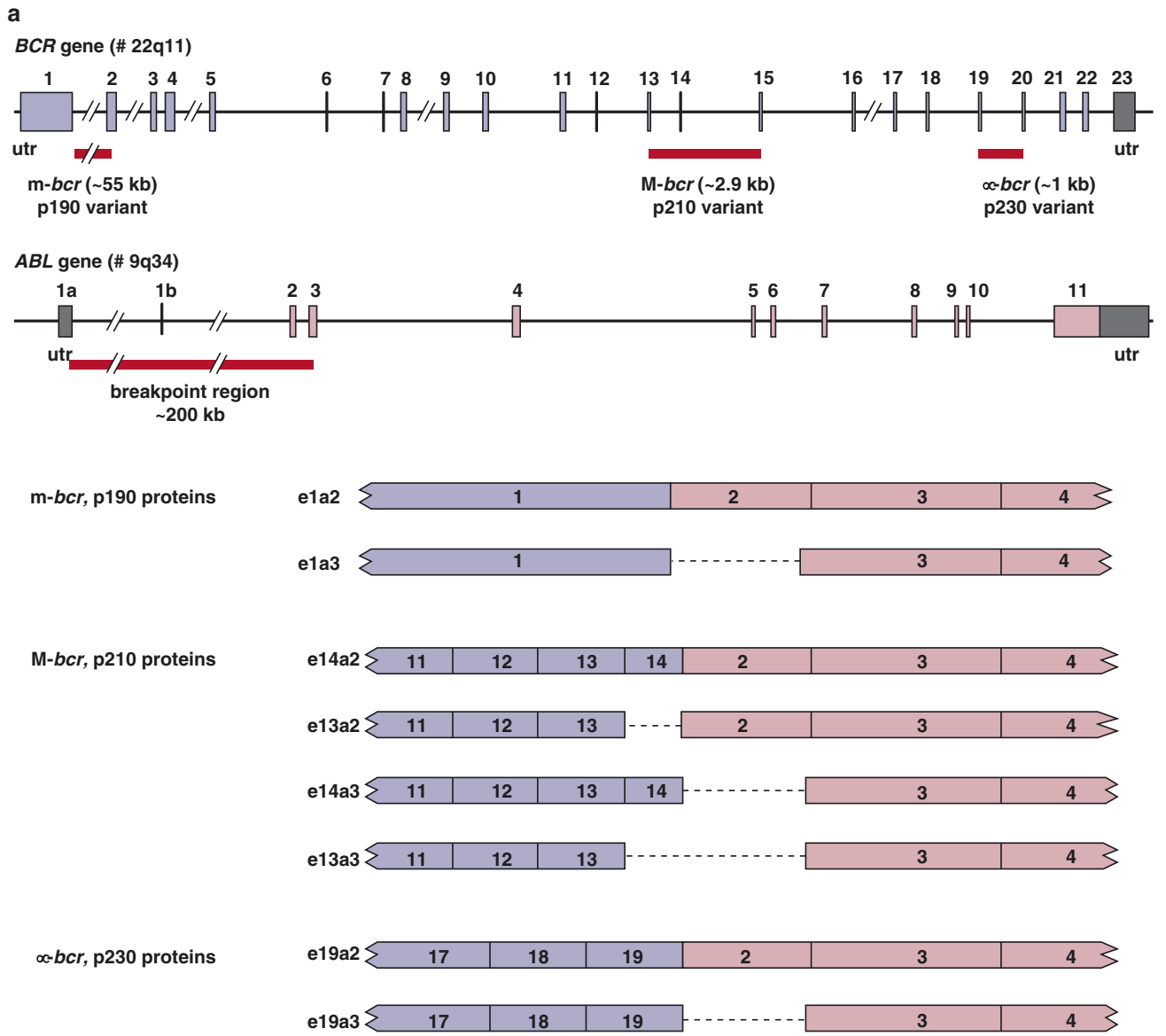


Fig. 21.4 Structure of the *BCR* and *ABL1* genes with breakpoint regions and corresponding fusion gene transcripts. (**a**) The *ABL1* gene contains one large breakpoint region (~200 kb), whereas three breakpoint regions have been found in the *BCR* gene: m-*bcr*, M-*bcr*, and μ-*bcr*, which are associated with the p190, p210, and p230 *BCR-ABL1* fusion proteins, respectively. The three well-defined breakpoint regions in the *BCR* gene can produce at least eight different fusion transcripts, because of alternative splicing in the *ABL* gene (splicing to exon 2 or exon 3) and because the M-*bcr* consists of two intronic regions (intron 13 and intron 14). (**b**) Sample A shows amplification of the intact *ABL1* gene product (*purple line* labeled "Sample A *ABL1*") with no detection of the fusion transcript. Sample B demonstrates amplification of the *BCR-ABL1* fusion transcript. The *green line* labeled "Sample B *ABL1*" shows the amplification of the intact *ABL1* gene product, while the *second green line* labeled "*BCR-ABL1*" shows the amplification of the *BCR-ABL1* fusion transcript. (From Dekking et al. [17], Fig. 3a, 3b, with permission of Elsevier)

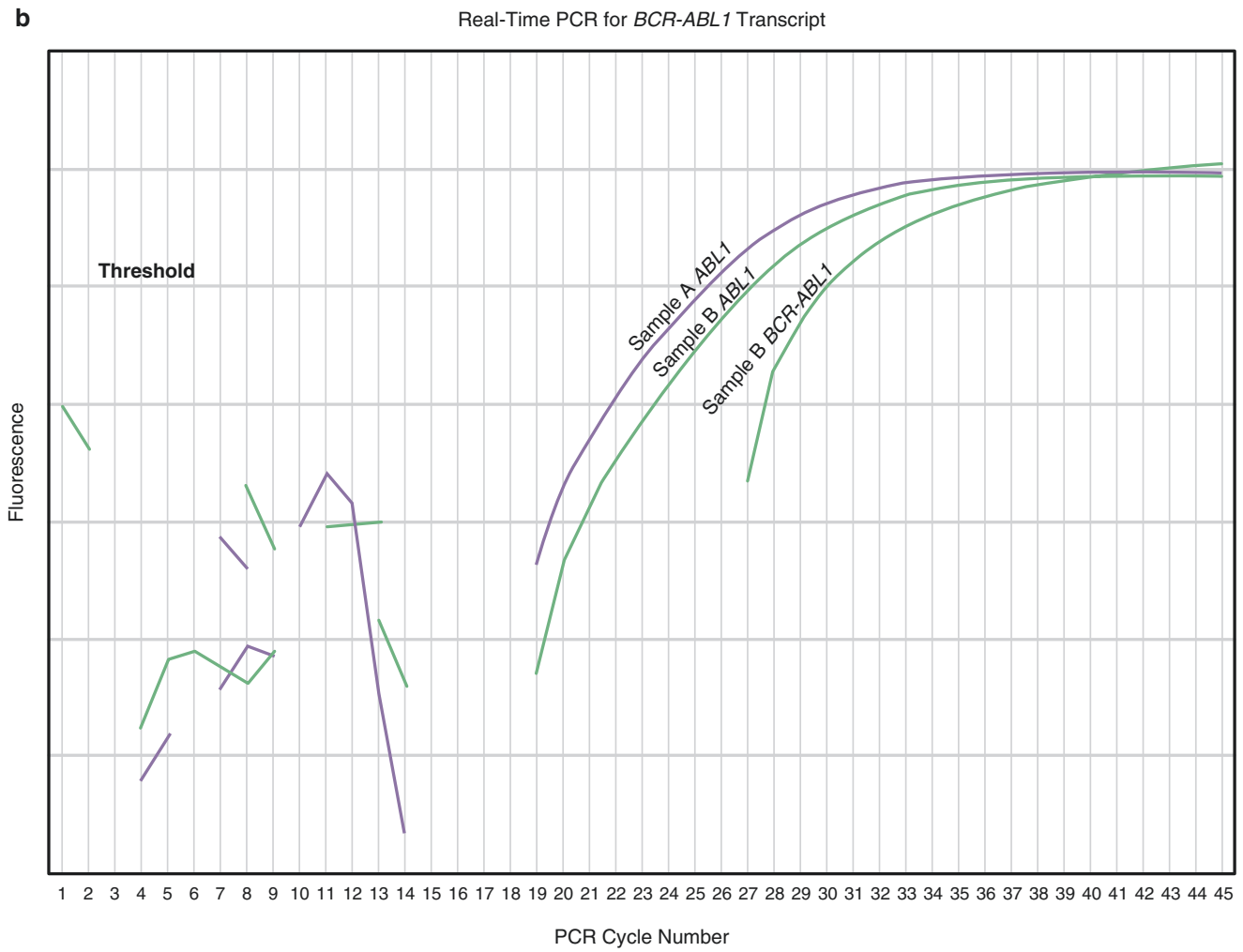


Fig. 21.4 (continued)

chromosome structure and define disease entities. Many hematologic cancers require the demonstration of a particular chromosomal abnormality for diagnosis, and the presence of certain chromosomal abnormalities provides prognostic and therapeutic information. Standard methodologies [conventional cytogenetics and fluorescence in situ hybridization (FISH)] and two newer methodologies, single-nucleotide polymorphism (SNP) arrays and array comparative genomic hybridization (aCGH), will be discussed.

Conventional Cytogenetics

Conventional cytogenetic techniques are the most commonly used modalities in clinical laboratories to assess chromosomal composition for the diagnosis and therapeutic evaluation of hematologic malignancies [21]. Routine cytogenetic analysis (karyotyping) can achieve resolutions sufficient to detect alterations of a few megabases (Mb) [22]. This range of detection is useful to assess both gains and losses of large regions of the genome, as well as rearrangements within and among chromosomes. Cytogenetic analysis can identify numerous diagnostic chromosomal abnormalities in leukemia and lymphoma such as the Ph chromosome in CML, 8q24 *MYC* translocations in Burkitt lymphoma, recurrent genetic abnormalities in acute myeloid leukemias, and deletions of 5q in MDS (Fig. 21.5).

To perform routine karyotypic analysis, live cells are cultured and stimulated to divide, in order to promote mitoses. The cells are arrested in metaphase with a pharmacologic microtubule inhibitor (e.g. colchicine) and treated with trypsin followed by a counterstain (Giemsa or equivalent). This produces differential staining of chromosomal regions leading to light (replicating) and dark (condensed chromatin) bands, each corresponding to a specific area of chromosomal DNA. A cytotechnologist then interprets the chromosomes

and their respective bands for any abnormalities, deletions, or duplications. There are, however, some major limitations to routine cytogenetic analysis. In order to perform the assay, cells must be received in a non-fixed, fresh state in order to stimulate mitoses and to perform successful staining. Also, karyotyping is not suited for the detection of microdeletions, cryptic translocations, or small genetic alterations due to the limit of detection which is at best a few Mb. More sensitive methodologies such as FISH are available to detect abnormalities not seen on routine cytogenetic analysis.

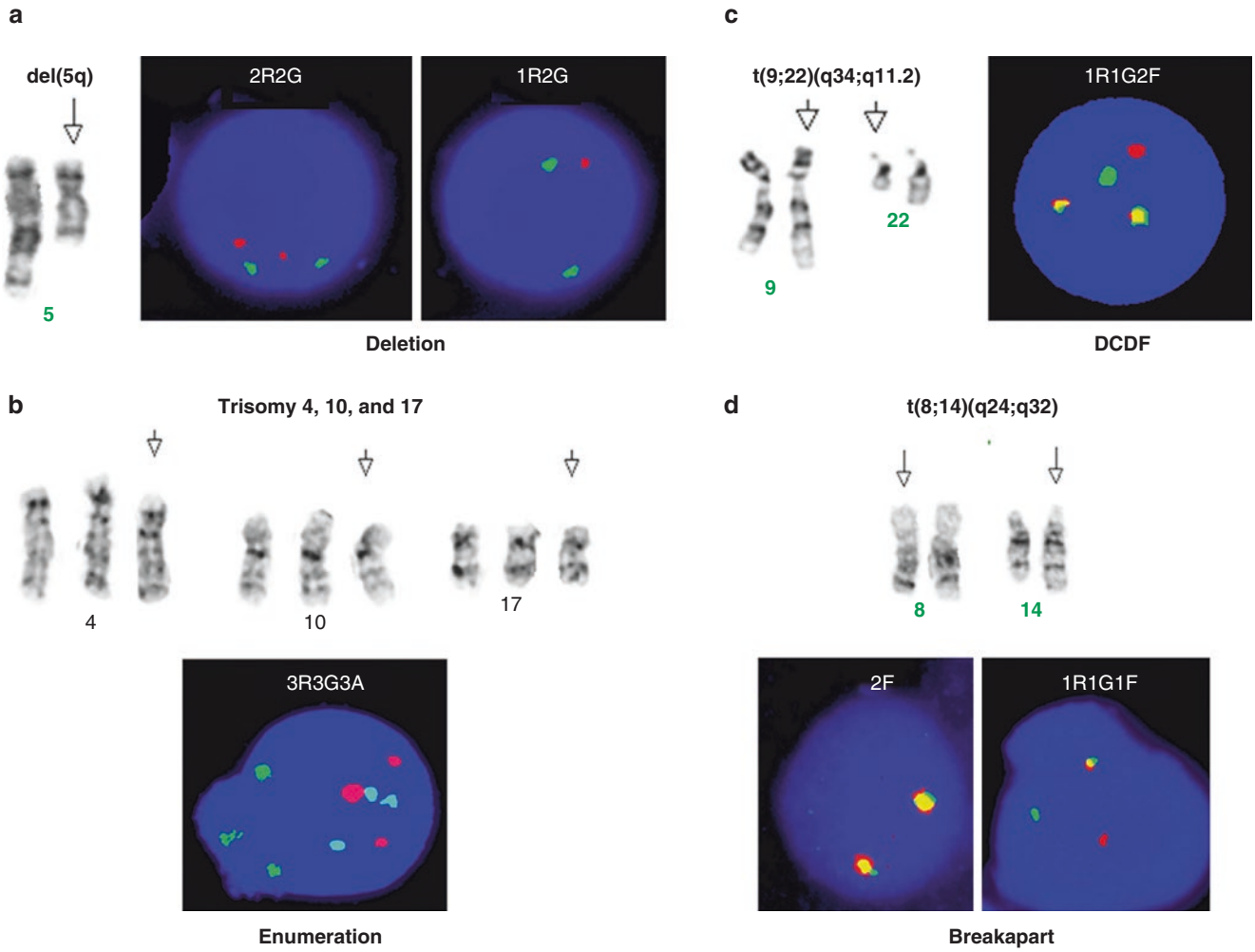


Fig. 21.5 Examples of classic cytogenetic abnormalities in hematologic malignancies visualized by conventional karyotyping or fluorescence in situ hybridization. (**a**) Deletion of the short arm of chromosome 5q [del(5q)]. (*Left*) A conventional karyotype shows the absence of chromosomal material from the long arm of chromosome 5. (*Right*) The expected 2R2G (two *red*, two *green*) signal indicating two copies of the long arm of chromosome 5q (R = red) and two centromeres for chromosome 5 (G = green) is shown in the nucleus (*blue circle*) on the *left*. Chromosomal material from the long arm of chromosome 5 is lost from the nucleus on the *right*. This results in an aberrant 1R2G signal. (**b**) Detection of the Philadelphia chromosome [t(9;22)(q34;q11.2)]. (*Left*) A conventional karyotype showing the translocation of genetic material from chromosome 22 (*BCR*) to chromosome 9 (*ABL1*). (*Right*) An aberrant one *red*, one *green*, two fusion (1R1G2F) signal is shown. The appearance of the *yellow* fusion color indicates juxtaposition of chromosomal material from chromosomes 9 and 22. The non-rearranged chromosome is shown with the expected 1R1G signal. (**c**) Trisomy 4, 10, and 17 in a patient with acute lymphoblastic leukemia (ALL). (*Left*) The karyotype revealed an extra chromosome 4, 10, and 17. This combination of cytogenetic abnormalities is prognostically favorable in some ALLs. (*Right*) Enumeration FISH probes with different fluorophores highlight the three copies of each chromosome [R = red, chromosome 7; G = green, chromosome 10; A = aqua, chromosome 17]. (**d**) Cytogenetic and FISH results from a patient with Burkitt lymphoma showing the translocation between *MYC* on chromosome 8 and *IGH* on chromosome 14. (*Left*) The characteristic translocation seen in Burkitt lymphoma involving the *MYC* locus on the long arm of chromosome 8 and the *IGH* locus on the long arm of chromosome 14 is shown [t(8;14)(q24;q32)]. (*Right*) An example of a FISH breakapart probe. The expected 2F (two fusion) signal indicates an intact *MYC* locus (chromosome 8, shown on the *left*). In normal cases, the contiguous configuration of the *MYC* locus juxtaposes the *red* and *green* probes to produce a fusion signal. The loss of the fusion signal (*right*) results in an aberrant one *red*, one *green*, one fusion (1R1G1F) signal. The loss of a fusion signal denotes a "break" in the genetic material from one set of chromosome 8. The appearance of both *red* and *green* signals indicates the genetic material from chromosome 8 is still detectable, but it is no longer in its contiguous configuration. (Image courtesy of Dr. Debra F. Saxe, PhD, Emory University School of Medicine Department of Pathology and Laboratory Medicine)

Fluorescence In Situ Hybridization (FISH)

The introduction of FISH in routine clinical diagnostics has enabled molecular cytogenetics and is now a widely used tool for the diagnosis and monitoring of patients with hematologic malignancies. FISH has aided the identification of structural chromosome rearrangements and is frequently used as a supplemental test to karyotypic analysis [20]. There are a few distinct advantages to FISH compared to conventional chromosome analysis. FISH can be performed on formalin-fixed paraffin-embedded (FFPE) tissue and, therefore, can visualize both metaphase and interphase chromosomes. FISH can visualize DNA segments between 100 kilobases (kb) and 1.5 Mb versus 2–5 Mb. Also, FISH has a relatively short turnaround time (as little as 16 h in some cases) compared to days for routine cytogenetic analysis. However, in order for FISH analysis to have any utility, one must first select FISH probes that interrogate the chromosome or region of interest.

FISH probes are fluorescently labeled DNA probes that bind to complementary regions of the chromosome of interest. The DNA content of the target probes is made of either locus-specific indicators (LSI) or centromere enumeration probes (CEP). LSI probes are specific to unique DNA sequences (target-specific loci) and are composed of 200–600 bp segments that span the 100 kb–1.5 Mb locus of interest. CEP probes are composed of highly repetitive DNA sequences that target common regions of all centromeres or are directed to specific centromeres of particular chromosomes. Each probe is tagged directly or indirectly with a fluorophore. The probes and the target DNA are denatured to yield single-stranded DNA, which allows for the annealing of complementary DNA sequences. The signal probes bind the target of interest, and the attached fluorophores are evaluated by fluorescence microscopy. For hematologic malignancies, several types of FISH probes are used, and each of these probes has utility in the appropriate diagnostic context (Table 21.1).

Enumeration (centromere) probes are utilized to examine whether neoplastic cells show a loss or gain of chromosomal number and serve as internal controls for deletion or amplification probes. The loss or gain of chromosome copy number or deletions of certain regions of the chromosome provide diagnostic and prognostic information. For example, characteristic hypo-lobated or non-lobated megakaryocytes are visualized in bone marrow preparations of patients with MDS associated with isolated del (5q) (Fig. 21.5a), whereas triple trisomy 4, 10, and 17 is considered a prognostically favorable genetic event in some ALLs (Fig. 21.5b) [23].

Some FISH probes are designed such that different fluorescent signals juxtapose to emit a single "fusion" color. The presence or absence of the fusion color can be detected in dual-color dual-fusion (DCDF) or breakapart probes, respec-

Table 21.1 Commonly used FISH probes for hematologic malignancies

Probe type	Locus	Identifies	Abnormal signal	Disease
Enumeration (centromere)	Telomeres or centromeres of chromosomes 3, 4, 7,10, 12, and 17	Gain or loss of chromosome number	nR, nG, nB where n ≠ 2	ALL, MM, CLL, and AMLs with complex karyotypes
Deletion	5p15.2 (G) 5q31 *EGR1* (R)	Loss 5, del(5q)	RG, RGG	MDS, AML
	7p11.1-q11.1 (G) 7q31 (R)	Loss 7, del(7q)		
	13q34 *LAMP1*(G) 3q14 (R)	Loss 13, del(13q)	RG, RGG	MM, CLL
	17p11.1-q11.1 (G) 17p13.1 *TP53* (R)	Loss 17, del(17p)		
Dual-color Dual-fusion	9q34 *ABL1* (R) 22q11.2 *BCR* (G)	t(9;22) *BCR/ABL1*	Any F	CML, Ph + ALL
	15q22 *PML* (R) 17q21.1 *RARA* (G)	t(15;17) *PML/RARA*	Any F	APL
	11q13 *CCND1* (R) 14q32 *IGH* (G)	t(11;14) *IGH/CCND1*	Any F	MCL, MM
Breakapart	11q23 *MLL* (F)	*MLL* rearrangements	Any R or G	AML, MDS, mixed-lineage lymphomas
	16q22 *CBFB* (F)	inv [16], t(16;16), del(16q)	Any R or G	AML with inv [16] or t(16;16)
	8q24 *CMYC* (F)	*MYC* rearrangements	Any R or G	Lymphoma (Burkitt)

R red, *G* green, *B* blue, *F* fusion (yellow), *MM* multiple myeloma, *CLL* chronic lymphocytic leukemia/small lymphocytic lymphoma, *AML* acute myeloid leukemia, *MDS* myelodysplastic syndrome, *CML* chronic myelogenous leukemia, *Ph+* Philadelphia chromosome, *ALL* acute lymphoblastic leukemia, *APL* acute promyelocytic leukemia, *MCL* mantle cell lymphoma, *MM* multiple myeloma

tively. DCDF probes can distinguish balanced translocations such as t(9;22)(q34;q11) involving *BCR* and *ABL1*. The red (R) fluorophore attached to the probe that targets the *ABL1* gene juxtaposes with the green (G) fluorophore attached to the probe that targets the *BCR* gene. This juxtaposition causes the fluorescent signal to fluoresce yellow. The presence of the yellow fusion (F) color is scored as a positive rearrangement (Fig. 21.5b). In contrast, breakapart probes start as a fusion signal, and when the locus is disrupted to give a single G or single R signal, it implies that a portion of the locus has rearranged to another chromosome. Breakapart probes are useful for detecting gene rearrangements in loci that rearrange with multiple chromosomes such as the 11q23 *MLL* locus in ALL and AML or the *MYC* locus in Burkitt lymphoma (Fig. 21.5d) [24].

Though FISH is a highly effective diagnostic tool, there are some limitations. Technically, the scoring of FISH probes and determining cutoffs for positivity can be problematic. A cytogenetic technologist, geneticist, or pathologist counts anywhere from 100–200 cells, and cutoffs must be well-defined to avoid making inappropriate calls. This is particularly true for deletion probes. In addition, like any nucleic acid-based test, sample integrity is very important and can lead to poor hybridization. Fusion signals may be small, weak, or absent due to biologic and technical reasons. Also, quality control is critical to determine that the FISH process is working appropriately.

By its nature, FISH analysis can only interrogate preselected chromosomal regions. By focusing on a specific region of the genome, other significant prognostic or biologic information in other regions of the same or other chromosomes may remain undetected. Theoretically, FISH probes targeting all regions of all chromosomes could be employed, but this is not a practical alternative for most clinical laboratories. For this reason, array comparative genomic hybridization (aCGH) or single-nucleotide polymorphism (SNP) arrays that combine comprehensive coverage of the entire genome with the ability to assess specific regions of interest are attractive new methods for whole-chromosome analysis.

Array Comparative Genomic Hybridization (aCGH)

Array CGH, also called molecular karyotyping, is a technique that uses competitive hybridization of fragmented tumor and control DNA to comprehensively interrogate hundreds of discrete genomic loci for DNA copy number gains and losses [25, 26]. The resolution aCGH depends on the spacing and length of the interrogating DNA probes on the microarray. Whole-genome aCGH platforms exist that utilize equally spaced probes to interrogate the genome at 6 kb–70 kb intervals, although most of the clinically available aCGH platforms use targeted arrays that detect previously characterized aneuploidies or chromosomal abnormalities/rearrangements [27].

In aCGH, equal amounts of fluorescently labeled sample DNA (e.g., using Cy3, a green dye) and control DNA (e.g., using Cy5, a red dye) are co-hybridized to an array containing complementary DNA targets. When the sample or control DNA anneals, the labeling intensities or "spots" of fluorescence are measured. The resulting ratio of the fluorescence intensities is proportional to the ratio of the copy numbers of DNA sequences in the sample and control genomes. If the intensities of the fluorescent red and green dyes are equal, that region of the patient's sample genome is interpreted as having equal quantity of DNA to the control sample. If there is an altered green:red ratio, this indicates a loss or a gain of the sample DNA at that specific genomic region (Fig. 21.6) [26].

One of the most commonly reported applications of aCGH in hematologic oncology is the detection of chromosomal abnormalities in "cytogenetically normal" malignancies [28], that is, detecting copy number alterations (CNAs) or chromosomal abnormalities that are not detected by conventional cytogenetics. Because aCGH can reach genomic resolution of approximately 6 kb, aCGH can detect chromosomal alterations that G-banding or FISH probes lack the resolution to detect. In one study, aCGH detected new cytogenetic abnormalities not seen by karyotype or FISH analyses in 80% of MDS patients [29]. Other studies have also demonstrated genomic imbalances, cryptic CNAs, or karyotypic alterations in cytogenetically normal MDS patients [30–32]. Similar results have been described in pediatric ALL [33], blastic plasmacytoid dendritic cell neoplasms [34], chronic lymphocytic leukemia [35], and AML [36]. These studies suggest that aCGH is a useful adjunct to conventional chromosomal analyses to assess for both diagnostic and prognostic cytogenetic abnormalities in hematologic malignancies.

Despite its broad and sensitive cadre of applications, aCGH is limited as a comprehensive diagnostic tool because of an inability to detect balanced translocations, copy-neutral loss of heterozygosity (CN-LOH, a process whereby a lost portion of the chromosome is reduplicated from the sister chromatid), and uniparental disomy (UPD, a process similar to CN-LOH that involves an entire chromosome). The results of aCGH are also influenced by the amount of tumor sample present, the presence of tumor subclones, and the resolution of the microarray platform [25]. Not surprisingly, aCGH platforms that target the entire genome are more expensive and are likely to detect genomic imbalances of unclear significance [26].

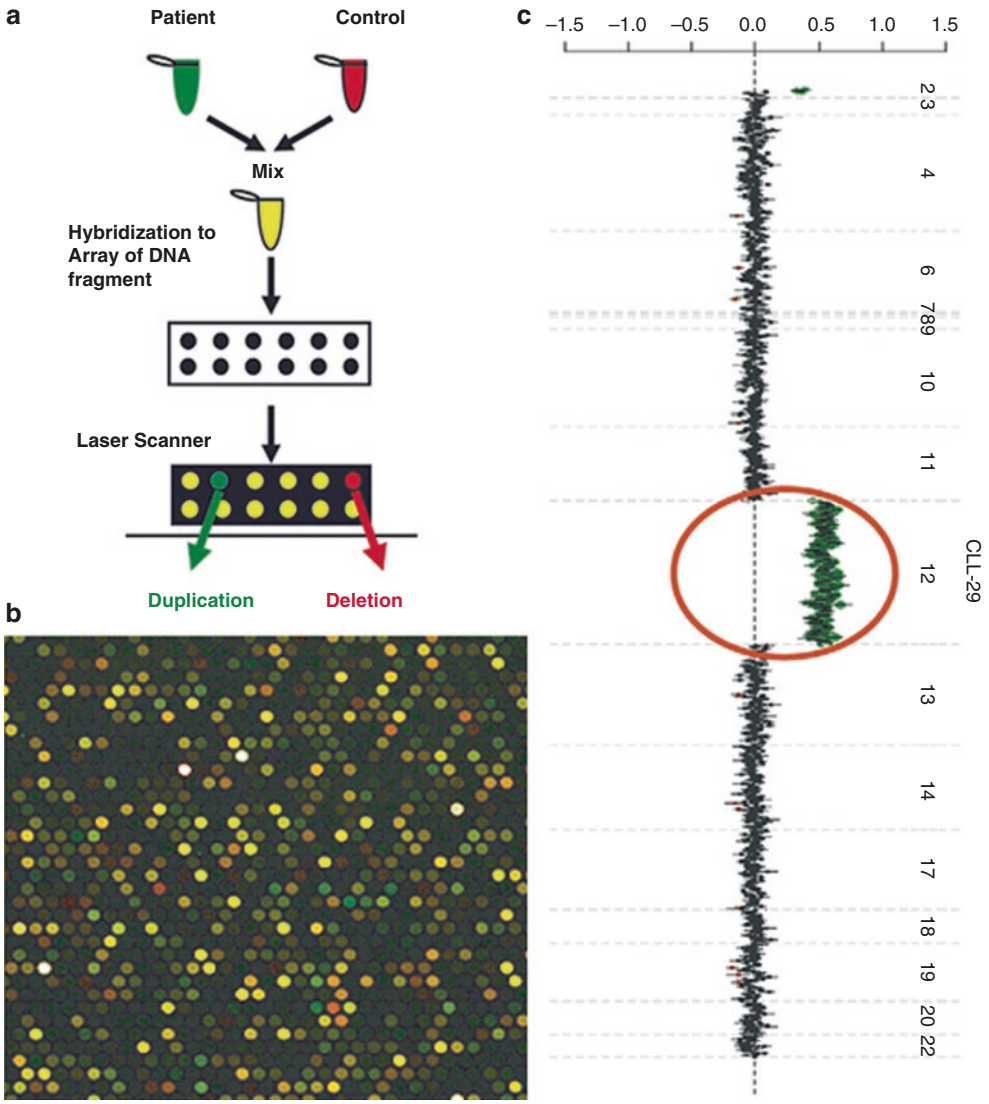


Fig. 21.6 An example of array comparative genetic hybridization with gain of chromosome 12 in chronic lymphocytic leukemia. (**a**) Sample DNA is labeled with a green-fluorescent dye (Cy3), and reference DNA is labeled with red (Cy5). The two are mixed and competitively co-hybridized to an array containing genomic DNA targets that have been fixed to a glass slide. The areas on the slide that appear *green* indicate extra chromosomal material (duplication) in the test sample at that particular region. Areas on the slide that appear *red* indicate relatively less test DNA (deletion) in the sample at that specific spot. *Yellow areas* indicate equal amounts of sample and reference DNA. (**b**) The slides are scanned into image file, and an output of scanning depicts hundreds of spots with different ratios of the fluorescence intensities. (**c**) Microarray image files are quantified using software that detects the fluorescent signals and maps them to specific regions of the chromosome. The signals are converted to the data output format shown here. A gain of genetic material from chromosome 12 from a CLL patient sample is indicated in the *red circle.* (Adapted from Shinawa and Cheung [26], with permission of Elsevier)

Single-Nucleotide Polymorphism (SNP) Arrays

SNPs are single-base-pair changes in genomic DNA that occur (on average) every 1000–2000 bases [37]. Because SNPs have a low rate of recurrent mutation, studies have "mapped out" the location of common SNPs along the human genome. These SNP maps serve as reference sequences to allow comparison between DNA of interest (sample) and normal DNA (reference DNA or uninvolved tissue) at the single-nucleotide level. The utility of SNP analysis is finding favor in virtually every facet of medicine: pharmacogenetics, neuropsychiatric disorders, and forensics, to name a few [38–40].

For hematologic malignancies, SNP array karyotyping takes advantage of very large numbers of allele-specific probes synthesized on microarrays to detect genome-wide copy number alterations and allelic imbalances. SNP array karyotyping represents the only platform currently available

for genome-scale detection of CN-LOH or UPD. However much like aCGH, SNP arrays are not designed to detect balanced translocations, which, as noted previously, are commonly found in hematopoietic malignancies [41]. That said SNP array karyotyping is a tool for the diagnosis and monitoring of hematopoietic neoplasms.

SNP arrays interrogate genomic loci to determine the DNA copy number and the genotype. The most common SNP array platforms include Illumina and Affymetrix arrays, which utilize bead [42, 43] or chip technology [44], respectively. In the bead-based Illumina platform, whole-genome amplification and fragmentation steps are followed by hybridization to an oligonucleotide bead array (Fig. 21.7). In the Affymetrix technology, genomic DNA is digested by restriction endonucleases, amplified and labeled and hybridized to oligonucleotides on a microarray chip (Fig. 21.8) [45]. SNP arrays offer superior resolution to conventional karyotypic analysis, can detect genetic lesions less than 100,000 bp in size, and nullify the need for mitotically active cells. They can also detect genes involved in unbalanced copy number changes and determine genetic targets of amplifications and deletions [41].

During the past decade, this resolution sensitivity has yielded a wealth of information regarding genomic alterations in hematologic malignancies. For example, SNP arrays identified recurrent abnormalities in the *EBF1* and *PAX5* genes in childhood ALL [46, 47]. CLL and plasma cell myeloma are amenable to SNP array analysis, because there sometimes is difficulty obtaining metaphase chromosomes for conventional karyotyping [41]. SNP arrays identified 24 large (>10 Mb) copy-neutral regions with LOH in some cases of CLL that were not detectable by alternative methods

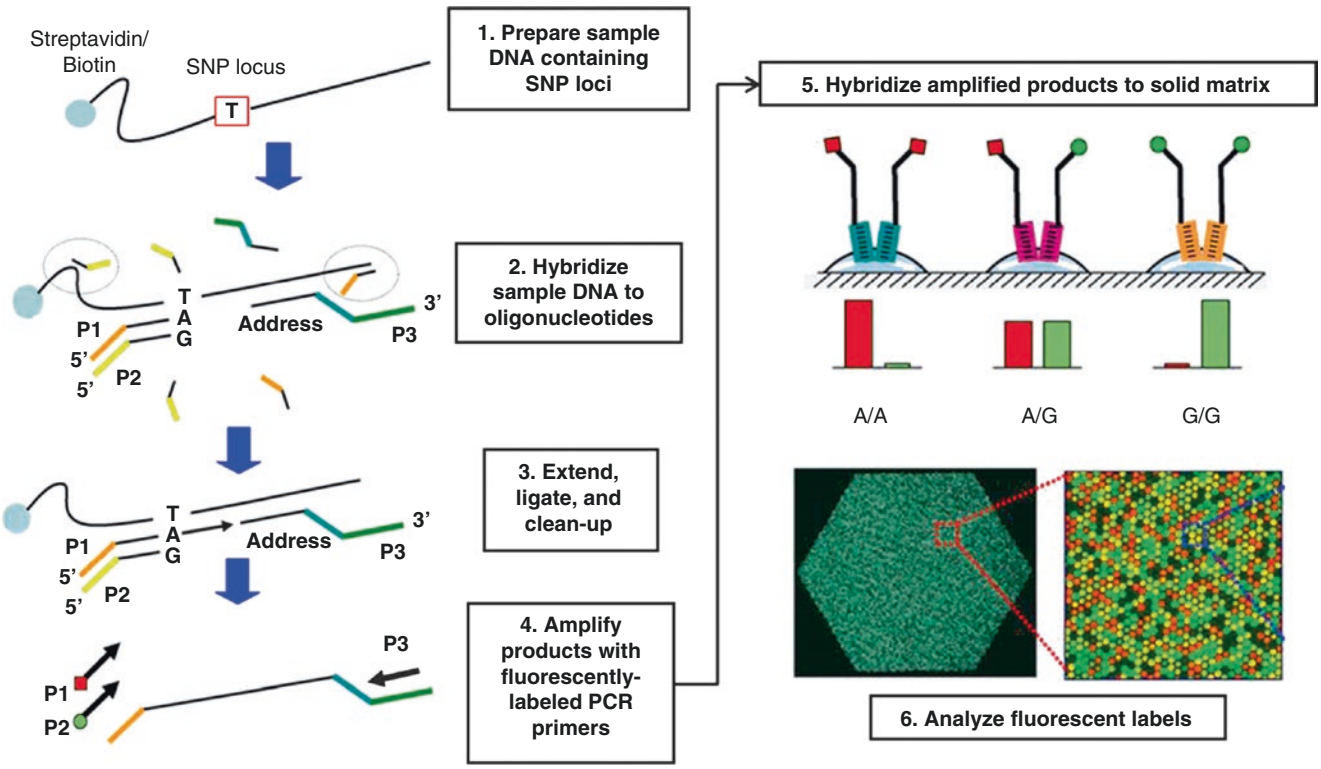


Fig. 21.7 Principles of a bead-based (Illumina) SNP array [1]. Genomic sample DNA is first "activated" by biotinylation [2]. Oligonucleotides that correspond to specific SNPs are then combined with the activated DNA in the oligonucleotide/target annealing step, in which the query oligonucleotides hybridize to the genomic DNA that binds to the paramagnetic particles (via the biotin). Two allele-specific oligonucleotides (ASOs designated *orange* or *yellow*) and one locus-specific oligonucleotide (LSO designated *green*) are designed for each SNP. The ASO contains a 5′ universal sequence that serves as a universal primer for all beads. The LSO contains a unique sequence complementary to a particular bead type designated as the "address." [3] DNA polymerase with high specificity for a perfectly matched target sequence at the SNP adds nucleotides between the ASO and the LSO. DNA ligase is used to seal the nick between the extended ASO and the LSO to form PCR templates that can be amplified with universal PCR primers (Step 4) [4]. PCR amplification is performed with three universal PCR primers (P1, P2, and P3) labeled with Cy3 (*green*), Cy5 (*red*), and biotin, respectively [5]. Double-stranded dye-labeled PCR products are converted to single-stranded DNA (ssDNA) on paramagnetic beads (again via binding with biotin). These ssDNA are removed and hybridized to their complement bead type via their unique "address." [6] The bound DNA is containing Cy3 and/or Cy5. The dyes are excited by lasers at different wavelengths. Based on the intensities detected from the two channels for the two respective alleles of each SNP, genotypes are designated using computer software. (From Shen et al. [87], with permission of Elsevier)

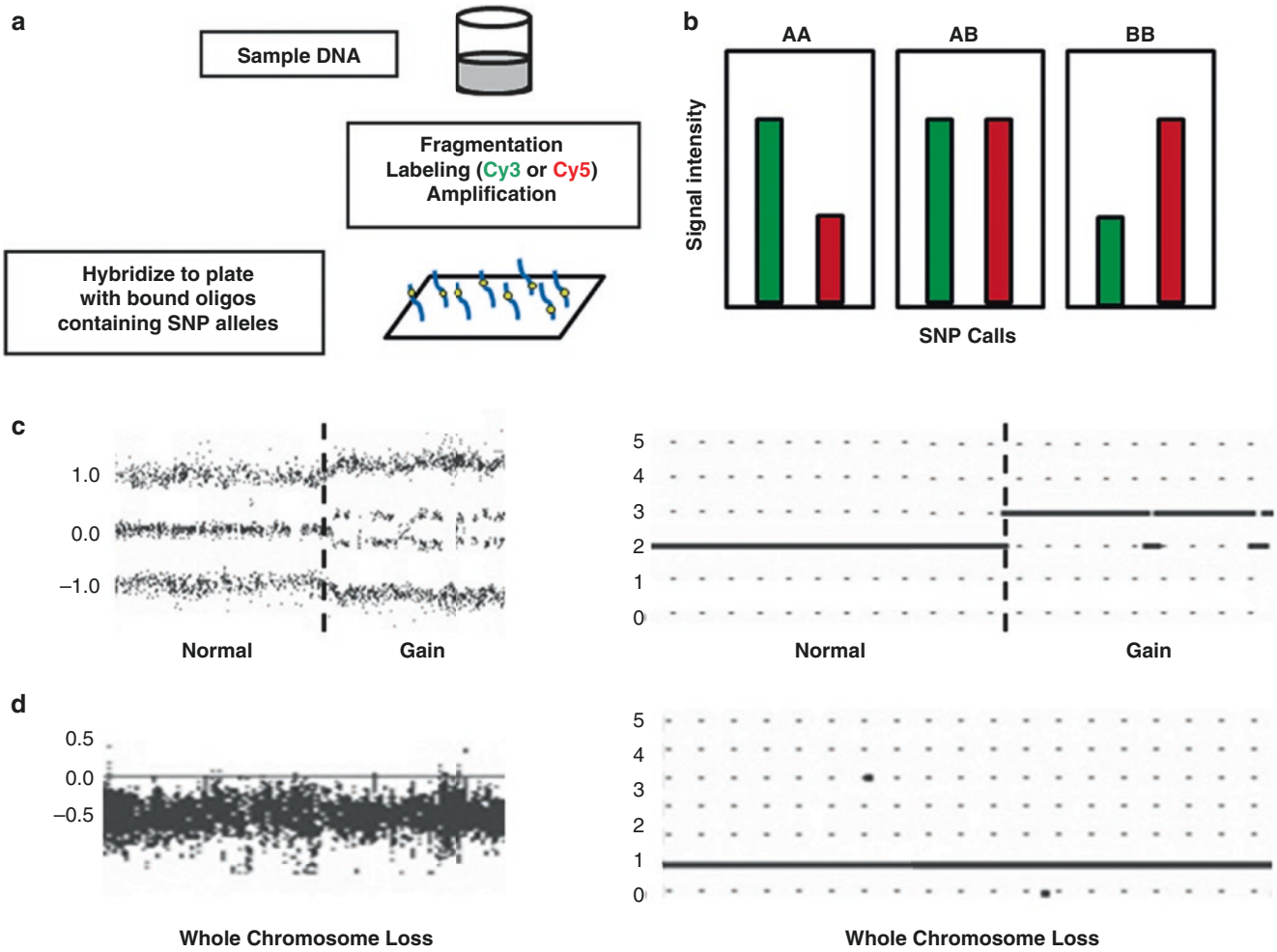


Fig. 21.8 Principles of a chip-based (Affymetrix) SNP array. (**a**) Sample DNA is fragmented and labeled with fluorescent dye (e.g., Cy3 and Cy5). DNA is hybridized to oligonucleotide probes (*blue*) corresponding to the individual SNP alleles (*yellow circles*). (**b**) Amplified and labeled DNA is hybridized to probes corresponding to alleles for each SNP locus. This results in a genotyping pattern allowing for determination of the heterozygosity or homozygosity for each allele. At the same time, intensity of the hybridization signals allows for determination of copy number changes. Various software packages allow for generation of karyotyping maps. (**c**) Example data from an Affymetrix 6.0 SNP array showing segmental gain of chromosomal material. The *left panel* is an example of an "allele peak" view. On the *left side* of the *dotted line*, there is a mixture of signal intensities of the arbitrarily designed "B" allele. In the normal state (*left* of the *dotted line*), there are equal mixtures of loci homozygous for the B SNP allele (+1), heterozygous for the AB allele (0), and homozygous for the A SNP allele (−1). A segmental gain of chromosomal material is shown to the *right* of the *dotted line*. The "allele peak" view now shows the presence of three alleles: BBB (*top line*, +1.5), BBA (*second line from top*, +0.75), BAA (*third line from top*, −0.75), and AAA (*bottom line*, −1.5). The "copy number state" is shown in the *right panel*. This assists in the identification of three alleles. (**d**) Example data from an Affymetrix 6.0 SNP array showing complete loss of chromosomal material from 5q. The *left panel* is an example of a "weighted log2" view indicating overall loss of fluorescent signal. The "copy number state" is shown in the *right panel*. This shows one copy of the allele and is interpreted as a complete loss of genetic material

[48]. Also in CLL, investigators found novel mutations in MAX pathway genes that are involved in regulatory mechanisms for cell proliferation, differentiation, and apoptosis [49]. In plasma cell myeloma, SNP arrays consistently recapitulate the findings of FISH analysis and provide further information regarding the particular genes deleted within chromosomal regions associated with poor prognoses [e.g., *CYLD* and *WWOX* in del(16q)] [50]. Perhaps the most significant application for SNP array analysis in hematologic malignancies has been the discovery of prognostically relevant genomic alterations in MDS and AML without recurrent or defined chromosomal translocations [51–53].

SNPs exist in a binary fashion (either SNP A or SNP B). Therefore, a single-bead or chip oligonucleotide corresponds to a single allele at the SNP locus. Each SNP allele produces a specific color indicator (e.g., red for SNP A and blue for SNP B) when sample DNA is bound. When genomic DNA binds, the relative intensities of red and blue signal at individual SNP loci are evaluated to determine three possible genotypes: homozygous A/A or B/B or heterozygous A/B

[54]. Importantly, the hybridization signals of tumor DNA across more than a million SNP loci are compared to normal diploid DNA (usually buccal mucosa DNA) at the individual probe sites [55]. The comparison of tumor sample to matched normal DNA allows for evaluation of chromosomal abnormalities. The abundance of one color (SNP allele) in the tumor sample could represent LOH, either secondary to chromosomal deletion or to CN-LOH (Fig. 21.8d).

The source of reference DNA is important in SNP analysis. When a SNP array sample is compared to reference DNA instead of paired normal sample from the same individual, there is a greater risk of a "miscall." In the most common scenario, an individual harbors an inherited copy number variation (CNV, a region of the genome with polymorphous gene segments). CNVs appear frequently in the genome, but the same CNVs are usually not seen among many persons. This has complicated the validation of these regions in reference databases such as the database of genomic variants (DGV) [56]. Virtually all CNV and SNP databases are incomplete and lack comprehensive validation. Thus, it is recommended to simultaneously analyze normal, nonneoplastic DNA from the same patient [41, 54, 57]. Buccal mucosal epithelial cell swabs are routinely used as a source of normal DNA, but buccal mucosa can be contaminated with blood, thus introducing tumor DNA into the supposed "normal sample." A skin biopsy may be a source of optimal normal DNA but is a more invasive procedure than a simple mouth swab [54].

Although SNP arrays interrogate a significant portion of the genome in terms of SNP distribution, they still cover less than 0.1% of the whole genome. With current technology, SNP arrays cannot be designed to distinguish every single-base-pair change throughout the genome [58]. Thus, they cannot be used to evaluate point mutations in genes in which these are common, such as *FLT3*, *CKIT*, or *PDGFR*. SNP arrays also are unable to detect balanced chromosomal translocations. To resolve this, SNP array analysis is applied subsequent to tests that readily identify translocations and point mutations such as routine cytogenetic analysis/FISH or single-gene assays, respectively.

The ability to interrogate all base pairs in hematologic malignancies would provide a comprehensive assessment of the neoplastic genotype and ensure that all clinically relevant information would be obtained from a patient sample. Assays that interrogate the whole genome at the base pair level have recently been introduced to assess this information.

Multigene and Whole-Genome Assays

As the name implies, whole-genome assays (at least theoretically) interrogate all nucleotides present in the entire genome or, in the case of whole-exome sequencing (WES), interrogate all the base pairs in the coding regions of the genome. Multigene panels typically are accomplished through selective capture or amplification of specific genes to be sequenced. Next-generation sequencing (NGS) platforms can sequence the entire genome in a single experimental run, because modern computers possess the massive computing power necessary to manipulate billions of data points simultaneously. The time and cost required to sequence the entire genome have dropped precipitously. In 2001, it took a staggering 10 years and 2.7 billion US dollars to sequence the entire genome. Approximately a decade later, it took an equally staggering few weeks and a few thousand US dollars to perform the same task [59]. The expenses and time needed continue to drop and now can be performed for less than 1000 USD and completed in less than 1 week [60]. By coupling enormous clinical potential with legitimate affordability and reasonable turnaround time, NGS techniques are projected to become a mainstay for the diagnosis and treatment of hematologic malignancies in the imminent future.

Next-Generation Sequencing (NGS)

Two benchtop NGS manufacturers, Illumina and Life Technologies, have emerged as the industry leaders and are used in both academic and commercial laboratories for hematolymphoid malignancy testing, although other manufacturers are developing clinical-grade sequencing platforms. Regardless of the platform, intact genomic DNA must be partitioned into workable amplifiable fragments such that template DNA can undergo massively parallel DNA sequencing. Current methods generally involve randomly breaking genomic DNA into smaller sizes (either by sonication or restriction digestion), amplifying the template DNA (e.g., creating a DNA library), and anchoring the fragments to solid-phase components. These methods can be also used to sequence specific genes by capturing or amplifying the specific gene regions to be sequenced. The solid-phase anchored fragmented DNA is partitioned in such a way such that simultaneous sequencing reactions can occur. These sequencing reactions are performed on amplified DNA fragments (of the same template sequence) because most imaging systems cannot detect single template fluorescent or luminescent events [60, 61]. The ability to detect the sequence of millions of individual partitioned fragments of genomic DNA simultaneously is the sine qua non of NGS [60–65].

Each benchtop NGS platform performs massive parallel sequencing using a different methodology, and as a result, each of the two has differential performance characteristics. At least for bacterial genomes, the Illumina MiSeq tends to generate the highest throughput per run with the lowest number of errors but delivers short reads. The Ion Torrent PGM produces the shortest reads with the most errors but with the fastest throughput and shortest run time [66, 67]. Both NGS platforms are currently widely used in clinical laboratories,

but it is unclear whether one platform (or another platform in development) will emerge as the "gold standard" for clinical use. Still, for the purposes of this chapter, the application of NGS to hematologic malignancies is the same, regardless of which platform is used.

Applications of NGS to Hematologic Malignancies

The goal of NGS in hematologic malignancies is to identify and interpret genetic variation between the neoplastic population and the matched germline DNA [68]. When optimized, NGS can detect point mutations, insertions, deletions, and chromosomal rearrangements making it an ideal platform to assess known clinically relevant mutations and for the discovery of new mutations or mutations previously undetected by standard methodologies [69].

One of the first documented applications of NGS to hematologic malignancies was performed in 2008 when investigators used NGS to sequence "cytogenetically normal" (by routine karyotyping and FISH analysis) AMLs. They discovered ten somatic mutations in exon-coding regions. Of the ten mutations, two mutations, *FLT3*-ITDs and *NPM1* exon 12 insertions, were known to be mutated in AML, and the other eight novel mutations and the associated genes are currently under intense investigation to determine their contribution to the pathogenesis of AML [70]. This landmark paper served as a "proof-of-principle" study: one could feasibly use NGS to identify mutations in hematologic malignancies that other methods had failed to detect. A few years later, novel recurrent mutations in *DNMT3A* were discovered in approximately 20% of AML cases and found to be associated with poor overall survival but predictive of improved survival in patients less than 60 years of age treated with high-dose induction chemotherapy [68, 71]. In MDS, NGS identified mutations in *TP53*, *EZH2*, *ETV6*, *RUNX1*, and *ASXL1* that were found to be predictors of poor overall survival [72]. In multiple myeloma, investigators discovered previously unknown point mutations in *KRAS*, *BRAF*, and *NRAS* [73]. New mutations were also discovered in CLL, and NGS has been investigated as a tool to molecularly monitor clonal evolution in pediatric acute leukemia patients [74–76].

NGS has been reported to recapitulate the results generated by traditional single-gene assays, such as *FLT3* mutations and *JAK2* p.V617F mutations [70, 77, 78]. There are also reports that NGS can identify balanced translocations such as *BCR-ABL1* t(9;22)(q34;q11.2) and cryptic translocations not identified by routine cytogenetics [69, 79]. Combined NGS assays may also be able to detect and monitor clonal rearrangements of the immunoglobulin and T-cell receptor loci as well as translocations involving the immunoglobulin genes [80].

The use of NGS platforms has also found favor in monitoring residual disease and disease recurrence. NGS platforms are able to "look for every aberration" in contrast to conventional methods such as cyogenetics, FISH, or PCR that are designed to assess for predetermined chromosomal abnormalities or mutations. This unique property allows for the detection of tumor subclones that were either present at the initiation of therapy or are evolving in the presence of therapy. Examples of this "escape clone monitoring" were recently described in AML. Investigators found two major clonal evolution patterns during AML relapse: one, primary tumor clone acquired additional mutations that evolved into the relapse clone, and two, a subclone of the primary tumor clone survived initial therapy, gained additional mutations, and expanded at relapse. These data also suggested that AML cells routinely acquire additional mutations at relapse, and some of these mutations may contribute to clonal selection and chemotherapy resistance [81]. Clonal evolution monitoring with NGS is also described in drug-resistant *BCR-ABL1* mutants [82] and chronic lymphocytic leukemia [83]. As NGS studies accumulate, it is postulated that more sophisticated monitoring and treatment protocols will arise.

The comprehensive data acquired from NGS are obviously attractive for use as a clinical tool. However, the trade-off for generating many parallel short templates is loss of sequencing accuracy. NGS platforms have approximately tenfold higher error rates in base pair reads (1 in 1000 bases at 20-fold coverage) versus Sanger sequencing (1 in 10,000 bases). The depth of sequencing coverage (the number of times a single fragment is amplified) may be insufficient to identify single-point mutations in limited sample sizes; neoplastic tissue that comprises 25% of the input sample that is sequenced at 30-fold coverage still produces an error rate of 5% [84]. Certain highly repetitive regions of the genome are difficult to examine accurately, partly due to the algorithms used to align the sequencing data (68). WES allows greater depth of coverage to more accurately detect point mutations in the 1–2% of the genome that constitutes the protein-coding regions. However, certain portions of the exome are also subject to the same sequencing difficulties, and pertinent mutations in intronic DNA will be missed. To address the technological challenges inherit to NGS, consensus guidelines for test validation, quality control, proficiency testing, and quality control for NGS testing in clinical laboratories are available and continue to evolve [85].

Both WGS and WES generate tremendous amounts of data that pose unprecedented informatics challenges to analyze, interpret, retrieve, and store, particularly, in a Health Insurance Portability and Accountability Act (HIPAA)-compliant manner. A recent report highlighted some of the data analysis challenges that NGS sequencing poses. For the detection of *FLT3* internal tandem duplication, only one of seven software analysis packages (Pindel) reliably detected

the aberration with 100% sensitivity and specificity. Some of the software programs did not detect the duplication in any sample [86]. These data raise questions regarding the "in silico" interpretation of NGS data that are beyond the scope of this chapter.

Suffice it to say, guidelines and recommendations governing the broad clinical application of WGS or WES to hematologic malignancies will require input from pathologists, clinicians, and informatics specialists, among others, to meet the rigorous quality demands required for clinical laboratory testing.

Conclusion

The genomic applications to hematologic malignancies are diverse but are essential for both diagnosis and clinical management, and each testing methodology has utility in the appropriate clinical context. Choosing the most fitting test requires a fund of knowledge for both the disease entity and the testing methods. Single-gene testing, routine cytogenetic karyotyping, FISH, aCGH, SNP array, WES, and WGS analyses all possess innate utility and limitations when the clinical question is clearly delineated. As the discoveries and applications of innovative technologies in genomic medicine will continue to evolve, it is the medical professional's responsibility to become familiar with all genomic testing methodologies, in order to request the most suitable test for the diagnosis and management of patients with hematologic malignancies.

References

1. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006;24:541–70.
2. Bruns DE, Ashwood ER, Burtis CA. Fundamentals of molecular diagnostics. St. Louis: Saunders Elsevier; 2007.
3. Nikiforova MN, Hsi ED, Braziel RM, Gulley ML, Leonard DGB, Nowak JA, et al. Detection of clonal IGH gene rearrangements: summary of molecular oncology surveys of the College of American Pathologists. Arch Pathol Lab Med. 2007;131(2):185–9.
4. Mannu C, Gazzola A, Bacci F, Sabattini E, Sagramoso C, Roncolato F, et al. Use of IGK gene rearrangement analysis for clonality assessment of lymphoid malignancies: a single center experience. Am J Blood Res. 2011;1(2):167–74.
5. Tapia G, Sanz C, Mate JL, Munoz-Marmol AM, Ariza A. Improved clonality detection in Hodgkin lymphoma using the BIOMED-2-based heavy and kappa chain assay: a paraffin-embedded tissue study. Histopathology. 2012;60(5):768–73.
6. Jevremovic D, Viswanatha DS. Molecular diagnosis of hematopoietic and lymphoid neoplasms. Hematol Oncol Clin North Am. 2009;23(4):903–33.
7. Swerdlow SH, Cancer IAfRo, Organization WH. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: International Agency for Research on Cancer; 2008.
8. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. Leukemia. 2003;17(12):2257–317.
9. Leonard DGB. Diagnostic molecular pathology. Philadelphia: W.B. Saunders; 2003.
10. Patnaik MM, Tefferi A. Molecular diagnosis of myeloproliferative neoplasms. Expert Rev Mol Diagn. 2009;9(5):481–92.
11. Tefferi A, Skoda R, Vardiman JW. Myeloproliferative neoplasms: contemporary diagnosis using histology and genetics. Nat Rev Clin Oncol. 2009;6(11):627–37.
12. Hammond E, Shaw K, Carnley B, P'ng S, James I, Herrmann R. Quantitative determination of JAK2 V617F by TaqMan: an absolute measure of averaged copies per cell that may be associated with the different types of myeloproliferative disorders. J Mol Diagn. 2007;9(2):242–8.
13. Lippert E, Girodon F, Hammond E, Jelinek J, Reading NS, Fehse B, et al. Concordance of assays designed for the quantification of JAK2V617F: a multicenter study. Haematologica. 2009;94(1):38–45.
14. Barosi G, Birgegard G, Finazzi G, Griesshammer M, Harrison C, Hasselbalch HC, et al. Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. Blood. 2009;113(20):4829–33.
15. Tefferi A, Barbui T. Polycythemia vera and essential thrombocythemia: 2017 update on diagnosis, risk-stratification, and management. Am J Hematol. 2017;92(1):94–108. Frohling S, Scholl C, Gilliland DG, Levine RL. Genetics of myeloid malignancies: pathogenetic and clinical implications. J Clin Oncol 2005;23(26):6285–95.
16. Murphy KM, Levis M, Hafez MJ, Geiger T, Cooper LC, Smith BD, et al. Detection of FLT3 internal tandem duplication and D835 mutations by a multiplex polymerase chain reaction and capillary electrophoresis assay. J Mol Diagn. 2003;5(2):96–102.
17. Dekking E, van der Velden VHJ, Böttcher S, Brüggemann M, Sonneveld E, Koning-Goedheer A, et al. Detection of fusion genes at the protein level in leukemia patients via the flow cytometric immunobead assay. Best practice & amp. Res Clin Haematol. 2010;23(3):333–45.
18. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med. 2003;348(11):994–1004.
19. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med. 2006;355(23):2408–17.
20. White HE, Matejtschuk P, Rigsby P, Gabert J, Lin F, Lynn Wang Y, et al. Establishment of the first World Health Organization International Genetic Reference Panel for quantitation of BCR-ABL mRNA. Blood. 2010;116(22):e111–7.
21. Wang N. Methodologies in cancer cytogenetics and molecular cytogenetics. Am J Med Genet. 2002;115(3):118–24.
22. Bejjani BA, Saleki R, Ballif BC, Rorem EA, Sundin K, Theisen A, et al. Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more? Am J Med Genet A. 2005;134(3):259–67.
23. Sharathkumar A, DeCamillo D, Bhambhani K, Cushing B, Thomas R, Mohamed AN, et al. Children with hyperdiploid but not triple trisomy (+4,+10,+17) acute lymphoblastic leukemia have an increased incidence of extramedullary relapse on current therapies: a single institution experience. Am J Hematol. 2008;83(1):34–40.
24. Marschalek R. Mechanisms of leukemogenesis by MLL fusion proteins. Br J Haematol. 2011;152(2):141–54.

25. Jacoby MA, Walter MJ. Detection of copy number alterations in acute myeloid leukemia and myelodysplastic syndromes. Expert Rev Mol Diagn. 2012;12(3):253–64.
26. Shinawi M, Cheung SW. The array CGH and its clinical applications. Drug Discov Today. 2008;13(17–18):760–70.
27. Shaikh TH. Oligonucleotide arrays for high-resolution analysis of copy number alteration in mental retardation/multiple congenital anomalies. Genet Med. 2007;9(9):617–25.
28. Shao L, Kang S-HL, Li J, Hixson P, Taylor J, Yatsenko SA, et al. Array comparative genomic hybridization detects chromosomal abnormalities in hematological cancers that are not detected by conventional cytogenetics. J Mol Diagn. 2010;12(5):670–9.
29. Kolquist KA, Schultz RA, Furrow A, Brown TC, Han J-Y, Campbell LJ, et al. Microarray-based comparative genomic hybridization of cancer targets reveals novel, recurrent genetic aberrations in the myelodysplastic syndromes. Cancer Genet. 2011;204(11):603–28.
30. Paulsson K, Heidenblad M, Strombeck B, Staaf J, Jonsson G, Borg A, et al. High-resolution genome-wide array-based comparative genome hybridization reveals cryptic chromosome changes in AML and MDS cases with trisomy 8 as the sole cytogenetic aberration. Leukemia. 2006;20(5):840–6.
31. Thiel A, Beier M, Ingenhag D, Servan K, Hein M, Moeller V, et al. Comprehensive array CGH of normal karyotype myelodysplastic syndromes reveals hidden recurrent and individual genomic copy number alterations with prognostic relevance. Leukemia. 2011;25(3):387–99.
32. Vercauteren SM, Sung S, Starczynowski DT, Lam WL, Bruyere H, Horsman DE, et al. Array comparative genomic hybridization of peripheral blood granulocytes of patients with myelodysplastic syndrome detects karyotypic abnormalities. American Journal of Clinical Pathology. 2010;134(1):119–26.
33. Dawson AJ, Yanofsky R, Vallente R, Bal S, Schroedter I, Liang L, et al. Array comparative genomic hybridization and cytogenetic analysis in pediatric acute leukemias. Curr Oncol. 2011;18(5):e210–7.
34. Lucioni M, Novara F, Fiandrino G, Riboni R, Fanoni D, Arra M, et al. Twenty-one cases of blastic plasmacytoid dendritic cell neoplasm: focus on biallelic locus 9p21.3 deletion. Blood. 2011;118(17):4591–4.
35. Higgins RA, Gunn SR, Robetorye RS. Clinical application of array-based comparative genomic hybridization for the identification of prognostically important genetic alterations in chronic lymphocytic leukemia. Mol Diagn Ther. 2008;12(5):271–80.
36. Okada M, Suto Y, Hirai M, Shiseki M, Usami A, Okajima K, et al. Microarray CGH analyses of chromosomal 20q deletions in patients with hematopoietic malignancies. Cancer Genet. 2012;205(1–2):18–24.
37. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928–33.
38. Ersland KM, Christoforou A, Stansberg C, Espeseth T, Mattheisen M, Mattingsdal M, et al. Gene-based analysis of regionally enriched cortical genes in GWAS data sets of cognitive traits and psychiatric disorders. PLoS One. 2012;7(2):e31687.
39. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int. 2006;164(1):20–32.
40. Tzvetkov M, von Ahsen N. Pharmacogenetic screening for drug therapy: from single gene markers to decision making in the next generation sequencing era. Pathology. 2012;44(2):166–80.
41. Sato-Otsubo A, Sanada M, Ogawa S. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? Semin Oncol. 2012;39(1):13–25.
42. Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, et al. [3] Illumina Universal Bead Arrays. In: Alan K, Brian O, editors. Methods in enzymology. New York: Academic Press; 2006. p. 57–73.
43. Murray SS, Oliphant A, Shen R, McBride C, Steeke RJ, Shannon SG, et al. A highly informative SNP linkage panel for human genetic studies. Nat Meth. 2004;1(2):113–7. https://doi.org/10.1038/nmeth712.
44. Maciejewski JP, Tiu RV, O'Keefe C. Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. Br J Haematol. 2009;146(5):479–88.
45. Maciejewski JP, Mufti GJ. Whole genome scanning as a cytogenetic tool in hematologic malignancies. Blood. 2008;112(4):965–74.
46. Kawamata N, Ogawa S, Zimmermann M, Niebuhr B, Stocking C, Sanada M, et al. Cloning of genes involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray. Proc Natl Acad Sci U S A. 2008;105(33):11921–6.
47. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007;446(7137):758–64.
48. Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007;109(3):1202–10.
49. Edelmann J, Holzmann K, Miller F, Winkler D, Buhler A, Zenz T, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. Blood. 2012;120(24):4783–94.
50. Jenner MW, Leone PE, Walker BA, Ross FM, Johnson DC, Gonzalez D, et al. Gene mapping and expression analysis of 16q loss of heterozygosity identifies WWOX and CYLD as being important in determining clinical outcome in multiple myeloma. Blood. 2007;110(9):3291–300.
51. Parkin B, Erba H, Ouillette P, Roulston D, Purkayastha A, Karp J, et al. Acquired genomic copy number aberrations and survival in adult acute myelogenous leukemia. Blood. 2010;116(23):4958–67.
52. Tiu RV, Gondek LP, O'Keefe CL, Elson P, Huh J, Mohamedali A, et al. Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. Blood. 2011;117(17):4552–60.
53. Yi JH, Huh J, Kim HJ, Kim SH, Kim YK, Sohn SK, et al. Adverse prognostic impact of abnormal lesions detected by genome-wide single nucleotide polymorphism array-based karyotyping analysis in acute myeloid leukemia with normal karyotype. J Clin Oncol. 2011;29(35):4702–8.
54. Heinrichs S, Li C, Look AT. SNP array analysis in hematologic malignancies: avoiding false discoveries. Blood. 2010;115(21):4157–61.
55. Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, et al. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet. 2004;36(3):299–303.
56. Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet Genome Res. 2006;115(3–4):205–14.
57. Le Scouarnec S, Gribble SM. Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. Heredity (Edinb). 2012;108(1):75–85.
58. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 2009;37(13):4181–93.
59. Venter JC. Multiple personal genomes await. Nature. 2010;464(7289):676–7. https://doi.org/10.1038/464676a.
60. https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/. Accessed 10 Mar 2018; Metzker ML. Sequencing technologies [mdash] the next generation. Nat Rev Genet. https://doi.org/10.1038/nrg2626. 2010;11(1):31–46.
61. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348–52. https://doi.org/10.1038/nature10242.

62. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53–9.
63. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A. 2003;100(15):8817–22.
64. Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD, et al. A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. Electrophoresis. 2003;24(21):3769–77.
65. Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science. 1998;281(5375):363–5.
66. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
67. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotech. 2012;30(5):434–9. https://doi.org/10.1038/nbt.2198.
68. Merker JD, Valouev A, Gotlib J. Next-generation sequencing in hematologic malignancies: what will be the dividends? Ther Adv Hematol. 2012;3(6):333–9.
69. Welch Js WPDL, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011;305(15):1577–84.
70. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66–72.
71. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med. 2012;366(12):1079–89.
72. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011;364(26):2496–506.
73. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471(7339):467–72. https://doi.org/10.1038/nature09837.
74. Campana D. Minimal residual disease monitoring in childhood acute lymphoblastic leukemia. Curr Opin Hematol. 2012;19(4):313–8.
75. Gawad C, Pepin F, Carlton VE, Klinger M, Logan AC, Miklos DB, et al. Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. Blood. 2012;120(22):4407–17.
76. Ramsay AJ, Martinez-Trillos A, Jares P, Rodriguez D, Kwarciak A, Quesada V. Next-generation sequencing reveals the secrets of the chronic lymphocytic leukemia genome. Clin Transl Oncol, 2013;15(1):3–8;
77. Benichou J, Ben-Hamo R, Louzoun Y, Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. Immunology. 2012;135(3):183–91.
78. Kohlmann A, Grossmann V, Haferlach T. Integration of next-generation sequencing into clinical practice: are we there yet? Semin Oncol. 2012;39(1):26–36.
79. Shaffer LG, Schultz RA, Ballif BC. The use of new technologies in the detection of balanced translocations in hematologic disorders. Curr Opin Genet Dev. 2012;22(3):264–71.
80. Wren D, Walker BA, Brüggemann M, Catherwood MA, Pott C, Stamatopoulos K, et al. Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. Haematologica. 2017;102(2):e57–60.
81. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506–10. https://doi.org/10.1038/nature10738.
82. Soverini S, De Benedittis C, Machova Polakova K, Brouckova A, Horner D, Iacono M, et al. Unraveling the complexity of tyrosine kinase inhibitor-resistant populations by ultra-deep sequencing of the BCR-ABL kinase domain. Blood. 2013;21:2013.
83. Landau Dan A, Carter Scott L, Stojanov P, McKenna A, Stevenson K, Lawrence Michael S, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714–26.
84. Gullapalli RR, Lyons-Weiler M, Petrosko P, Dhir R, Becich MJ, LaFramboise WA. Clinical integration of next-generation sequencing technology. Clin Lab Med. 2012;32(4):585–99.
85. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotech. 2012;30(11):1033–6. https://doi.org/10.1038/nbt.2403.
86. Spencer DH, Abel HJ, Lockwood CM, Payton JE, Szankasi P, Kelley TW, et al. Detection of FLT3 internal tandem duplication in targeted, short-read-length, next-generation sequencing data. J Mol Diagn. 2013;15(1):81–93.
87. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Garcia EW, McBride C, Steemers F, Garcia F, Kermani BG, Gunderson K, Oliphant A. High throughput SNP genotyping on universal bead arrays. Mutat Res. 2005;573(1–2):70–82.

Genomic Applications in Brain Tumors

22

Matija Snuderl

Introduction

The field of neuro-oncology has evolved significantly over the past two decades. Because brain tumors are difficult to culture, discoveries of genomic rearrangements made by classic cytogenetics in other fields of oncology have not been possible. A new era has started with the implementation of molecular techniques, which led to discoveries of novel diagnostic, prognostic, and predictive molecular markers. Whereas histological classification of tumors relies on the morphological features, targeted and genome-wide approaches led to deeper understanding of brain tumor biology and molecular subclassification of morphologic entities. Recent advances in genome-wide techniques have also discovered genes underlying previously well-known aberrations. For example, a co-deletion of 1p/19q has long been accepted in the field as a predictive marker in oligodendroglioma and has recently been adopted as a defining feature. Whole-genome sequencing discovered the underlying role of the *CIC* gene in oligodendroglioma biology. While many methods are still for research use only, there has been increased implementation of molecular tests in diagnosis and management of brain tumors. A variety of assays have been designed to analyze chromosomal rearrangements, copy number changes, point mutations, and epigenetic changes. This is particularly important because most malignant brain tumors have largely resisted standard chemotherapy and radiation therapy and will require more targeted approaches based on the specific biology of the tumors. Genome, transcriptome, and epigenome analyses will likely become a focus for diagnostics and for identifying therapeutic targets.

Gliomas are the most common tumors of the central nervous system (CNS) and often require additional molecular workup, either for diagnosis or for clinical management. In clinical practice, a single gene or target region is often evaluated. The most commonly used assays include analyses of 1p/19q, *MGMT* methylation, and *IDH1/2* mutation status [1–3]. From a technical point of view, they include fluorescence in situ hybridization (FISH), the polymerase chain reaction (PCR), a variety of methylation-specific assays, and sequencing or immunohistochemistry (IHC), DNA arrays, and DNA or RNA next-generation sequencing. These approaches can provide important diagnostic and prognostic or predictive information, particularly in diffuse gliomas. Clinical practice has been revolutionized by introduction of mutation-specific antibodies including IDH1 R132H, BRAF V600E, ATRX, and histone H3 K27M, which enable laboratories without easy access to sequencing to assess the most common genetic alterations for diagnosis and prognosis. Genome-based analysis of the expression profile of medulloblastomas, supported later by copy number and mutation analysis, has pioneered subclassification of a single disease based on molecular characteristics. The molecular complexity of brain tumors, such as gliomas, meningiomas, ependymomas, and medulloblastomas, requires large panel next-generation DNA or RNA sequencing. One of the most promising strategies has been whole-genome DNA methylation profiling. Large panel, exome, and genome scales are becoming more feasible and cost-effective solutions. Utilizing epigenetic signature of DNA methylation allows us to accurately classify brain tumors. With the costs of whole-genome analyses decreasing, one can expect that a many specific assays designed for particular targets will be replaced by a panel that is able to evaluate numerous genes of interest.

M. Snuderl (✉)
Department of Pathology, NYU Langone Medical Center and Medical School, New York, NY, USA
e-mail: matija.snuderl@nyumc.org

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_22

Targeted Genomic Assays Used in the Clinical Evaluation of Brain Tumors

1p/19q loss, IDH1/2, and TERT Promoter Mutations

Loss of chromosomal arms 1p and 19q is the defining molecular feature of oligodendroglial neoplasms.

Numerous studies have confirmed an association between 1p/19q co-deletion and a favorable response to chemotherapy, initially to procarbazine, lomustine, and vincristine, and later to temozolomide, as well as to radiotherapy. Therefore, testing for 1p/19q loss is considered the standard of care, and most neuro-oncologists will use 1p/19q status to make therapeutic decisions. Many will withhold radiation therapy upfront, even in case of a small residual tumor after surgery, in order to avoid the risk of long-term toxicity and choose chemotherapy with temozolomide, or even careful monitoring alone. Radiation, therefore, remains as an option in case of progression. From a diagnostic point of view, 1p/19q loss can help to distinguish oligodendrogliomas from morphologically similar neoplasms such as neurocytomas, clear cell ependymomas and meningiomas, dysembryoplastic neuroepithelial tumors (DNETs), or small cell variant of glioblastoma. The loss of 1p and 19q is mediated by formation of a balanced whole-arm translocation involving chromosomes 1 and 19, with subsequent loss of the derivative chromosome der(1;19)(p10;q10) and maintenance of der(1;19)(q10;p10). The genes responsible for tumorigenesis of oligodendroglioma were enigmatic until recently, when several whole-genome sequencing studies have identified *CIC* (Fig. 22.1) and *FUBP1* gene mutations [4, 5].

In laboratory practice, PCR-based loss of heterozygosity studies and FISH are the most commonly used methods to detect 1p/19q loss. Other less common methods would include arrayed comparative genomic hybridization (aCGH) and multiplex ligation-dependent probe amplification (MLPA). MLPA only requires standard PCR instrumentation and capillary gel electrophoreses; however, it does not require the patient's normal DNA sample. Both FISH and PCR methods are technically straightforward but have some advantages and disadvantages. Loss of heterozygosity analysis is a PCR-based method; the major disadvantage of which is the necessity of obtaining a normal blood sample. This can be complicated if blood is not collected at the time of surgery and the patient is discharged when the diagnostic dilemma arises. FISH scoring can be time-consuming but can provide additional prognostic information. By FISH, tumor cells with 1p/19q present would show two signals for 1p and two for 19q and two control signals of 1q and 19p, respectively (Fig. 22.2a). A typical co-deletion pattern would have nuclei with two signals for 1q and one for 1p and nuclei with two signals for 19p and only one for 19q (Fig. 22.2b). This is called an absolute deletion [6]. However, some tumors are characterized by polysomy, i.e., gains of either chromosome 1 or 19 or both, with concurrent loss of 1p/19q [7]. These are sometimes referred to as relative deletions. Nuclei will have four or more 1q signals and two or more 1p signals (Fig. 22.2c) or four or more 19p signals and two or more 19q signals. Several studies have confirmed that the additional data about polysomy and relative deletions seen by FISH but not by loss of heterozygosity analysis provide important prognostic information. Concurrent loss of 1p/19q and polysomy predicts early recurrence and poor survival [7, 8]. Therefore, for the clinical assessment of 1p/19q loss, FISH provides better predictive value compared to loss of heterozygosity.

Importantly, the 1p/19q status is an early event and is shared among tumor cells throughout the tumor. Therefore, tumor heterogeneity is not an issue for testing, and any focus of the tumor can be evaluated. Interestingly, association between a brain site and 1p/19q status has been well documented. Frontal lobe IDH1/2-mutated tumors are significantly more likely to also carry 1p/19q loss than IDH1/2-mutated gliomas from the temporal lobe. The biological reasons for that are currently unknown. Also, 1p/19q loss is exceedingly rare in the pediatric population and does not have the same diagnostic or prognostic importance. For these reasons, molecular results have to be interpreted together with histological and clinical data, and cannot completely replace a morphological diagnosis. 1p/19q status, WHO grade, morphological diagnosis, patient age, and performance score are independent statistically significant prognostic variables. A combination of molecular tests, using 1p/19q, IDH1/2, and TERT promoter status, enables to confidently diagnose majority of oligodendrogliomas, and diagnosis of mixed glioma and oligoastrocytoma can largely be abandoned unless molecular features are atypical [9].

IDH1/2, *ATRX*, and *TP53* Mutation Status in Diffuse Gliomas

The classification of diffuse gliomas has changed remarkably using molecular understanding of the disease. While previously diffuse gliomas were classified histologically into astrocytoma, oligodendroglioma, and mixed oligoastrocytoma, current classification largely follows the molecular criteria. Diffuse gliomas are classified primarily into IDH-mutated gliomas and IDH wild-type gliomas. IDH1/2 mutation leads to a gain of function phenotype and overproduction of oncometabolite 2-hydroxyglutarate. The effects

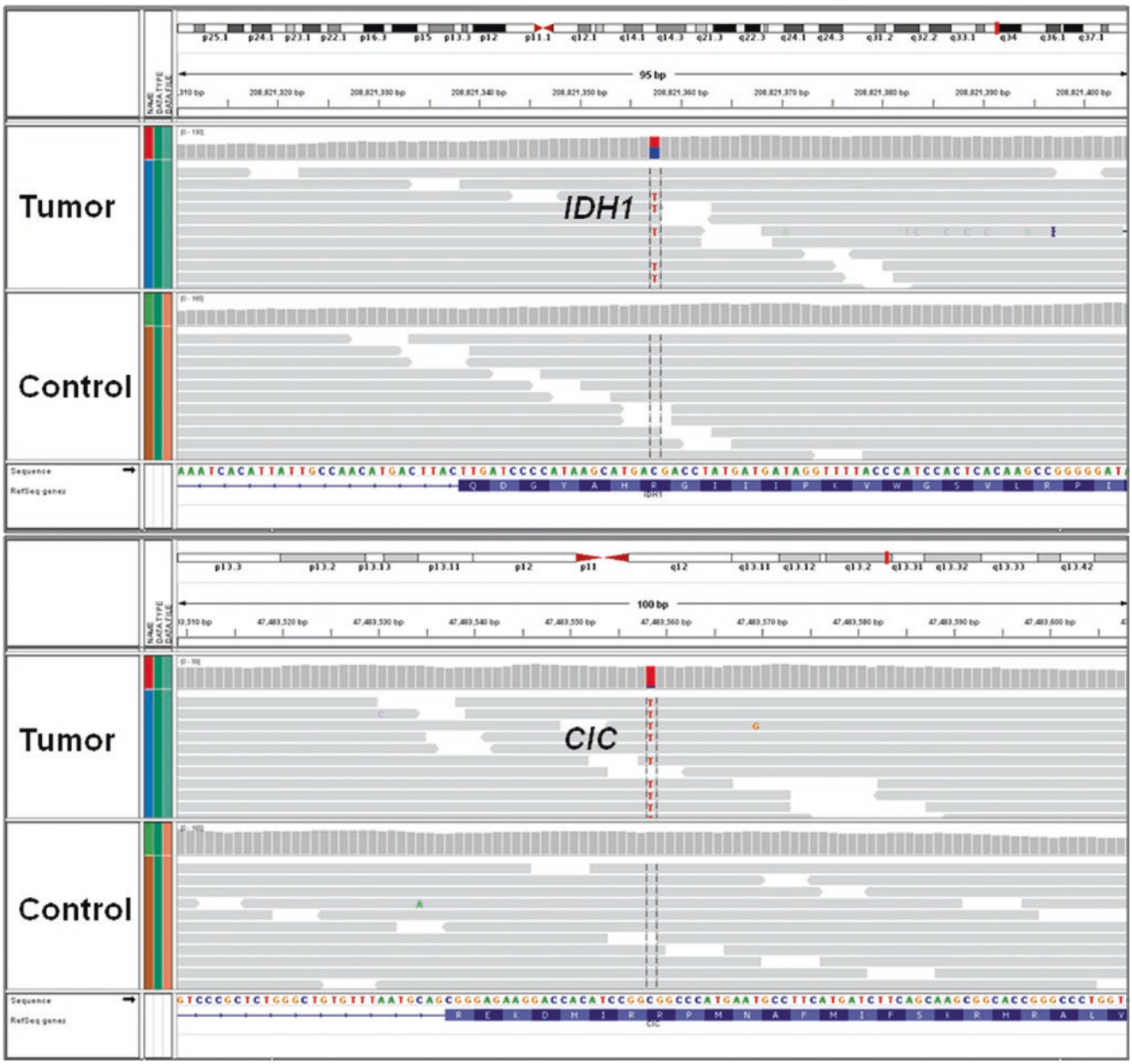


Fig. 22.1 Identification of *IDH1* and *CIC* mutations in oligodendroglioma. In oligodendroglioma, whole-genome sequencing (Illumina platform) identifies concurrent mutations in the *IDH1* (c. 395C > T, p.R132H) and *CIC* (c.604C > T, p.R202W) genes. The majority of *IDH1* mutations in gliomas are p.R132H. The majority of mutations in oligodendrogliomas with 1p19q loss and *IDH1* or *IDH2* mutations occur within exons 5 and 20 of the *CIC* gene. The example shown, in the form of the Integrative Genomics Viewer (IGV, Broad Institute) browser view, is from exon 5, which is a highly conserved DNA-interacting HMG domain. Novel non-synonymous mutations can be identified by filtering against the normal sequence pileup and by comparison with the dbSNP database. (Figure courtesy of Dr. Stephen Yip, BC Cancer Agency)

of 2-hydroxygluatarate are widespread and include the effect on tumor epigenome and induction of the G-CIMP phenotype with diffuse hypermethylation. In addition to cell-specific mechanism, IDH-mutated gliomas also have systemic effect, affecting systemic and local coagulation [10] and inducing seizures [11]. Most importantly, gliomas with mutations in IDH1 or IDH2 genes have significantly better outcome than wild-type tumors. However, the prognostic role of various *IDH1* and *IDH2* mutations is currently not known. IDH1/2-mutated tumors are further subclassified based on the presence of 1p/19q co-deletion, which defines an oligodendroglioma. *IDH1/2*-mutated tumors that show

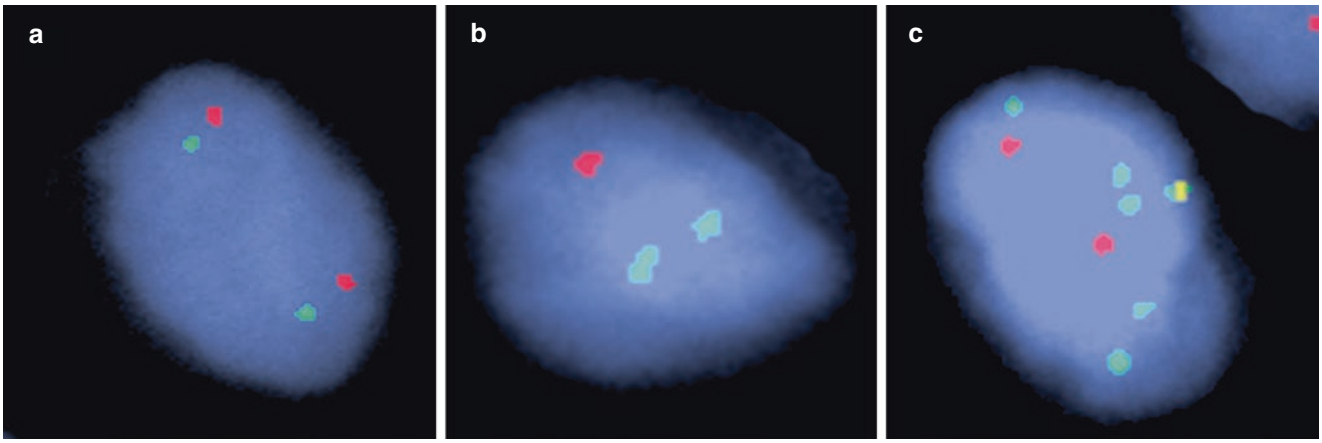


Fig. 22.2 Testing for 1p19q in oligodendroglioma by FISH. 1p19q testing is the standard of care for tumors with a suspected oligodendroglial component and serves as a diagnostic and predictive marker. FISH analysis includes control probes for 1q or 19p (*green signal*) and probes of interest for 1p or 19q (*red signal*). Panel (**a**) reflects the presence of 1p with two *red* and two *green* signals. In Panel (**b**), a classic absolute deletion with loss of one copy of 1p is demonstrated, while two 1q signals remain. Panel (**c**) illustrates the so-called relative deletion, also known as superloss, with numerous hybridized control probes of 1q indicating polysomy and loss of ~50% of the 1p signals. This tumor, by PCR, would appear to have loss of heterozygosity (LOH) similar to the tumor with the absolute deletion illustrated in Panel (**b**). However, tumors with superloss tend to be more aggressive

maintenance of chromosomal arms 1p and 19q usually carry mutations in both *ATRX* and *TP53*. While there certainly are tumors that carry molecular features of both oligodendroglioma and astrocytoma, i.e., bona fide molecular oligoastrocytomas [12], the utilization of histological diagnosis of mixed glioma oligoastrocytoma has significantly decreased with widespread introduction of molecular markers.

While histological grading criteria are still practiced, their significance is unclear. It has been well-documented that the presence or absence of IDH1/2 mutation is a better predictor of survival than histological grade or proliferation activity [13]. Despite better prognosis, certain molecular features are associated with rapid progression of *IDH1/2*-mutated gliomas. First is the hypermutation phenotype, which is most frequently therapy- induced [14]. The second is the presence of focal copy number aberrations such as *MYC* and *MDM2* [15].

IDH1/2 mutations are associated with younger patients and with the diagnosis of secondary glioblastoma (sGBM), which develops from a preexisting low-grade glioma (LGG). The majority of diffuse astrocytomas WHO grade II and anaplastic astrocytomas WHO grade III carry *IDH1/2* mutations [16]. *IDH1/2* mutations are also pathognomonic for oligodendroglioma and always associated with 1p/19q loss. This suggests that both low-grade astrocytomas and oligodendrogliomas might develop from a common precursor. The overall *IDH1/2* mutation frequency in astrocytomas is between 50% and 80%. In contrast, *IDH1* mutation is exceedingly rare in primary GBMs, which arise without a known low-grade precursor lesion. Because the frequency in other tumors is low, *IDH1* mutation is a very useful diagnostic marker. In the vast majority of cases, *IDH1* mutation affects codon 132 and is heterozygous, with the other gene copy remaining wild type. In gliomas, the most common mutation is p.R132H which represents about ~90% of mutations (Fig. 22.1), followed by p.R132C (4%), p.R132S, and p.R132G in approximately 1.5% of cases, each. Mutations in the *IDH2* gene are present in approximately 3% of gliomas. Gliomas with an *IDH1* mutation have a significantly better outcome than wild-type tumors, independent on type and grade. In addition to strong prognostic value, *IDH1* is also useful for diagnosis, where it can help in several ways. First, it distinguishes oligodendroglioma from other similar-looking neoplasms such as neurocytoma, clear cell ependymoma, pilocytic astrocytoma with a prominent oligodendroglial-like component, and dysembryoplastic neuroepithelial tumor (DNT), which all lack *IDH1* mutations. Second, it is very useful in distinguishing between diffuse glioma and reactive gliosis. The existence of a robust antibody to the mutant IDH1 p.R132H protein allows the establishment of this diagnosis even in very small samples and in samples with few infiltrating tumor cells [17]. Although the antibody is specific for only one type of mutated protein, it can identify about 90% of mutated cases due to the predominance of the p.R132H mutation. Sequencing should be performed to evaluate the remaining cases if clinically warranted. Mutation-specific antibody for IDH1 R132H can be combined with the ATRX antibody, which will show loss of staining in mutated cells (Fig. 22.3).

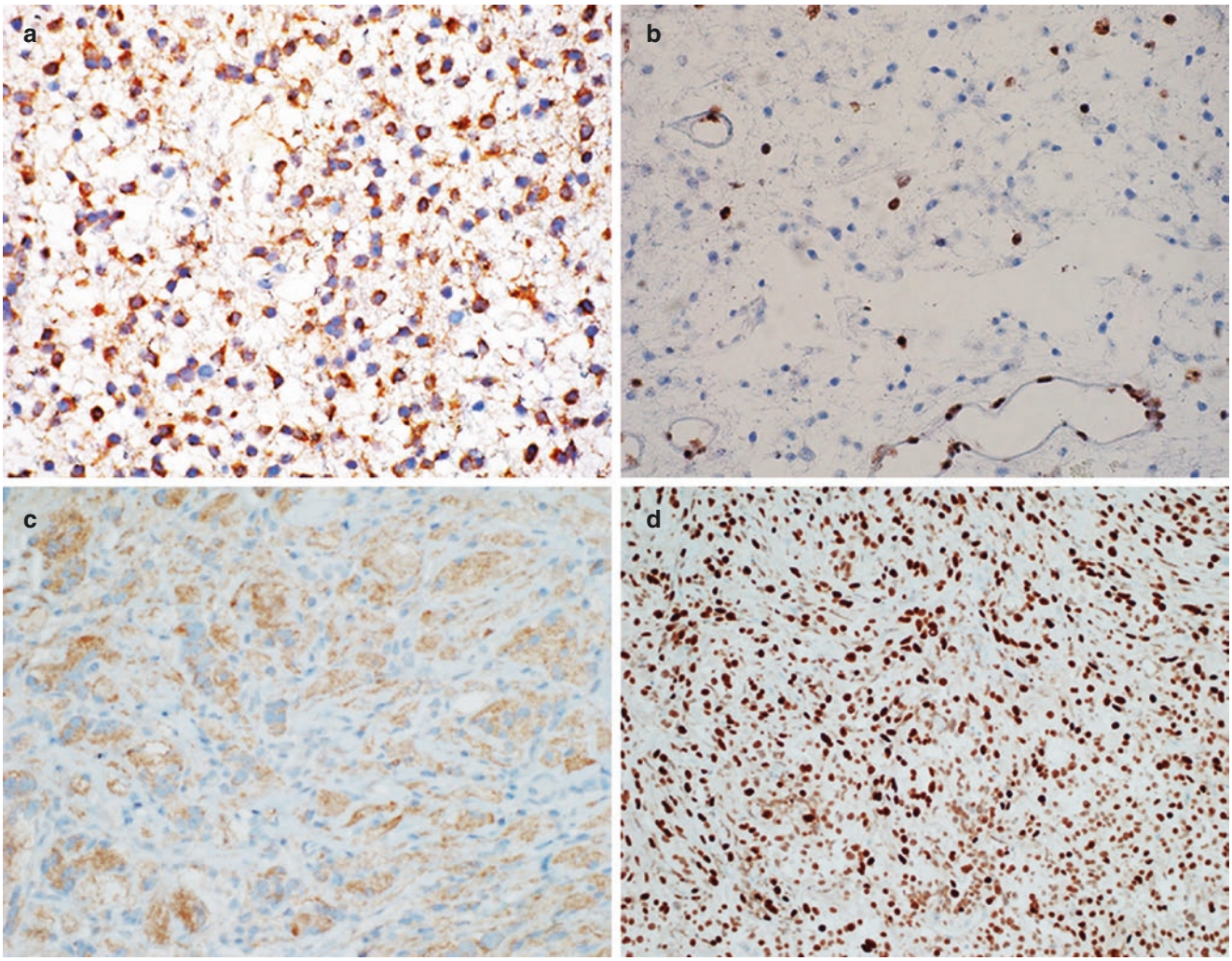


Fig. 22.3 Mutation-specific antibodies in diagnosis of brain tumors. Development of mutation-specific antibodies has revolutionized diagnostics of brain tumors. A combination of IDH1 R132H-positive tumor (**a**) with ATRX loss (**b**) is virtually diagnostic of a diffuse astrocytoma, regardless of the histological features even in an extremely small biopsy. Positivity for BRAF V600E by immunohistochemistry is helpful in diagnosis of ganglioglioma (**c**), pilocytic astrocytoma, and some GBM. Histone H3 K27 M mutation can be reliably identified in small samples using a mutation-specific antibody (**d**)

Amplification of *EGFR* and Other Receptor Tyrosine Kinases

Amplifications of receptor tyrosine kinase (RTK) genes play a crucial role in tumorigenesis of malignant glioma and are considered major drivers of tumor growth. Up to 50% of high-grade gliomas have amplification of a RTK gene, most commonly *EGFR*, *KIT*, *VEGFR2*, *PDGFRA*, and *MET* [18–21]. Compared to *TP53* mutation, *IDH1* mutation, and 1p/19q loss, the high-level amplification of RTK genes is a relatively late event in the tumorigenesis of glioma. Typically, only one RTK will have a high-level amplification in any given tumor. *EGFR* is the most commonly amplified RTK in adult GBMs (~40%), while in children *PDGFRA* amplification seems to be the most common with frequencies ranging from 5% to 12%. Up to 30% of diffuse intrinsic pontine gliomas in children have *PDGFRA* amplification. *EGFR* amplification is a hallmark of primary GBM and is more common in older patients, whereas secondary GBMs that develop from lower-grade gliomas are much less likely to have *EGFR* amplification. In addition to amplification, however, ~50% of all *EGFR* amplified GBMs include a truncated mutant variant, *EGFRvIII*, with constitutively upregulated tyrosine kinase activity. Not surprisingly, given the prevalence of *EGFR* alterations, significant but thus far unsuccessful efforts have been focused on development of therapies targeting *EGFR*. In addition, amplification of *EGFR* has unclear significance as a prognostic marker. Some

studies found no association with survival, others reported a negative impact, and some suggested a favorable impact on patient survival. Interestingly, it has been reported that the level of *EGFR* amplification influences response to therapy [22].

There are several reasons to perform *EGFR* testing in clinical practice [6]. From a diagnostic perspective, *EGFR* amplification seen in the context of a brain tumor is pathognomonic of GBM, and due to the high level of amplification, cells can be easily identified, even in samples with low cellularity. *EGFR*-amplified cells have been shown to be associated with the invasive edge of gliomas. *EGFR* amplification is also almost mutually exclusive with 1p/19q deletion and *IDH1/2* mutations and is therefore helpful in the differential diagnosis of anaplastic oligodendroglioma versus the small cell variant of GBM, because there is significant morphological overlap between these clinically very different glioma subtypes. *EGFR* amplification is encountered in most small cell GBMs, and tumors lack 1p/19q co-deletion, while anaplastic oligodendrogliomas often show 1p19q loss but never *EGFR* amplification [6]. Despite numerous studies, it remains primarily a diagnostic marker, as this alteration does not seem to provide independent prognostic information. Also, despite numerous clinical trials, it does not seem to predict response to EGFR inhibitors or antibodies. The role of other RTKs in clinical diagnostics is less clear. Although associated with distinct molecular subtypes of GBM, *PDGFRA* amplifications can be seen also in lower-grade gliomas, and *MET* amplifications are present in a small subset of GBMs and LGGs. For clinical practice, one can consider performing FISH or using a DNA array such as aCGH (Fig. 22.4) or methylation array. The advantage of

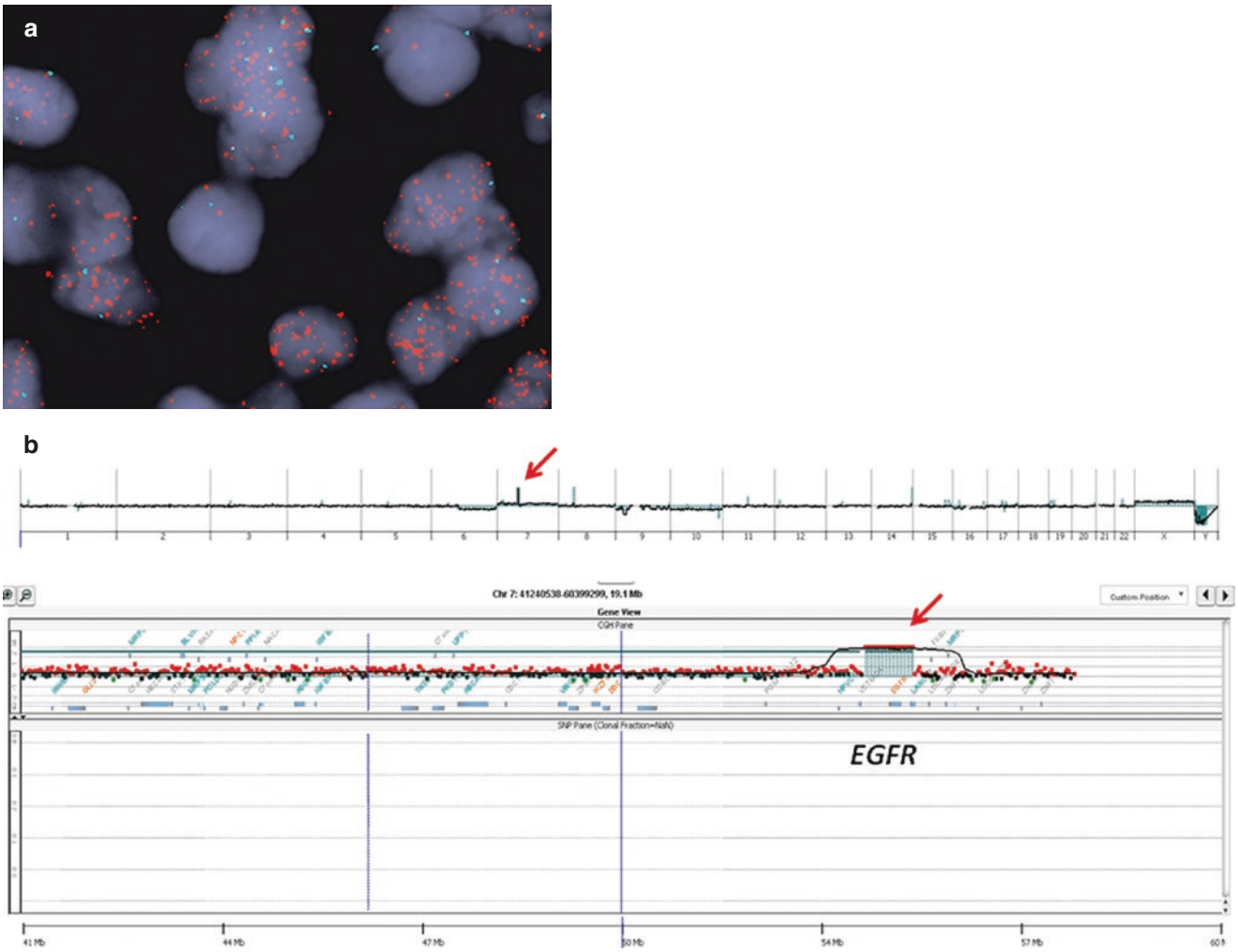


Fig. 22.4 *EGFR* amplification in GBM. Amplification of RTKs is one the most common driver mutations in tumorigenesis of GBM. *EGFR* is the most commonly amplified gene in adults. High-level amplification can be easily detected by FISH (Panel (**a**), *red*, *EGFR*; *blue*), control probe for chromosome 7, nuclei stained by DAPI) or by aCGH (**b**). Whereas the overall genome is relatively stable, chromosome 7 shows a distinct peak of focal amplification (*Panel* (**b**), *top*, *red arrow*). A gain of signal from probes covering *EGFR* gene is seen in the magnified view of the amplified region on chromosome 7 (*Panel* (**b**), *bottom*, *red arrow*)

array or NGS approach is the ability to evaluate multiple targets with one assay. However, whole-genome arrays might not be able to detect minor clones with amplifications and might be less successful with samples in which tumor cell density is low.

BRAF Mutation and Duplication Analysis

BRAF is a member of the serine/threonine protein kinases family. RAF kinases are part of the MAPK cascade, which regulates cell proliferation, differentiation, and apoptosis and therefore plays a wide variety of biological roles in a tumor cell. Activating mutations in *BRAF* are common in variety of solid tumors including papillary thyroid carcinoma, melanoma, adenocarcinoma of the colon, and pediatric LGGs. In addition to point mutations, pilocytic astrocytomas have a unique mechanism of *BRAF* activation via single-copy gain of *BRAF*, which results from a tandem repeat leading to fusion product *BRAF-KIAA1549* [6, 23]. This is a molecular hallmark of pilocytic astrocytoma and is identified in approximately 80% of pilocytic tumors in the cerebellum. Several other breakpoint variants have been described such as *KIAA1549* exon16–*BRAF* exon 9, *KIAA1549* exon 15–*BRAF* exon 9, *KIAA1549* exon19–*BRAF* exon 9, *KIAA1549* exon18–*BRAF* exon 10, and *KIAA1549* exon 16–*BRAF* exon 11. The resulting fusion proteins are highly tumorigenic. However, the constitutive activation of *BRAF* can also lead to senescence, particularly in slow-growing neoplasms.

Another mechanism of MAPK pathway activation is a tandem duplication at 3p25 leading to an in-frame fusion between *SRGAP3* and *RAF1*, which has high sequence homology with *BRAF*. The fusion gene retains the activation segment and kinase domain but lacks the inhibitory domain, resulting in constitutive upregulation. Constitutive activation of the MAPK pathway is seen in *NF1*- and non-*NF1*-associated pilocytic and pilomyxoid astrocytomas arising in the cerebellum. Duplication of the *BRAF* locus at 7q34 was identified in more than half of these tumors. In contrast, *BRAF* gene fusion is unusual in diffuse gliomas, which often contain the *BRAF* p.V600E point mutation, present in 25% of pediatric astrocytomas. Testing for *BRAF* should therefore include testing for both point mutations and gene fusions, based on the location of the tumor [23]. Patients with BRAF V600E LGG exhibit worse outcomes after chemotherapy and radiation therapies than BRAF wild-type LGG [24]. BRAF V600E is a potentially highly targetable mutation detected not only in LGG but also PXA and a subset of GBM, such as epithelioid GBM, and can be efficiently targeted by small molecule inhibitors. BRAF V600E mutations can be efficiently tested via immunostaining using a mutation-specific antibody (Fig. 22.3).

MGMT Promoter Methylation

The O6-methylguanine-DNA methyltransferase (*MGMT*) gene has been one of the most commonly tested molecular markers in neuropathology and neuro-oncology because of the clear association between *MGMT* promoter hypermethylation and an increased response to alkylating agents [25]. GBM patients with methylated *MGMT* promoter have a significant survival benefit with temozolomide treatment, and *MGMT* promoter hypermethylation is one of the strongest prognostic factors for patients with newly diagnosed GBM, including elderly patients [26, 27]. Patients with a hypermethylated *MGMT* promoter who are treated with concomitant and adjuvant temozolomide and radiotherapy had survival rates of ~50 and 15% at 2 and 5 years, respectively. However, 2- and 5-year survival rates in patients treated with radiotherapy alone were only ~25 and ~5%, respectively. In patients with GBM lacking *MGMT* promoter hypermethylation, 2- and 5-year survival rates were 15 and 8% when treated with combined radiochemotherapy, but 2 and 0% when treated with radiotherapy alone. Although the response to temozolomide is best in the methylation-positive group, one might argue that there appears to be some benefit in patients with nonmethylated tumors. In the pediatric GBM population, the data are less clear.

The frequency of *MGMT* promoter hypermethylation in the glioma literature varies widely, ranging from ~30 to 70% in GBM. This is due to technical aspects of the testing, but also to tumor heterogeneity, necrosis, and normal tissue contamination. Overall, *MGMT* promoter hypermethylation was observed in ~50–80% of anaplastic gliomas WHO grade III and 40–90% of the diffuse gliomas WHO grade II. The *MGMT* gene is located on 10q26 has a CpG-rich region of 763 bp with 98 CpG sites within the first exon. Promoter and enhancer regions are also located within the CpG island. CpG sites are not methylated in the normal tissue. In tumors, however, the cytosine in CpG sites can be methylated, which leads to altered chromatin structure and prevents binding of transcription factors. The result of this is silencing of the gene expression. *MGMT* is a DNA repair protein that in normal tissue catalyzes the transfer of a methyl group from the O6-position of a guanine DNA nucleotide to a cytosine residue. This is a one-way process, and alkylated MGMT is degraded. In tumors for which alkylating chemotherapy is used, such as temozolomide in malignant gliomas, this process leads to the binding of an alkyl group to the O6-position of guanine, which induces DNA mismatching and DNA double-strand breakage, resulting in apoptosis. A normally functioning MGMT protein neutralizes the lethal effects of alkylating agents by repairing DNA damage. When *MGMT* is silenced by hypermethylation of the promoter, however, reduced *MGMT* expression is thought to result in tumor cells not being able to repair DNA damage. This enhances the

cytotoxic effects of temozolomide. Interestingly, patients with a hypermethylated *MGMT* promoter exhibited a survival benefit even when treated with radiotherapy alone. Therefore, it is possible that MGMT also plays a role in radiotherapy-induced DNA damage repair. Another possibility is that *MGMT* methylation is an overall marker of genome methylation status in the tumor and that other DNA repair genes are silenced by promoter hypermethylation as well. Because other genetic alterations associated with a favorable prognosis, such as 1p/19q loss and *IDH1* mutation, often coexist with *MGMT* promoter hypermethylation, the contribution of each remains to be determined. Given the observed effect of temozolomide, even in the nonmethylated subgroup, and the lack of other options, it remains the first drug of choice regardless of *MGMT* status. *MGMT* methylation testing is frequently requested in clinical practice, but the impact of this testing on clinical management is unclear because the therapy remains similar, regardless of the result. *MGMT* testing, however, plays an important role in clinical trials to properly stratify patients. There are several assays that can be used for testing. The most frequently utilized are methylation-specific PCR and real-time methylation-specific PCR [28]. Other possibilities include methylation-specific pyrosequencing and methylation-specific MLPA. Contraintuitively, expression of the protein by IHC does not correlate well with the DNA results and cannot be recommended for clinical practice [28].

INI1 Loss

Atypical teratoid/rhabdoid tumor (AT/RT) is characterized by a combination of the presence of a primitive embryonal component and mesenchymal and epithelial components. Rhabdoid cells are not always detectable at the time of diagnosis, and the tumor can closely mimic medulloblastoma. The hallmark of AT/RT is a loss of chromosome arm 22q, which carries the S*MARCB1* gene, also known as *INI1* or *hSNF5*, at 22q11. Altered by deletion or mutation, loss of *INI1* is a defining molecular event in this tumor. Families with germ-line mutations of the *SMARCB1* gene have an inherited disposition to rhabdoid tumors everywhere in the body, including the brain AT/RT. Therefore, family members of children with these tumors should be tested for mutations to assess the potential risk. Presence of a reliable antibody for IHC of SMARCB1/INI1 is widely used in clinical diagnostics and has dramatically decreased the number of misdiagnosed tumors [29]. The current standard of care includes testing for INI1 by IHC in all medulloblastomas, primitive neuroectodermal tumors (PNETs), and choroid plexus carcinomas to avoid misdiagnosis. Patients with AT/RTs have an extremely poor outcome, although regimens using high-dose chemotherapy suggest potential benefit. In addition to AT/RT, mosaic loss of *INI1* was described in neurofibromatosis type 2 (NF2)-associated schwannoma and in schwannomatosis-associated schwannoma but is rarely seen in a sporadic schwannoma. This suggests a role for *INI1* in syndromes associated with multiple schwannomas. IHC is available and is the test method of choice. Sequencing can be performed in some cases or for genetic testing of family members, if a familial syndrome is suspected. Recently, three molecular subclasses of AT/RT have been identified, described in the separate section below.

PTEN

PTEN is a tumor suppressor and its loss is common in gliomagenesis. Inactivation of *PTEN* either by mutation or deletion is a frequent feature in many high-grade gliomas and leads to upregulation of the AKT pathway. Up to 80% of all GBMs show a loss of 10q23 containing *PTEN*, and up to 40% of primary GBMs will carry *PTEN* mutations. In the small cell variant of GBM, 10q is almost always lost, and this, together with *EGFR* amplification and 1p/19q preservation, comprises a useful molecular panel to diagnose GBM and distinguish it from an anaplastic oligodendroglioma. *PTEN* loss is present in both primary and secondary GBMs and is associated with shorter survival in the pediatric population, making it an interesting diagnostic as well as prognostic marker. It does not, however, seem to be a prognostic marker in adult GBM. *PTEN* loss is most often hemizygous, and testing can be easily performed using FISH [6] or by DNA arrays or next-generation sequencing.

CDKN2A

CDKN2A is located on 9p21 and encodes the p16 protein, which is a key inhibitor of the cell cycle in Rb pathway signaling. This pathway is one of the most commonly affected pathways in cancers, including gliomas. In gliomas, oligodendroglioma, and astrocytoma, *CDKN2A* is usually lost via homozygous deletion and associated with high-grade tumors and decreased survival. Interestingly, a small subset of pilocytic astrocytoma also demonstrates loss of *CDKN2A*. Large studies have shown that by multivariable clinical and molecular stratification, the CDKN2A deletion contributed independently to poor outcome in BRAF V600E mutant low-grade gliomas. Similar to *PTEN*, testing can be performed using FISH [6] or by DNA arrays or next-generation sequencing.

RELA Fusion and Ependymoma

Whole-genome profiling efforts have remarkably changed our understanding of ependymomas. While, histologically, ependymomas are relatively uniform and histological features carry low prognostic value, molecularly, ependymomas can be separated into distinct entities. One of the newly described subtypes of supratentorial ependymoma is characterized by fusion of RELA gene and is believed to be associated with poor outcome [30]. RELA fusion ependymomas represent majority of supratentorial ependymomas in children and adults. Interestingly RELA fusion is not a driver of ependymomas arising in the posterior fossa or the spinal cord. *RELA-C11orf95* is the most common variant and arises as a result of chromotrypsis and genomic reassembly. The RELA fusion leads to constitutive activation of NF-kappaB pathway. The presence can be detected by several molecular methods, such as break-apart FISH, methylation, and RNAseq. Immunohistochemically, the expression of L1CAM correlates strongly with the presence of RELA fusion.

Histone H3 K27M and G34V/R Mutations in Gliomas

Mutation landscape of gliomas of children has long remained a mystery. Several sequencing studies have identified recurrent mutations in the histone-coding genes *H3F3A*, *HIST1H3B*, and *HIST1H3C*, with *H3F3A* being the most commonly mutated. These mutations occur in position K27 and result in K27M amino acid change. The mutation leads to diffuse loss of histone H3K27 trimethyl mark (H3K27me3), which is believed to decrease the PRC2 activity although the exact mechanism of action is unknown. Since biopsies are small, the diagnosis of H3 K27M mutation has been significantly improved by introduction of mutation-specific antibody (Fig. 21.3). The midline H3K27M-mutated tumors usually carry poor prognosis and are resistant to chemotherapy and radiation. However, while most of the tumors are aggressive and WHO recommends grading all H3K27M tumors as WHO grade IV, several studies have reported tumors with prolonged growth and less aggressive behavior [31]. These tumors often have concurrent mutations in *BRAF* and *FGFR1*. Therefore, additional analysis using the BRAF V600E mutation-specific antibody is paramount, and one should be cautious uttering the grade IV solely based on the presence of the H3K27M positivity particularly in a small biopsy and in complete absence of other imaging, clinical, or histological features of GBM. In addition to midline gliomas characterized by K27M mutation, G34R/V mutations in the tail of histone H3.3 have been identified in pediatric malignant gliomas. In addition, these tumors are characterized by loss of function of *ATRX* or *DAXX* and associated with alternative lengthening of telomeres phenotype (ALT). K27 and G34 mutations lead to distinctly different tumors. Using DNA methylation profiling tumors form distinct epigenetically defined clusters. Tumors with K27 M mutation arise almost exclusively in the midline, from spinal cord to hypothalamus, while tumors with G34R/V mutations are limited to gliomas of the cerebral hemispheres. Patients with K27M-mutated tumors are also significantly younger (median age 10) compared to patients with G34-mutated gliomas (median age 18). While some reports suggested that K27M-mutated tumors are more aggressive, this may be due to their midline location and inability to achieve a significant resection. However, both subtypes are radio- and chemotherapy resistant, and there are currently no effective therapies [32].

Whole-Genome Molecular Classification of Brain Tumors

Medulloblastoma

Medulloblastoma is the most common malignant brain tumors of childhood. Medulloblastomas, by definition, arise in the posterior fossa, while similarly looking tumors are called PNETs elsewhere in the brain. The WHO classification recognizes several subtypes of medulloblastoma based on morphology: classic, desmoplastic/nodular, medulloblastoma with extensive nodularity, and large cell/anaplastic medulloblastoma. These subtypes can be associated with age (such as medulloblastoma with extensive nodularity in infants) or with better versus poor prognosis (such as desmoplastic medulloblastoma and large cell/anaplastic medulloblastoma, respectively). Although several of the morphologic subtypes are no longer regarded as separate entities by the WHO, molecular classification is playing an increasingly important role in the classification of this disease.

Medulloblastoma is a prototypic brain tumor in which molecular tools have provided better understanding of the disease biology by classifying a relatively uniform appearing neoplasm into distinct biological entities and identifying potential therapeutic targets. Association of medulloblastoma with rare entities such as Turcot and Gorlin syndromes not only suggested the role of Wnt and Hedgehog signaling but also suggested that medulloblastomas can have a relatively simple “one pathway” oncogenesis. The real extent of the molecular diversity in medulloblastoma was first revealed using expression profiling. Expression profile studies first identified medulloblastoma, PNET, and AT/RT as distinct molecular entities, and later studies divided

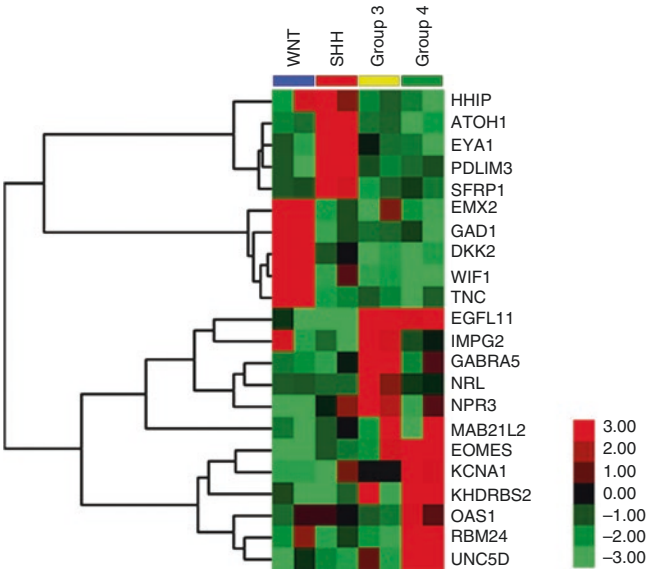


Fig. 22.5 Expression profiling of medulloblastoma. Whole-genome expression profiling pioneered the molecular classification of medulloblastoma. Focused panels of selected genes specific for each subgroup can be created to classify tumors [37]. Each tumor profile is represented by a column and each gene expression level by a line. Color coding represents the level of increased (*red*) or decreased (*green*) expression compared to control genes. Two tumors of each subgroup are shown to cluster together. Shh and Wnt tumors are very distinct, but there is some overlap between Group 3 and 4 expression, even with this selected group of genes. Nevertheless, tumors can still be easily categorized. Immunohistochemistry (IHC) for SFRP1, DKK2, NPR3, and KCNA1 has been proposed as a practical panel for classification in clinical practice. (Figure courtesy of Joanna Triscott and Dr. Sandra E. Dunn, University of British Columbia)

medulloblastomas into four to six distinct subgroups, depending on the study. There are several reports of distinct molecular pathways and distinct groups of medulloblastoma [33–36]. Whereas all groups identified the Sonic hedgehog (Shh) and Wnt subgroups as relatively distinct, separation of non-Shh and non-Wnt tumors varied based on classifications schemes. The current consensus molecular classification distinguishes four types: Wnt, Shh, Group 3, and Group 4 (Fig. 22.5) [38, 39].

Sequencing of the genes involved in the oncogenic pathways identified Wnt activation in sporadic medulloblastomas via mutations in *CTNNB1*, *AXIN*, and *APC* and mutations activating the Shh pathway including *PTCH1*, *SUFU*, and *SMO*. Whole-genome sequencing studies have revealed novel genes that are mutated in medulloblastoma including *MLL2* in Shh and Wnt groups, *MLL3* in Group 3 and 4, *SMARCA4* in Wnt and Group 3, *DDX3X* in Wnt, *LDB1* and *BCOR* in Shh, and many others. Among the most frequently mutated genes in medulloblastoma are, not surprisingly, *CTNNB1* and *PTCH1*, but also *DDX3X*, *MLL2*, *SMARCA4*, and *KDM6A*. In addition to mutations, chromosomal and smaller copy number changes were identified early on. Isochromosome 17q, formed as a result of loss of chromosome 17p and gain of 17q, was found in approximately 30% of medulloblastomas, and large chromosomal or small copy number changes have been associated with certain subtypes. Examples are loss of chromosome 6 in the Wnt group, loss of long arm of chromosome 9 in Shh group, and amplifications of *MYC* or *MYCN* loci associated with large cell/anaplastic medulloblastomas subtypes and poor outcome. Later, gains and amplifications of *OTX2* and *PVT2* were revealed in pathogenesis of Group 3, *SNCAIP* and *CDK6* in Group 4, and *GLI2* in Shh medulloblastoma. The amount of genomic data is significantly larger than the amount of functional data that would confirm the role of many of these changes; however, the consensus about the four main groups of medulloblastoma remains a practical framework for further studies.

Wnt Group

The importance of recognizing the Wnt group is the very good long-term prognosis in comparison to other groups. Whereas medulloblastoma is more common in males, in Wnt medulloblastomas the sex ratio is approximately 1:1. Wnt medulloblastomas occur at all ages, but are rare in infants and are usually not disseminated at the time of diagnosis. Germ-line mutations of the *APC* gene in the Wnt pathway are associated with Turcot syndrome. Unfortunately, this subtype is also the least common and represents only ~10% of all medulloblastomas. With long-term survival rates exceeding 90%, many patients suffer from the long-term therapy-associated complications including cognitive decline, endocrine insufficiencies, growth problems, and secondary neoplasms rather than from medulloblastoma recurrence. While in other subtypes molecular studies might identify new therapeutic targets, in Wnt patients, the goal instead might be to first modify unnecessary toxic treatments to decrease later morbidities and mortality associated with the treatment. The Wnt medulloblastoma genome is relatively stable, with only a few changes other than monosomy 6. Having said that, tumors with a clear Wnt expression profile and without monosomy 6 also have been described. Furthermore, overexpression of genes in the Wnt pathway has been detected in Shh and Group 3 medulloblastomas, as well. Although most Wnt tumors show classic morphology, a Wnt transcriptional signature is associated with excellent prognosis even in tumors with anaplastic/large cell features. Such tumors, however, are rare in the Wnt subgroup. Several methods have been proposed to diagnose Wnt medulloblastoma. It remains to be validated whether IHC for the DKK1 or CTNNB1 proteins, cytogenetic testing for monosomy 6, or a transcriptional signature should be used for clinical testing.

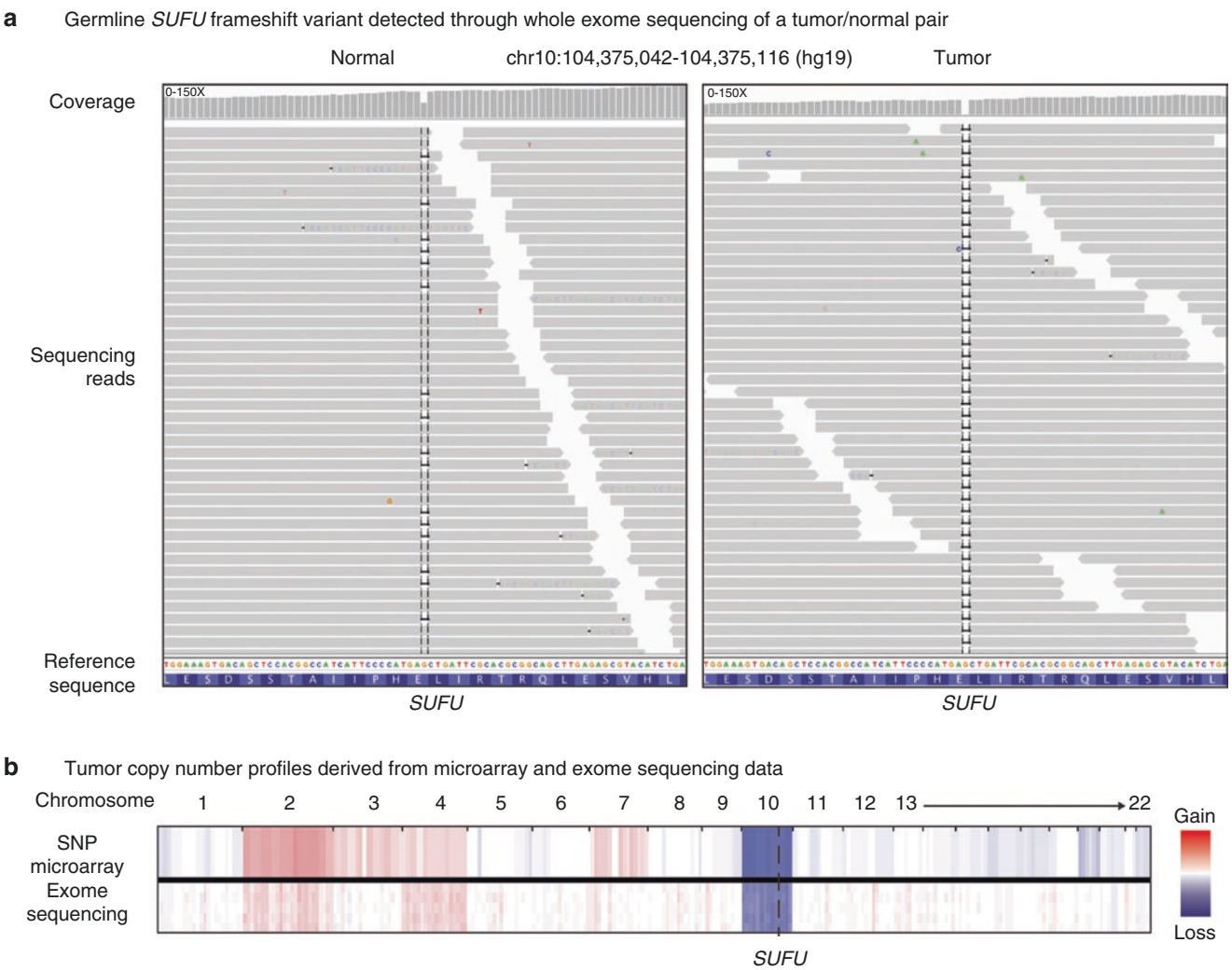


Fig. 22.6 Complex genomic features in medulloblastomas with germline mutation and somatic loss. (**a**) A patient with a germ-line heterozygous loss-of-function *SUFU* mutation (in the specimen labeled normal, shown as *empty spots* in some reads) developed a tumor via a somatic loss of chromosome 10, which resulted in complete loss of function of *SUFU* (completely empty column in all reads of the tumor specimen). The resulting medulloblastoma is therefore classified to be of Shh subtype. (**b**) The overlay of SNP microarray and exome sequencing data reveals the copy number profile of the tumor with loss of chromosome 10. Of note, both a SNP array and exome sequencing can provide copy number data in addition to information about point mutations and indels. (Figure courtesy of Dr. Trevor J. Pugh, Broad Institute of MIT and Harvard)

Shh Group

The Shh medulloblastomas are characterized by aberrations involving the Shh signaling pathway, which drive tumor initiation and progression. Germ-line mutations in the Shh receptor *PTCH1* lead to Gorlin syndrome, which includes in its phenotype a predisposition to medulloblastoma. Infantile medulloblastoma is associated with germ-line mutations of the Shh inhibitor *SUFU* (Fig. 22.6). Shh medulloblastomas have a bimodal age distribution with high frequency in infants 0–3 years old and in adults. There is no gender predominance in Shh medulloblastomas. Nodular/desmoplastic medulloblastomas almost invariably belong to the Shh subgroup. However, up to 50% of Shh subgroup medulloblastomas are not nodular/desmoplastic, and therefore histology alone is not sufficient to identify them. Overall, the prognosis of Shh medulloblastoma is similar to Group 4. Patients have a better prognosis than those with Group 3 medulloblastoma, but worse than those with Wnt tumors. Somatic mutations along the entire pathway have been identified, including in *PTCH1*, *SMO*, and *SUFU*, as well as amplifications of their downstream transcription factors *GLI1* and *GLI2*. Although Shh medulloblastomas have been successfully identified on the basis of transcriptional profiling, some have proposed a combination of immunohistochemical stains such as those for the SFRP1 or GAB1

proteins to be used in clinical practice. On the DNA level, loss of chromosome 9q, where *PTCH1* is located (9q22), is exclusive for the Shh group. These tumors can potentially be targeted by small molecule inhibitors against *SMO* [40]. Unfortunately, the effect is short-lived, and tumors rapidly develop resistance mutations [41]. Furthermore, Shh medulloblastomas that carry aberrations at the more downstream parts of the pathway, such as amplifications of *GLI1*, *GLI2*, and *MYCN* genes would be inherently resistant to such inhibitors.

Group 3

Overall, Group 3 and 4 medulloblastomas are characterized by overlapping features and are less distinct than Wnt and Shh medulloblastomas. Also, both groups exhibit a higher number of genetic changes, particularly more complex DNA rearrangements. Group 3 tumors occur more commonly in males than females, and are found in infants and children, but are almost never observed in adults. They have a high incidence of large cell/anaplastic histology although many of them are classic medulloblastomas. In addition, they are very frequently metastatic at the time of diagnosis. The Group 3 transcriptional profile is photoreceptor/GABAergic. A true hallmark of Group 3 tumors, however, is *MYC* overexpression/amplification, to the point that some have proposed to rename them *MYC* Group, instead. Immunohistochemical positivity for NPR3 has been suggested as a Group 3 marker and has been associated with aggressive disease. Group 3 tumors often show gains of chromosomes 1q, 7, and 17q and/or loss of chromosomes 11, 17p, 5q, 10q, and 16q. Gains of 18q are seen in both Groups 3 and 4. Amplification and overexpression of the medulloblastoma oncogene *OTX2* appears to be restricted to Group 3 and Group 4 tumors. Group 3 can be further stratified into Group 3a, which includes all medulloblastomas with *MYC* amplifications and contains most of the high-risk patients. Group 3b patients do not have *MYC* amplifications, and their clinical outcome is similar to Group 4 patients. Because Group 3 patients have the worst prognosis, they have, arguably, the highest need for the most aggressive therapy and for novel targeted therapies.

Group 4

Group 4 medulloblastomas have classic histology and are more frequent in males. Presence of the isochromosome 17q is a hallmark of Group 4, although it can be also seen in a minority of Group 3 medulloblastomas. Furthermore, isolated 17p deletion is seen in both Groups 3 and Group 4, but not in Wnt or Shh medulloblastomas. Recently it has been found that 17p harbors *CTDNEP1*, which is a novel candidate gene in the pathogenesis of medulloblastoma. The other interesting cytogenetic change is common loss of the X chromosome in females with Group 4 medulloblastoma, which is seen in ~80% of tumors. This is particularly interesting considering that the male–female ratio in Group 4 medulloblastoma is almost 3:1 and therefore suggests that the X chromosome contains potential tumor suppressor genes such as *KDM6A*, as discussed below. *MYCN* and *CDK6* are often amplified in Group 4 medulloblastomas, while they are usually not affected in Group 3 tumors. Tumors are frequently metastatic at presentation, and the prognosis is intermediate, similar to non-infant Shh medulloblastomas. Compared to other groups, their pathogenesis is the least understood, and there is no consensus about the driver mutation.

Group 4 medulloblastomas are reliably identifiable via their transcriptional profile and have a neuronal differentiation/glutamatergic profile. KCNA1 has been suggested as an immunohistochemical marker for this group. However, recent whole-genome studies identified several mutated genes in medulloblastoma that normally participate in histone modifications. *KDM6A* is an interesting example because it is altered by a nonsense mutation and can be considered a novel tumor suppressor gene. *KDM6A*, *BCOR*, *DDX3X*, and other genes mutated in medulloblastoma are located on Xp, which is commonly lost in females with this type of medulloblastoma. In addition to *KDM6A*, other chromatin remodeling genes such as *ZMYM3* and *CHD7* can be mutated in Group 4, suggesting that this group might be defined by mutations in genes responsible for epigenetic modifications [42].

Molecular methods are playing an increasing role in the diagnosis and stratification of medulloblastoma. Whereas the overall biology indicated above is suggestive of a high level of complexity, the actual list of prognostic molecular factors associated with poor or improved survival in medulloblastoma patients is surprisingly short. Markers of good prognosis include Wnt subtype, increased TrkC mRNA expression, and nuclear staining of beta-catenin or positivity for *DKK1* indicating Wnt pathway activation. Markers indicating worse outcomes include *MYC* amplification and/or overexpression, 17p loss and i17q formation, as well as strong p53 immunoreactivity suggestive of underlying *TP53* mutation. Introducing molecular testing for subclassification of medulloblastomas has been challenging due to the numerous methods. Although a combination of immunohistochemical stains has been proposed, standardization across clinical laboratories has been challenging. IHC could be combined with a few FISH targets, most importantly *MYC*, but possibly also *MYCN* and 9q22 for *PTCH1* to increase the accuracy. A

relatively simple pathology stratification for clinical practice can combine a patient's age, medulloblastoma morphology, and a combination of a few immunohistochemical stains with FISH to identify Wnt and Shh pathway status, and *MYC* anomalies. Another possibility is to combine IHC with aCGH or to replace IHC altogether and perform a targeted expression profiling assay using a subset of genes. Alternatively, a focused expression profile array to distinguish among medulloblastoma subclasses can be performed (Fig. 22.5). Medulloblastoma subgroups were originally defined based on gene expression profiling using fresh-frozen samples [35]. However, in most cases only formalin-fixed paraffin embedded (FFPE) material is available, and RNA-based analysis of FFPE tissue for classification is inferior to data obtained from the frozen tissue. This is particularly prominent in older FFPE samples with significant RNA degradation [43]. The use of DNA-based assays for subgrouping has significant advantages due to the higher stability of DNA compared to RNA. DNA methylation profiling has recently been applied for the subgrouping of large series of, for example, glioblastoma and chronic lymphocytic leukemia [44, 45]. Numerous publications have shown that medulloblastomas can be reliably subclassified based on their DNA methylation profile [46] (Fig. 22.7).

Gliomas

Glioblastoma

Glioblastoma (GBM) is the most common malignant brain tumor of adults. It typically develops from astrocytes and mostly arises de novo without a previous low-grade precursor (primary GBM, pGBM). Approximately 10% of GBM arise from a preexisting LGG and are termed secondary GBM (sGBM). Survival of patients with sGBM is longer than that of patients with primary GBM. These two types arise along different molecular pathways and have different expression profiles.

A variety of studies have attempted to identify individual genes as well as signaling pathways by combining expression profiling and structural DNA data to identify prognostic and possibly predictive markers. The most commonly affected genes and pathways in GBM include *EGFR* and other receptor tyrosine kinases, the *PI3K/PTEN/AKT* pathway, and the *TP53/MDM2/p14* pathway. The most common focal DNA changes are amplification of *EGFR*; amplification of 4q12 which contains *PDGFRA*, *KIT*, and *VEGFR2*; and deletion of the *CDKN2A* gene. The large chromosomal variants include loss of 10q, 19q, 22q, and 1p. Pediatric

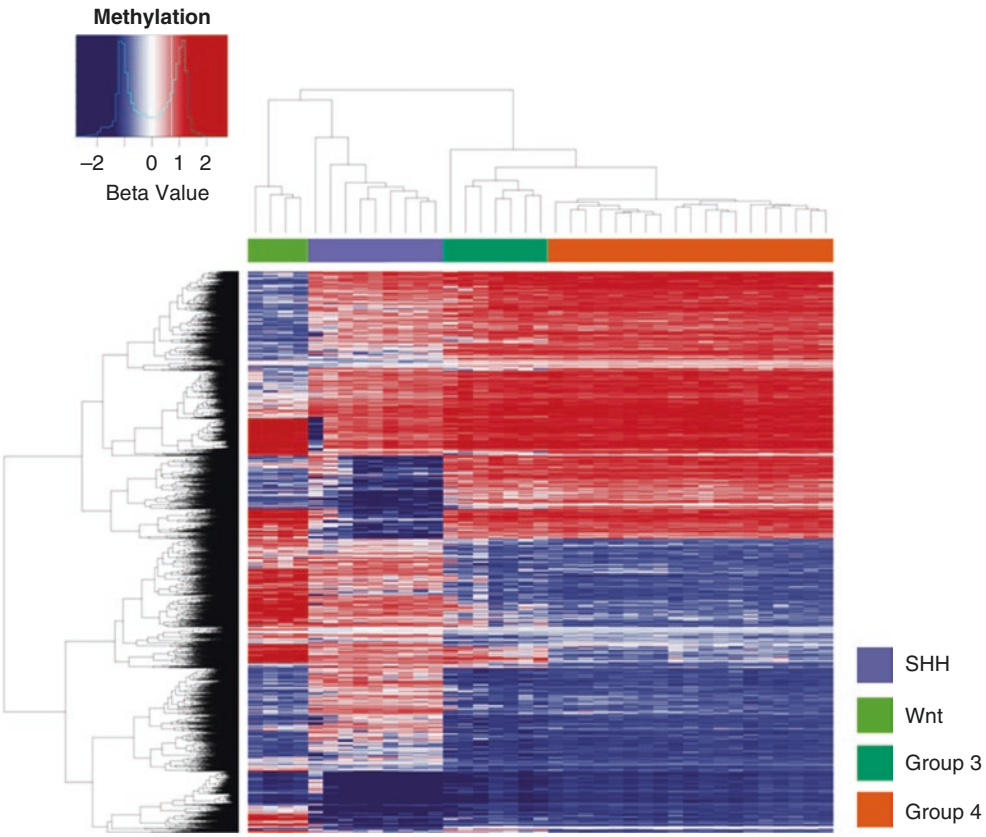


Fig. 22.7 Medulloblastoma subclassification using DNA methylation signatures. Despite similar histological features, DNA methylation-based signatures can reliably stratify medulloblastomas in molecular distinct subgroups. Four major subgrouos sonic hedgehog (SHH), WNT, Group 3, and Group 4

GBMs also commonly show microsatellite instability due to DNA mismatch repair defects, which is uncommon in adult tumors, and have a different spectrum of copy number changes such as gain of 1q, 3q, and 16p as well as loss of 8q and 17p. Based on expression profiling, gliomas were classified into three main groups: proneural, mesenchymal, and proliferative, based on analysis of gene ontology. This classification demonstrated prognostic value and has been confirmed by several studies, including The Cancer Genome Atlas (TCGA) project. Discovery of novel mutations, such as those in the *IDH1*gene, in high-grade gliomas and a combination of expression profiling classification studies and DNA alterations led to subsequent identification of additional subtypes of GBM and a more recent classification into proneural, neural, mesenchymal, and classical subtypes [47]. Each subtype is defined by a combination of expression parameters as well as DNA aberrations such as *EGFR* amplification, *NF1* loss, and *PDGFRA*/*IDH1* alterations.

Similarly to medulloblastomas, the original subclassification of gliomas has been performed using RNA expression profiling relying heavily on fresh frozen tissues. However, superiority of DNA in regards to the quality and stability has enabled DNA methylation based approaches to classify GBM based on their epigenetic signatures [45]. It has been further shown that methylation-based diagnosis using methylation signatures as a surrogate for diagnosis stratifies gliomas more accurately than histological diagnoses [37]. Based on epigenetic signatures, alternative grouping was proposed. This subclassified the proneural subtypes into IDH, K27M, and RTKI subgroup. Classical GBM corresponds to RTKII subgroup in this classification scheme, while G34-mutated GBM have mixed gene expression profile and expression and epigenetic signatures of mesenchymal subtype did not differ.

While a consensus which classification scheme is most optimal for clinical management is still pending, it is clear that GBM is a group of molecularly distinct diseases rather than a single entity, despite similar histological features.

GBM Classical/RTKII

High-level *EGFR* amplification, often accompanied by *EGFRvIII* mutation and paired with *EGFR* overexpression, is a hallmark molecular change in classical GBMs. Another typical finding is loss of chromosome 10, containing *PTEN*. Classical GBMs also lack mutations in *TP53*, which is one of the most commonly mutated genes in GBM. Homozygous deletion of the 9p21.3 locus containing the *CDKN2A* gene that encodes p16INK4A and p14ARF is another frequent event in classical GBM. Loss of *CDKN2A* is mutually exclusive with loss of other RB pathway genes, such as *RB1*, *CDK4*, or *CCDN2*, suggesting that the RB pathway is almost exclusively affected through homozygous deletion of *CDKN2A*. Expression profiling identified Notch and Shh signaling pathways to be overexpressed in the classical group.

GBM Mesenchymal

The molecular hallmark of the mesenchymal subgroup is a heterozygous deletion of the *NF1*-containing region 17q11.2. The majority of these tumors have decreased *NF1* expression. In addition to the heterozygous deletion, *NF1* mutations are also common in this subgroup. Expression profiling indicated upregulation of mesenchymal markers, including *YKL40* and *MET*, as well as genes in the tumor necrosis factor super family pathway and in the NF-kB pathway, both possibly due to high levels of necrosis and inflammation in these tumors.

GBM Proneural: Including IDH1/2, K27, and RTKI Subtypes

GBMs that classify as proneural by gene expression studies include tumors driven by *IDH1/2* mutations (secondary GBM); aberrations of *PDGFRA*, either by amplification or point mutations; and K27 M-mutated tumors, which are discussed in more detail in a separate section. Similar to *EGFR* amplifications in the classical group, focal amplifications of the locus at 4q12 that harbors *PDGFRA*, *VEGFR2*, and *KIT* are present in all subtypes of GBM but are observed much more frequently in the proneural group. GBM with concurrent *PDGFRA* amplification accompanied by high levels of *PDGFRA* expression is a molecular hallmark of proneural GBMs. Another common genetic event in this group is loss of *TP53* function. As evidenced by expression profiling, overexpression of oligodendrocytic developmental genes, such as *SOX*, *DCX*, *DLL3*, *ASCL1*, *NKX2-2*, and *OLIG2*, and decreased *CDKN1A* expression can be present. High expression of *OLIG2* was previously shown to downregulate the tumor suppressor *CDKN1A*, leading to increased proliferation. *PIK3CA*/*PIK3R1* were also identified in proneural tumors without *PDGFRA* abnormalities.

IDH1/2- and K27M-mutated tumors represent distinct subsets of proneural GBMs which lack *PDGFRA* aberrations and in fact represent distinct entities that should not be classified together with *PDGFRA*-driven tumors (Fig. 22.8).

Neural GBM

The originally described neural subtype seems to have the least identifiable molecular features including aberrations

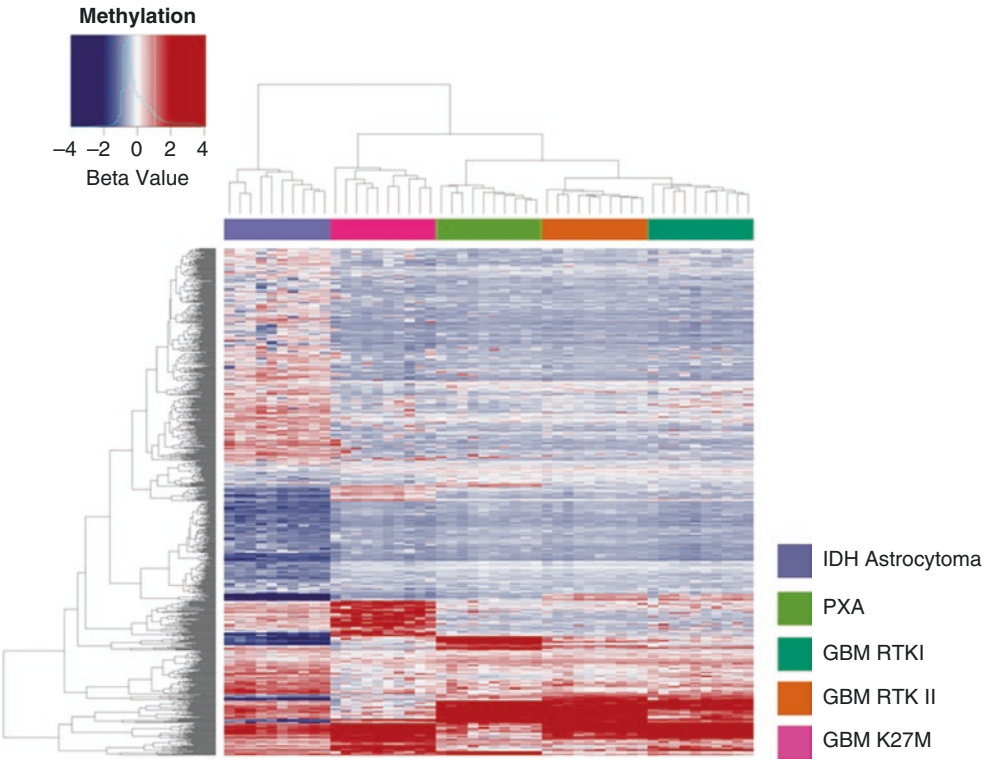


Fig. 22.8 Glioma subclassification using DNA methylation signatures. Gliomas can share histological features; however, they differ markedly in clinical behavior and outcome. Epigenetic signatures using DNA methylation show distinct methylation-based subgroups separating tumors driven by IDH1/2 mutations (IDH astrocytoma), GBM with histone H3 K27M mutation (GBM K27M), pleomorphic xanthoastrocytomas (PXA) which are driven by BRAF V600E mutations, and RTKI and RTKII subtypes of GBM, which correspond to proneural and classical subtypes by RNA expression profiling

of *PTEN*, *TP53*, *EGFR*, *NF1*, and *ERBB2*, as well as homozygous deletions of *CDKN2A*. None of these, however, dominate. Expression profiling showed upregulation of neuronal markers, including *SLC12A5*, *NEFL*, *GABRA1*, and *SYT1*.

Due to the lack of effective targeted therapy against either group, it is unclear whether classification has added prognostic or predictive value to currently performed clinical tests. There also seems to be an association between glioma grade and molecular subtype. While GBMs are composed of a mix of subtypes, grade II and grade III diffuse gliomas are almost exclusively proneural *IDH1/2*- or *PDGFRA*-driven tumors. Furthermore, sGBM are usually proneural subtype. These tumors are diagnosed at a younger age and have a high rate of *IDH1* and *TP53* mutations and lower rates of *EGFR* amplification and chromosome 10 loss. Pediatric GBMs are characterized by two distinct mutations of histone H3.3, each defining an epigenetic subgroup of GBM with a distinct global methylation pattern. These mutations were also mutually exclusive with *IDH1* mutations, which characterizes a third mutation-defined subgroup of pediatric GBM [45].

Low-Grade and K27 Wild-Type Diffuse Gliomas of Childhood

LGGs are the most common brain tumors of childhood and in children seem to display different aberrations than the LGGs that are precursors of GBM in adults [48]. Whole-genome studies have shown that LGG have relatively stable genomes, with loss of *CDKN2A* that is common in adult and pediatric high-grade gliomas. The most common focal gain is amplification of *PDGFRA*, followed by *MET*, *IGF1R*, *ERBB4*, and *EGFR*, in contrast with adult GBMs in which *EGFR* amplification is the most common focal gain. By FISH it was also observed that some tumors contained mutually exclusive subclones, with amplifications of *PDGFRA* or *MET*. Whole-genome sequencing identified recurrent mutations in *BRAF*, *RAF1*, and *ATRX*, rearrangements of *MYB* or *MYBL1*, and mutations and duplications of *FGFR1*, all of which seem to be mutually exclusive on the cellular level [48–51]. The number of non-silent mutations and rearrangements is very low with the median number of one mutation per tumor, which suggests that very few alterations are necessary for tumorigenesis. Furthermore, *FGFR2* and *FGFR3* fusions have been identified as a common driver in low-grade neuroepithelial tumors [52].

As discussed in the section about *BRAF*, mutations and duplications are characteristic for low-grade pilocytic astrocytomas and pleomorphic xanthoastrocytomas, and as discussed in the section about H3K27M mutations, histone K27 mutation is typical for diffuse midline gliomas. BRAF- and K27M-driven tumors can be distinguished by combination of sequencing, mutation-specific antibodies, or DNA methylation-based epigenetic signatures (Figs. 22.3 and 22.8).

Genetic Mosaicism and Intratumoral Heterogeneity in Gliomas

Classification studies divided high-grade gliomas into distinct subgroups. However, several studies identified a genetic mosaicism of receptor tyrosine kinase (RTK) amplifications in GBMs, which leads to the question of how well defined these groups in fact are. RTKs are commonly amplified in GBMs, the most common being *EGFR* amplification, which is present in ~40% of cases, and *PDGFRA* amplification in 10–15% of cases. *PDGFRA* amplification is often accompanied by amplifications of *VEGFR2* and *KIT*, which reside in the same region. The third most common is amplification of *MET*. Each of these RTKs is associated with a particular molecular subtype of GBM as described above, *EGFR* with classic, *PDGFRA* with proneural, and *MET* with the mesenchymal subtype.

Studies using FISH illustrated that some GBMs contain a mix of up to three intermingled subpopulations of GBM cells with mutually exclusive amplifications of *EGFR*, *PDGFRA*, and *MET*, which arose from the same precursor (Fig. 22.9) [53–55]. GBM cells in which *EGFR* and *PDGFRA* amplification was present within the same tumor cell have also been observed [53, 54]. Different clones tend to inhabit different microenvironments and might play a different role in the growth and progression of GBM [54]. Mosaic amplification has also been described in LGGs [56] and pediatric gliomas [49], further hampering the idea that gliomas reliably can be stratified into distinct subtypes. The authors of one of the studies, which quantified subclones with different amplifications, reported that ratios of subclones are highly variable, and minor subclones might not be picked up by whole-genome approaches. One can assume that the heterogeneity observed on the level of RTK can be also present on the level of point mutations, which would make the level of intratumoral heterogeneity even higher [57]. Finally, the presence of different subclones within the tumor raises several issues with regard to sampling of the brain tumors and testing at the time of diagnosis versus later, as well as issues pertaining to the development of therapies that are to be specific for certain molecular subtypes.

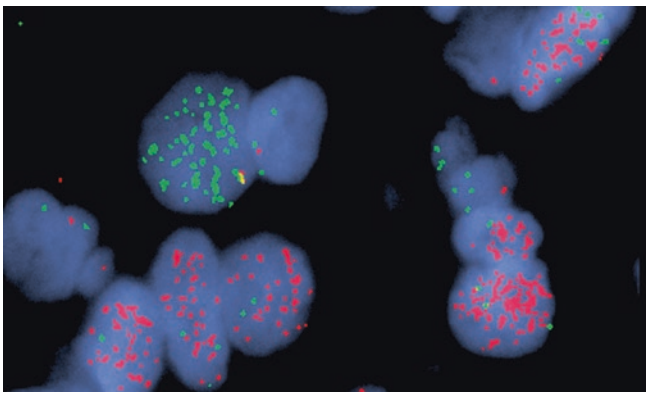


Fig. 22.9 Mosaic amplification of RTKs in GBM. Intermingled subpopulations of GBM cells carry high-level amplification of *EGFR* (*red*) or *PDGFRA* (*green*). Amplification of an RTK is a strong driver mutation, and *EGFR* and *PDGFRA* amplification are considered mutually exclusive on the cellular level. Furthermore, *EGFR* amplification is strongly associated with the classical subtype and amplification of *PDGFRA* with the proneural subtype of GBM, two supposedly distinctly different molecular categories of GBM. In this tumor, however, these two populations arose from the same precursor, sharing the same early mutation events such as *TP53* mutation and homozygous deletion of *CDKN2A*. Subclones later developed different, usually mutually exclusive RTK amplifications. This type of heterogeneity can be identified by in situ methods, whereas a whole-genome approach would likely not be able to distinguish whether *EGFR* and *PDGFRA* amplifications occur in the same cell or in different subclones. Furthermore, it might not identify minor subclones

Meningioma

Meningiomas represent approximately 30% of primary CNS tumors of adults. Although most are benign, these tumors tend to recur and require multiple resections. Hence, they are associated with significant morbidity and mortality. Currently, the prognostic criteria for tumor behavior are largely based on histology, but molecular studies helped identify several potential markers of aggressive behavior.

Meningiomas have a complex karyotype. Not surprisingly, the loss of 22q is the most frequent change seen in meningiomas. The long arm of chromosome 22 harbors the *NF2* gene, and meningiomas are commonly associated with neurofibromatosis type 2. *NF2* loss is also present in sporadic meningiomas. Although this can theoretically be useful for diagnosis, it is rarely used due to the sufficient material for histopathologic, immunohistochemical, and ultrastructural (electron microscopy) analysis. Loss of 22q in combination with 1p and 14q loss can be used to distinguish meningioma from other dural-based tumors, such as hemangiopericytomas and solitary fibrous tumors. Loss of 1p, gain of 1q, and loss of 14q have also been associated with shorter progression-free survival in adults. In pediatric cohorts, loss of 22q is commonly observed due to association with neurofibromatosis type 2, and tumors often show loss of 1p and 14q, as well, although the correlation with survival in this population is not clear. Although in gliomas the loss of *CDKN2A* is an early event of tumorigenesis, in meningiomas it is associated with higher grade and short survival.

Other genes associated with meningioma have long been elusive. Whereas loss of 17p has been observed, meningiomas rarely carry *TP53* mutations, which seem to be restricted to anaplastic tumors. The *INI1* gene that resides near *NF2* on 22q has been another candidate gene; however, it is rarely altered in meningiomas. Rare alterations of *PTEN* and *PTCH1* have been described. Whole-genome studies recently

succeeded in identifying driver mutations in non-NF2 meningiomas. Meningiomas without *NF2* aberrations can carry mutations in *TRAF7*, *KLF4*, *AKT1*, and *SMO* [58]. Mutations in these genes appear mutually exclusive with *NF2* aberrations. There is also a striking spatial distribution: meningiomas associated with *NF2* loss are found in the hemispheres, cerebellum, and spinal cord, whereas other tumors originate in the skull base. Meningiomas with *SMO* mutations are frequently present around the skull base midline, which is particularly interesting considering the role of the Shh pathway in midline brain development and its failure resulting in holoprosencephaly. Finally, there is also a striking association between histological type and mutation status. Secretory meningiomas have been defined by concurrent mutations of *TRAF7* and *KLF4* [59]. *NF2* aberrant meningiomas seem to be associated with more aggressive behavior. Considering how many distinct histological variants of meningioma exist, it will be interesting to see whether future studies will associate other morphological subtypes with specific mutations, as well. Similar to other tumors, DNA methylation has shown superior performance for predicting tumor recurrence and prognosis compared to the WHO classification [60].

Ependymoma

Ependymoma is the second most common malignant brain tumor of childhood and the most common spinal cord tumor of adults. Despite histological relatively uniform features, ependymomas from different regions of the nervous system are biologically and clinically distinct. Histological grading of ependymomas is notoriously unreliable, and, given their potential for recurrences after many years of disease-free survival and resistance to current therapies, ependymomas represent an ideal target for molecular studies [61]. Many ependymomas have complex genomes with large chromosomal gains and losses, but clear diagnostic, prognostic, or predictive markers have not been yet identified. Common genetic abnormalities in ependymoma involve losses of chromosomes 1p, 3, 6q, 9p, 10q, 13q, 16p, 17, 21, and 22q and gains of 1q, 4q, 5, 7, 8, 9, 12q, and 20. Chromosome 22 loss is probably the most frequent overall genetic abnormality in sporadic ependymoma and in *NF2*-associated ependymomas. Patients with neurofibromatosis type 2 develop a variety of central nervous system malignancies including ependymomas and meningiomas, which both show loss of 22q. Interestingly, *NF2* mutations have been identified in spinal ependymomas, but are not common in intracranial tumors. The analysis of ependymomas has been significantly enhanced by using whole-genome DNA methylation profiling and RNA sequencing [30]. Whole-genome and whole-exome sequencing of posterior fossa ependymomas revealed an extremely low mutation rate and zero significant recurrent somatic single nucleotide variants [62]. However, transcriptional profiling of posterior fossa ependymomas identified two distinct subgroups [62, 63]. Group A patients are younger, with laterally located tumors and a balanced genome. Group A tumors are also more invasive and metastatic and are more likely to recur. Group B ependymomas often arise in the posterior fossa of adults and have an unbalanced genome and grow in the midline with minimal invasion, rare metastasis, and good survival. DNA methylation profiling identified nine molecular subgroups across the CNS with supratentorial ependymomas driven by distinct gene fusions including *YAP1* and *RELA* [30]. RELA ependymoma is now recognized as a distinct WHO entity and is discussed in a separate section.

Primitive Neuroectodermal Tumors (PNET)

The diagnosis of PNET has been for decades reserved for supratentorial primitive small round blue cell tumors. This descriptive diagnosis was in contrast to small round blue cell tumors in other compartments of the brain, which were defined by their site of origin: medulloblastoma in the cerebellum, retinoblastoma in the eye, and pineoblastoma in the pineal gland region. PNET are highly aggressive tumors and diagnostically challenging as they may exhibit variety of morphological features. Using DNA methylation-based epigenetic signatures, it has been shown that a large number of tumors diagnosed as PNET are actually misdiagnosed and represent other entities [64]. From the remaining "true PNETs," four distinct molecularly defined tumor groups have been defined so far. These include CNS neuroblastoma with FOXR2 activation (NB-FOXR2), CNS Ewing sarcoma family tumor with CIC alteration (EFT-CIC), CNS high-grade neuroepithelial tumor with MN1 alteration (HGNET-MN1), and CNS high-grade neuroepithelial tumor with BCOR alteration (HGNET-BCOR). However, they are still tumors that have not been classified, and it is likely that additional molecular subtypes will be identified.

Atypical Teratoid/Rhabdoid Tumors (AT/RT)

AT/RT have long been defined by loss of *INI1* (*SMARCB1*) gene; see the section on *INI1* loss above. However, despite the similar genetic landscape, it has been recently shown that using epigenetic signatures, AT/RT can be subclassified into three distinct molecular subtypes. AT/RT appear to be comprised of three epigenetically distinct subgroups, which were termed TYR, SHH, and MYC [65]. However, whether this subclassification results in prognostic biomarker or identifies novel therapeutic targets remains unclear.

Conclusions

Current molecular neuropathology provides several tests that help with diagnosis and clinical management of patients with brain tumors. Molecular tests for 1p19q and *MGMT* and testing for *IDH1/2* by IHC and sequencing are well established and are incorporated in clinical practice as well as in clinical trials. Genetics research significantly improved our knowledge about different molecular subtypes within the tumor types, which were previously defined solely based on morphology. Reflecting this heterogeneity will help to better design clinical trials and toward the development of targeted therapies. The current amount of data clearly surpasses our understanding of it, and functional studies are needed to identify potential targets. Implementation of new diagnostic technologies in clinical laboratories will play a crucial role in identifying molecular subtypes and correct therapeutic targets. Advanced whole-genome approaches have significantly increased our ability to diagnose and classify brain tumors more accurately. However, with constantly increasing number of molecular subtypes, it is becoming challenging to envision clinical trials and management. The best example is a medulloblastoma, a prototypical molecularly defined brain tumor. While the original description included 6 molecular subclasses [66], it has been replaced by a consensus classification of 4 subtypes [38], just to be replaced by 12 molecular subtypes currently [67]. Whether any of these will actually stand the test of time and become a clinically relevant, robust, accurate, and reproducible biomarker influencing clinical management and outcome remains to be seen. However, from the practical standpoint, a constantly changing terminology and subclassification of a disease can negatively affect the feasibility of clinical laboratories, regulatory agencies, and insurance payers to develop optimal strategy for molecular profiling of brain tumors and their introduction into clinical care.

References

1. Riemenschneider MJ, Jeuken JW, Wesseling P, Reifenberger G. Molecular diagnostics of gliomas: state of the art. Acta Neuropathol. 2010;120(5):567–84.
2. Tabatabai G, Stupp R, van den Bent MJ, Hegi ME, Tonn JC, Wick W, et al. Molecular diagnostics of gliomas: the clinical perspective. Acta Neuropathol. 2010;120(5):585–92.
3. von Deimling A, Korshunov A, Hartmann C. The next generation of glioma biomarkers: MGMT methylation, BRAF fusions and IDH1 mutations. Brain Pathol. 2011;21(1):74–87.
4. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH, et al. Mutations in CIC and FUBP1 contribute to human oligodendroglioma. Science. 2011;333(6048):1453–5.
5. Yip S, Butterfield YS, Morozova O, Chittaranjan S, Blough MD, An J, et al. Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. J Pathol. 2012;226(1):7–16.
6. Horbinski C, Miller CR, Perry A. Gone FISHing: clinical lessons learned in brain tumor molecular diagnostics over the last decade. Brain Pathol. 2011;21(1):57–73.
7. Snuderl M, Eichler AF, Ligon KL, Vu QU, Silver M, Betensky RA, et al. Polysomy for chromosomes 1 and 19 predicts earlier recurrence in anaplastic oligodendrogliomas with concurrent 1p/19q loss. Clin Cancer Res. 2009;15(20):6430–7.
8. Wiens AL, Cheng L, Bertsch EC, Johnson KA, Zhang S, Hattab EM. Polysomy of chromosomes 1 and/or 19 is common and associated with less favorable clinical outcome in oligodendrogliomas: fluorescent in situ hybridization analysis of 84 consecutive cases. J Neuropathol Exp Neurol. 2012;71(7):618–24.
9. Sahm F, Reuss D, Koelsche C, Capper D, Schittenhelm J, Heim S, et al. Farewell to oligoastrocytoma: in situ molecular genetics favor classification as either oligodendroglioma or astrocytoma. Acta Neuropathol. 2014;128(4):551–9.
10. Unruh D, Schwarze SR, Khoury L, Thomas C, Wu M, Chen L, et al. Mutant IDH1 and thrombosis in gliomas. Acta Neuropathol. 2016;132(6):917–30.
11. Chen H, Judkins J, Thomas C, Wu M, Khoury L, Benjamin CG, et al. Mutant IDH1 and seizures in patients with glioma. Neurology. 2017;88(19):1805–13.
12. Huse JT, Diamond EL, Wang L, Rosenblum MK. Mixed glioma with molecular features of composite oligodendroglioma and astrocytoma: a true "oligoastrocytoma"? Acta Neuropathol. 2015;129(1):151–3.
13. Olar A, Wani KM, Alfaro-Munoz KD, Heathcock LE, van Thuijl HF, Gilbert MR, et al. IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas. Acta Neuropathol. 2015;129(4):585–96.
14. Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science. 2014;343(6167):189–93.
15. Richardson TE, Snuderl M, Serrano J, Karajannis MA, Heguy A, Oliver D, et al. Rapid progression to glioblastoma in a subset of IDH-mutated astrocytomas: a genome-wide analysis. J Neuro-Oncol. 2017;133(1):183–92.
16. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med. 2009;360(8):765–73.
17. Horbinski C, Kofler J, Kelly LM, Murdoch GH, Nikiforova MN. Diagnostic use of IDH1/2 mutation analysis in routine clinical testing of formalin-fixed, paraffin-embedded glioma tissues. J Neuropathol Exp Neurol. 2009;68(12):1319–25.
18. Joensuu H, Puputti M, Sihto H, Tynninen O, Nupponen NN. Amplification of genes encoding KIT, PDGFRalpha and VEGFR2 receptor tyrosine kinases is frequent in glioblastoma multiforme. J Pathol. 2005;207(2):224–31.
19. Puputti M, Tynninen O, Sihto H, Blom T, Maenpaa H, Isola J, et al. Amplification of KIT, PDGFRA, VEGFR2, and EGFR in gliomas. Mol Cancer Res. 2006;4(12):927–34.
20. Schlegel J, Merdes A, Stumm G, Albert FK, Forsting M, Hynes N, et al. Amplification of the epidermal-growth-factor-receptor gene correlates with different growth behaviour in human glioblastoma. Int J Cancer. 1994;56(1):72–7.
21. Pierscianek D, Kim YH, Motomura K, Mittelbronn M, Paulus W, Brokinkel B, et al. MET gain in diffuse astrocytomas is associated with poorer outcome. Brain Pathol. 2013;23(1):13–8.
22. Hobbs J, Nikiforova MN, Fardo DW, Bortoluzzi S, Cieply K, Hamilton RL, et al. Paradoxical relationship between the degree of EGFR amplification and outcome in glioblastomas. Am J Surg Pathol. 2012;36(8):1186–93.
23. Horbinski C. To BRAF or not to BRAF: is that even a question anymore? J Neuropathol Exp Neurol. 2013;72(1):2–7.
24. Lassaletta A, Zapotocky M, Mistry M, Ramaswamy V, Honnorat M, Krishnatry R, et al. Therapeutic and prognostic implications

of BRAF V600E in pediatric low-grade gliomas. J Clin Oncol. 2017;35(25):2934–41.
25. Paz MF, Yaya-Tur R, Rojas-Marcos I, Reynes G, Pollan M, Aguirre-Cruz L, et al. CpG island hypermethylation of the DNA repair enzyme methyltransferase predicts response to temozolomide in primary gliomas. Clin Cancer Res. 2004;10(15):4933–8.
26. Hegi ME, Diserens AC, Godard S, Dietrich PY, Regli L, Ostermann S, et al. Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. Clin Cancer Res. 2004;10(6):1871–4.
27. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005;352(10):997–1003.
28. Cankovic M, Nikiforova MN, Snuderl M, Adesina AM, Lindeman N, Wen PY, et al. The role of MGMT testing in clinical practice: a report of the association for molecular pathology. J Mol Diagn. 2013;15(5):539–55.
29. Eberhart CG. Molecular diagnostics in embryonal brain tumors. Brain Pathol. 2011;21(1):96–104.
30. Pajtler KW, Witt H, Sill M, Jones DT, Hovestadt V, Kratochwil F, et al. Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. Cancer Cell. 2015;27(5):728–43.
31. Orillac C, Thomas C, Dastagirzada Y, Hidalgo ET, Golfinos JG, Zagzag D, et al. Pilocytic astrocytoma and glioneuronal tumor with histone H3 K27M mutation. Acta Neuropathol Commun. 2016;4(1):84.
32. Gajjar A, Bowers DC, Karajannis MA, Leary S, Witt H, Gottardo NG. Pediatric brain tumors: innovative genomic information is transforming the diagnostic and clinical landscape. J Clin Oncol. 2015;33(27):2986–98.
33. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature. 2012;488(7409):106–10.
34. Northcott PA, Jones DT, Kool M, Robinson GW, Gilbertson RJ, Cho YJ, et al. Medulloblastomics: the end of the beginning. Nat Rev Cancer. 2012;12(12):818–34.
35. Northcott PA, Korshunov A, Witt H, Hielscher T, Eberhart CG, Mack S, et al. Medulloblastoma comprises four distinct molecular variants. J Clin Oncol. 2011;29(11):1408–14.
36. Northcott PA, Shih DJ, Peacock J, Garzia L, Morrissy AS, Zichner T, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature. 2012;488(7409):49–56.
37. Korshunov A, Ryzhova M, Hovestadt V, Bender S, Sturm D, Capper D, et al. Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers. Acta Neuropathol. 2015;129(5):669–78.
38. Taylor MD, Northcott PA, Korshunov A, Remke M, Cho YJ, Clifford SC, et al. Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol. 2012;123(4):465–72.
39. Kool M, Korshunov A, Remke M, Jones DT, Schlanstein M, Northcott PA, et al. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. Acta Neuropathol. 2012;123(4):473–84.
40. Rudin CM, Hann CL, Laterra J, Yauch RL, Callahan CA, Fu L, et al. Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. N Engl J Med. 2009;361(12):1173–8.
41. Yauch RL, Dijkgraaf GJ, Alicke B, Januario T, Ahn CP, Holcomb T, et al. Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. Science. 2009;326(5952):572–4.
42. Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, et al. Dissecting the genomic complexity underlying medulloblastoma. Nature. 2012;488(7409):100–5.
43. Northcott PA, Shih DJ, Remke M, Cho YJ, Kool M, Hawkins C, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol. 2012;123(4):615–26.
44. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010;17(5):510–22.
45. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell. 2012;22(4):425–37.
46. Hovestadt V, Remke M, Kool M, Pietsch T, Northcott PA, Fischer R, et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol. 2013;125(6):913–6.
47. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010;17(1):98–110.
48. Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD, Tang B, et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet. 2013;45(6):602–12.
49. Paugh BS, Zhu X, Qu C, Endersby R, Diaz AK, Zhang J, et al. Novel oncogenic PDGFRA mutations in pediatric high-grade gliomas. Cancer Res. 2013;73(20):6219–29.
50. Khuong-Quang DA, Buczkowicz P, Rakopoulos P, Liu XY, Fontebasso AM, Bouffet E, et al. K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. Acta Neuropathol. 2012;124(3):439–47.
51. Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet. 2012;44(3):251–3.
52. Huse JT, Snuderl M, Jones DTW, Brathwaite CD, Altman N, Lavi E, et al. Polymorphous low-grade neuroepithelial tumor of the young (PLNTY): an epileptogenic neoplasm with oligodendroglioma-like components, aberrant CD34 expression, and genetic alterations involving the MAP kinase pathway. Acta Neuropathol. 2017;133(3):417–29.
53. Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc Natl Acad Sci U S A. 2012;109(8):3041–6.
54. Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell. 2011;20(6):810–7.
55. Little SE, Popov S, Jury A, Bax DA, Doey L, Al-Sarraj S, et al. Receptor tyrosine kinase genes amplified in glioblastoma exhibit a mutual exclusivity in variable proportions reflective of individual tumor heterogeneity. Cancer Res. 2012;72(7):1614–20.
56. Motomura K, Mittelbronn M, Paulus W, Brokinkel B, Keyvani K, Sure U, et al. PDGFRA gain in low-grade diffuse gliomas. J Neuropathol Exp Neurol. 2013;72(1):61–6.
57. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;1805(1):105–17.
58. Clark VE, Erson-Omay EZ, Serin A, Yin J, Cotney J, Ozduman K, et al. Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. Science. 2013;339(6123):1077–80.
59. Reuss DE, Piro RM, Jones DT, Simon M, Ketter R, Kool M, et al. Secretory meningiomas are defined by combined KLF4 K409Q and TRAF7 mutations. Acta Neuropathol. 2013;125(3):351–8.

60. Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, Schefzyk S, et al. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. Lancet Oncol. 2017;18(5):682–94.
61. Korshunov A, Witt H, Hielscher T, Benner A, Remke M, Ryzhova M, et al. Molecular staging of intracranial ependymoma in children and adults. J Clin Oncol. 2010;28(19):3182–90.
62. Mack SC, Witt H, Piro RM, Gu L, Zuyderduyn S, Stutz AM, et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. Nature. 2014;506(7489):445–50.
63. Witt H, Mack SC, Ryzhova M, Bender S, Sill M, Isserlin R, et al. Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. Cancer Cell. 2011;20(2):143–57.
64. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DTW, Capper D, et al. New brain tumor entities emerge from molecular classification of CNS-PNETs. Cell. 2016;164(5):1060–72.
65. Johann PD, Erkek S, Zapatka M, Kerl K, Buchhalter I, Hovestadt V, et al. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. Cancer Cell. 2016;29(3):379–93.
66. Cho YJ, Tsherniak A, Tamayo P, Santagata S, Ligon A, Greulich H, et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. J Clin Oncol. 2011;29(11):1424–30.
67. Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B, et al. Intertumoral heterogeneity within medulloblastoma subgroups. Cancer Cell. 2017;31(6):737–54. e6.

Genomic Applications in Head and Neck Cancers

23

Joseph A. Bellairs, Jessica Yesensky, Jamie Ahn Ku, and Nishant Agrawal

Introduction

Head and neck cancer is the sixth most common cancer worldwide, with more than 60,000 new cases diagnosed every year in the United States and 800,000 new cases globally. The most common type of cancer of the head and neck is squamous cell carcinoma (HNSCC) arising from the epithelial layer of the upper aerodigestive tract. Most patients present with locoregionally advanced disease with a less than 50% 5-year survival rate. Furthermore, those who present with early-stage disease are at a high risk of recurrence or development of a second primary tumor. Patients with recurrent or metastatic disease endure a significantly worse prognosis with a dismal overall survival of approximately 5–7 months. The primary modes of therapy include surgical resection or radiation therapy for early-stage disease and combination therapy with surgery, radiotherapy, and/or chemotherapy for advanced-stage disease. Surgical extirpation and chemoradiation protocols for the treatment of head and neck cancer often lead to severe functional deficits and cosmetic deformities. Despite all the advances in cancer therapy, the overall survival of patients with head and neck squamous cell carcinoma has remained essentially unchanged for the past 30 years. Most recently, with the identification of a human papillomavirus (HPV)-related subset of oropharyngeal SCC, the prognosis in this specific patient population has been significantly improved with approximately 80% 3-year survival in contrast to the continuing poor survival with HPV-negative oropharyngeal SCC [1, 2].

Advancements in molecular and genetic research techniques and bioinformatics have led to an explosion of new discoveries in the molecular biology and genetic alterations behind the pathogenesis of HNSCC. Gaining further insights into the mechanisms underlying tumorigenesis and treatment response as well as advancements in screening, diagnosis, and treatment of HNSCC will ultimately lead to improved clinical outcomes.

This chapter highlights key genetic alterations in HNSCC and their clinical implications and provides a preview of the future applications of this knowledge that are in development.

Risk Factors for HNSCC

The general principles underlying mechanisms behind tumorigenesis in HNSCC are thought to be similar to those in other solid tumors. Cancer arises from progressive accumulation of genetic or epigenetic alterations that lead to the development of malignant phenotypes. Prolonged, cumulative exposure to certain carcinogens is thought to be the leading cause of specific alterations acquired during tumor progression. Tobacco and alcohol represent the two predominant carcinogens that are synergistically responsible for HNSCC development [3–5]. In contrast, oral HPV infection, the main cause of the cancer of the oropharynx, is believed to act independently.

Tobacco and Alcohol

Numerous epidemiological studies conducted in different regions of the world have demonstrated a compelling association between sustained exposure to tobacco, tobacco-like products, and alcohol and the risk of HNSCC [5–9]. Several large case-control studies have shown neoplastic effects of

J. A. Bellairs
Department of Otolaryngology – Head and Neck Surgery, University of Washington Affiliated Hospitals, Seattle, WA, USA

J. Yesensky · N. Agrawal (✉)
Department of Surgery, Section of Otolaryngology – Head and Neck Surgery, University of Chicago Medicine, Chicago, IL, USA
e-mail: na@uchicago.edu

J. A. Ku
Head and Neck Institute, Head and Neck Surgery and Oncology, Cleveland Clinic Foundation and Lerner College of Medicine at Case Western Reserve University, Cleveland, OH, USA

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_23

tobacco use and alcohol consumption with a linear correlation with both duration and amount with odds ratio of 2- to 6-fold for alcohol and 7- to 20-fold for tobacco. When combined, these carcinogens had multiplicative or even supra-multiplicative effects with a greater than 35- to 200-fold risk for individuals who consume more than 2 or more packs of cigarettes and more than 4 drinks per day [6, 8, 9]. One explanation for the synergistic effect of alcohol and tobacco is that alcohol possibly acts as a solvent for penetration of other carcinogens through the mucosa of the upper aerodigestive tract [10].

According to the World Health Organization, tobacco use is the single most preventable risk factor for cancer deaths worldwide, responsible for 22% of all cancer mortality [11, 12]. There are over 4800 chemicals in processed tobacco, of which at least 250 are known to be harmful and more than 50 are known carcinogens. These include polycyclic aromatic hydrocarbons, *N*-nitrosamines, aromatic amines, aldehydes, volatile hydrocarbons, and metals [13]. Cigarette smoke also contains free radicals, nitric oxide, and other unstable oxidants that induce oxidative DNA damage [14, 15]. The human body responds to carcinogens by detoxifying and excreting them through a series of enzymatic processes. However, metabolites resulting from detoxification can also be reactive and cause DNA damage through the formation of DNA adducts [16, 17]. Therefore, it has been proposed that genetic susceptibility or polymorphisms in detoxifying enzymes, such as cytochrome P-450 and glutathione S-transferase, can lead to the development of cancer by either failing to deactivate carcinogens or to activate pro-carcinogenic intermediates [18–20].

The mechanisms by which alcohol exerts its carcinogenic effects have been linked to alcohol metabolism, DNA damage, and DNA methylation. Acetaldehyde, the primary metabolite of ethanol, can form adducts with DNA and thus result in DNA damage [21]. There is also growing evidence that genetic polymorphisms in enzymes for oxidation of ethanol into acetaldehyde modulate alcohol-related cancer risks, which further supports the mechanistic role of acetaldehyde [22, 23]. In addition, heavy alcohol intake leads to nutritional deficiencies, including that of vitamins B12, B6, and A and folate. This may also result in changes in DNA methylation and transcription patterns that promote tumorigenesis [21].

Human Papillomavirus

HPV infection is the most common sexually transmitted infection in the Unites States. There are over 100 HPV subtypes, and these are categorized into low-risk and high-risk subtypes. Among the high-risk group, HPV-16 and HPV-18 are the two leading subtypes responsible for cancer development [24]. In a recent large cross-sectional study conducted as part of the National Health and Nutrition Examination Survey (NHANES), the prevalence of oral HPV infection in the general population was determined to be about 6.9%, with a prevalence rate of 3.7% for high-risk HPV infection. The most prevalent HPV subtype detected was HPV-16 with a prevalence of 1.0% [25].

High-risk HPV has long been known to cause cervical cancer in women, penile cancer in men, and anal cancer in both men and women. Over the past 10 years, there has been overwhelming evidence that implicates HPV as a causative factor in a subset of HNSCC, mainly of the oropharynx, where up to 50–70% of the cases are associated with high-risk HPV [26, 27]. In contrast to cervical cancer, where HPV-16 and HPV-18 are together known to cause 70% of the cases, HPV-related HNSCC is exclusively caused by the HPV-16 subtype, with up to 90% of the cases being HPV-16 positive. While the overall incidence of head and neck cancer is on the decline over the last two decades, the incidence of oropharyngeal cancer is on the rise [28]. This steady incline in the incidence of oropharyngeal cancer is mirrored by the increase in the incidence of HPV-positive oropharyngeal cancer, while the rates of HPV-negative oropharyngeal cancer have been decreasing over the same period of time [27].

Molecular Biology of Head and Neck Cancers

The advances in our understanding of cancer genomics have further elucidated the biological complexity of HNSCC. The disease represents a heterogeneous collection of tumors in which multiple genes and pathways are altered (Table 23.1) (Fig. 23.1). In-depth understanding of the pathways implicated in HNSCC tumorigenesis is critical for the identification of new "personalized" therapeutic strategies.

TP53

The role of *TP53*, a tumor suppressor gene on chromosome 17p12, in HNSCC carcinogenesis is well established in the literature. *TP53* is the most commonly mutated gene in HNSCC, with approximately 50–70% all HNSCC tumors having a *TP53* mutation [29–32]. In normal cells, *TP53* plays a critical role in regulating the cell cycle in response to DNA damage. *TP53* is activated by exposure to cellular stress such as DNA damage, which results in the accumulation of active

Table 23.1 Common genetic alterations in primary head and neck squamous cell carcinoma

Gene symbol	Gene name	Location	Frequency	Function	Clinical applications
Tumor suppressor gene					
TP53	Tumor protein p53	17p13.1	47–72%	Assists in cell cycle arrest to allow DNA repair, apoptosis, or cell senescence	Biomarker for poor prognosis—decreased survival and therapy resistance; for analysis of margin status; adenoviral gene therapy
CDKN2A/p16	Cyclin-dependent kinase inhibitor 2A	9p21.3	9–22%	Regulates of G1-to-S phase transition in cell cycle and cell senescence	IHC for p16 as a surrogate marker for HPV-related tumor
FAT1	FAT atypical cadherin 1	4q35.2	12–23%	Poorly understood, regulates Wnt/β-catenin signaling	
Oncogene					
EGFR	Epidermal growth factor receptor	7p12	a	Activates critical signaling pathways in proliferation, migration, invasion, angiogenesis, and apoptosis	Fluorescent bioconjugated anti-EGFR molecules for intraoperative optical imaging; anti-EGFR-targeted therapies
HRAS	Harvey rat sarcoma viral oncogene homologue	11p15.5	4–5%	Promotes cell proliferation, differentiation, morphology, and survival	
PIK3CA	Phosphoinositide-3-kinase catalytic alpha polypeptide	3q26.32	6–21%	Promotes cell growth, survival, and cytoskeleton organization	
Both (tissue dependent)					
NOTCH1	Notch1	9p34.3	14–19%	Regulates of cell differentiation, lineage commitment, and embryonic development	Therapeutic inhibition or activation of NOTCH1 pathway

IHC immunohistochemistry, *HPV* human papillomavirus
[a]Overexpression in 80–90% of HNSCC

TP53 protein in the nucleus. Through transcriptional induction of downstream signaling pathways, it induces viable cell growth arrest or apoptosis. Therefore, this growth inhibitory effect of *TP53* is vital in preventing the proliferation of cells harboring damaged DNA or of cells with the potential for neoplastic transformation [33]. Functional loss of *TP53* is one of the most common genetic alterations in many types of human cancer, and mutations in this gene play a critical role in malignant transformation [34]. In HNSCC, alterations in *TP53* occur early in the premalignant squamous epithelium, before invasive transformation. For instance, premalignant oral lesions have been shown to harbor *TP53* mutations in up to 35% of cases [35, 36]. In fact, the incidence of mutations increases with histological progression from severe dysplasia to invasive carcinoma [37]. Furthermore, the frequency of *TP53* genetic alterations in patients with a history of tobacco and alcohol use is almost double that of those without such history [30, 38, 39].

FAT1

FAT atypical cadherin 1 (*FAT1*) is one of the newest genes implicated in HNSCC. In mammals, the FAT protocadherin family includes *FAT1*, *FAT2*, *FAT3*, and *FAT4* that are all closely related to the *Drosophila* tumor suppressor *Fat*. In *Drosophila*, *Fat* is involved in cell cycle regulation and proliferation [40–43]. In mammals, *FAT4* most closely resembles *Fat* and has been implicated in planar cell polarity and Hippo signaling [44–46]. *FAT1*, located on chromosome 4q35.2, is not thought to play a significant role in these processes. Historically, the function of *FAT1* was incompletely understood, but some reports tentatively described it as a tumor suppressor gene [40, 42]. More recently, there has been a renewed interest in *FAT1* after a study demonstrated that *FAT1* is mutated in multiple malignancies, including glioblastoma, colorectal cancer, and HNSCC. Furthermore, this study demonstrated that *FAT1* loss-of-function mutations result in Wnt pathway activation and tumorigenesis [47]. Aberrant activation of the Wnt/β-catenin signaling pathway is implicated in the development of many different malignancies, but genetic basis of aberrant pathway activation is incompletely understood in many malignancies [48–51]. This important study links genetic lesions of *FAT1* the aberrant Wnt/β-catenin signaling, thus solidifying the role of *FAT1* as a tumor suppressor gene that can drive oncogenesis in certain malignancies.

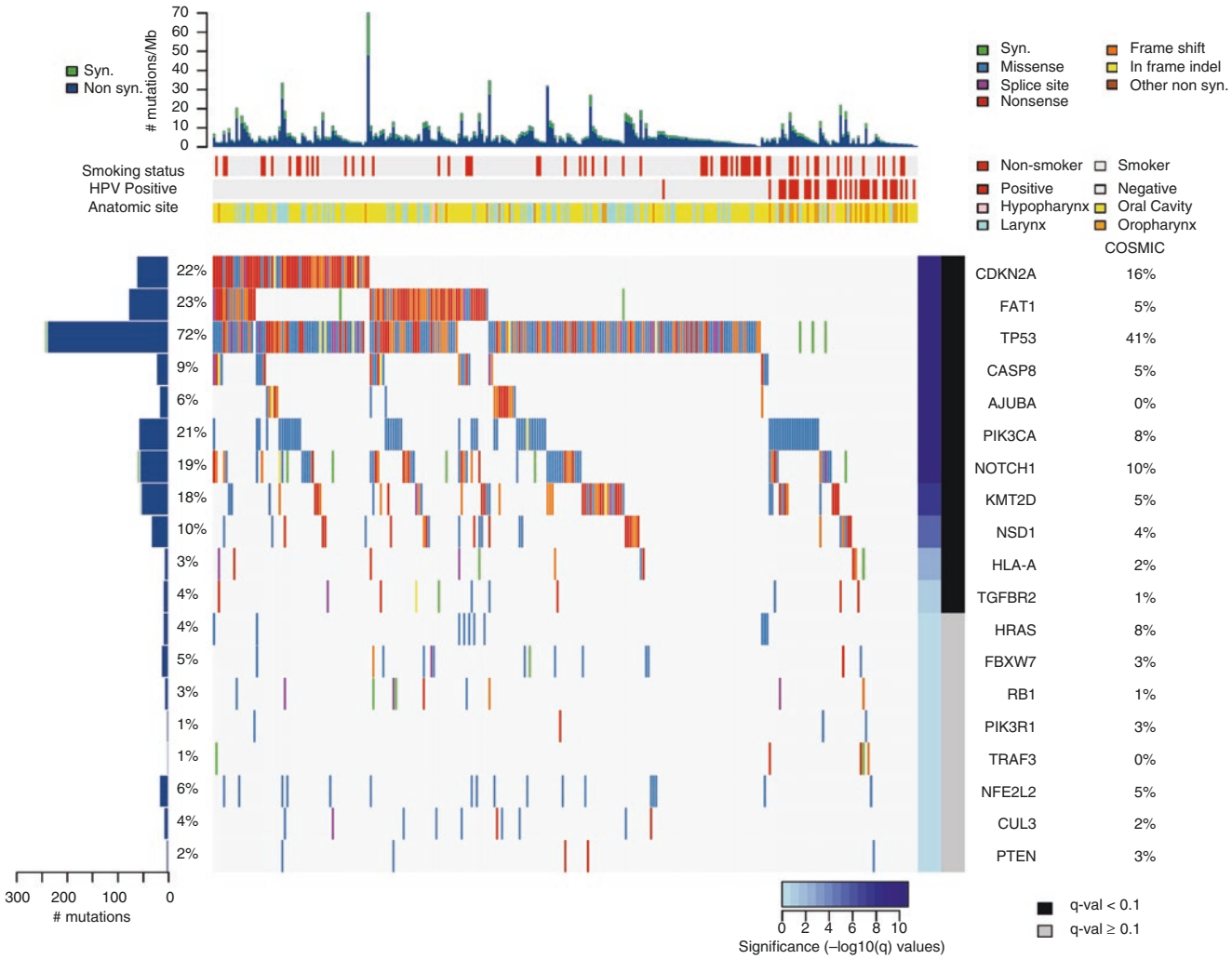


Fig. 23.1 Molecular mutation landscape in HNSCC. Genes (rows) with significantly mutated genes. Samples are arranged to emphasize mutual exclusivity among mutations. Left, mutation percentage in TCGA. Right, mutation percentage in "upper aerodigestive tract" tissue. Top, overall number of mutations per megabase. Color-coding indicates mutation type. (From Comprehensive genomic characterization of head and neck squamous cell carcinomas [29], with permission of Springer Nature)

The role of *FAT1* in HNSCC is evolving. In several recently published whole-exome next-generation sequencing (NGS) studies of HNSCC, *FAT1* mutations were identified in 12–23% of HNSCC patients [29, 32]. Preliminary studies are underway to determine if *FAT1* mutation status is associated with patient outcomes and to investigate if the association between *FAT1* and Wnt//β-catenin signaling pathway represents a new therapeutic approach in HNSCC [52, 53]. Regardless, the frequency of *FAT1* mutations in HNSCC warrants further research to better characterize the molecular mechanism that links *FAT1* with Wnt/β-catenin signaling.

CDKN2A/P16

Cyclin-dependent kinase inhibitor 2A (*CDKN2A*) located on chromosome 9p21 is a known tumor suppressor gene frequently disrupted in HNSCC. It is inactivated through deletion, point mutations, and epigenetic promoter methylation. Loss of heterozygosity on the short arm of chromosome 9 has been reported frequently, and this change has been recognized as an early event in the progression of premalignant lesions to HNSCC [54, 55]. In the previously mentioned whole-exome NGS projects, *CDKN2A* mutations were identified in 9–22% of all tumors [29, 31, 32]. When considering genetic and epigenetic alterations, *p16* inactivation has been detected in as much as 80% of HNSCC [56].

The *CDKN2A* gene encodes the protein product *p16* that plays an important role in regulating the G1 phase of the cell cycle. The p16 protein binds to cyclin-dependent kinase 4 (CDK4) and CDK6, inhibiting their association with cyclin D1. This inhibition of cyclin D1/CDK4/6 complex activity prevents pRb phosphorylation and the release of E2F transcription factor, leading to the inhibition of the G1-to-S phase transition and thus leading to cell senescence [57, 58].

Therefore, any genetic abnormalities inactivating the *p16* pathway may confer growth advantages in cells, contributing to the tumorigenic process.

NOTCH

NOTCH1 was recently identified as a new cancer gene implicated in HNSCC development. With a mutation frequency of 14–19%, *NOTCH1* is one of the most frequently mutated genes in HNSCC [29, 31, 32]. Previously, only a few functional studies described a role for *NOTCH1* in squamous cell oncogenesis, specifically of the skin [59, 60]. However, given the large size of the *NOTCH1* gene which comprises 34 coding exons, robust mutational data analysis was required to comprehensively detect these mutations.

There are four *NOTCH* family receptors in humans, *NOTCH1* to *NOTCH4. NOTCH1* encodes a transmembrane receptor that functions in regulating normal cell differentiation, lineage commitment, and embryonic development. After ligand binding, the *NOTCH1* intracellular domain (NICD) is cleaved, and the translocation of the NICD to the nucleus is necessary for transcriptional activation of downstream signaling. The *NOTCH1* ligands include Jagged 1 and 2 and Delta-like ligand 1, 3, and 4. After receptor activation through ligand binding, the release of NICD requires a two-step cleavage process. First, the extracellular portion of the protein is released by proteases TNF-alpha-converting enzyme (TACE) and a disintegrin and metalloprotease (ADAM). A second cleavage by gamma-secretase complex liberates the NICD from the membrane [61]. In the nucleus, NICD interacts with transcriptional regulators and activates downstream target genes, such as the *HRT* and *HES* family of genes, which are crucial for cell differentiation and normal embryonic development.

The role of *NOTCH1* in cancer has been recently described, with *NOTCH1* signaling having both oncogenic and tumor-suppressive roles depending on the cellular context. For instance, activating truncation mutations in *NOTCH1* have been reported in acute lymphoblastic leukemia and chronic lymphocytic leukemia, implicating *NOTCH1* as an oncogene in these hematopoietic cancers [62, 63]. In contrast, the initial findings of inactivating mutations in HNSCC and the observation that loss of *NOTCH1* in murine models led to skin carcinogenesis indicated that *NOTCH1* may also act as a tumor suppressor gene [59, 60]. The data thus far are conflicting with regard to the exact role of *NOTCH1* in HNSCC. Most *NOTCH1* mutations observed in HNSCC affect the epidermal growth factor (EGF)-like ligand-binding domain and are thought to lead to loss of function, suggesting the role of *NOTCH1* as a tumor suppressor [31]. Contrary to the genetic evidence, there is evidence that *NOTCH1* protein levels are elevated in HNSCC, and tumors expressing higher levels of *NOTCH1* protein are associated with reduced survival as well as with chemoresistance [64–67]. In support of the latter, it was demonstrated that inhibition of the *NOTCH1* pathway using gamma-secretase inhibitors (GSIs) prevented the growth of HNSCC cell lines [68]. These findings suggest that activated *NOTCH1* could function as an oncogene. Additional functional studies in vitro and in vivo are required to elucidate the exact role of *NOTCH1* in HNSCC.

EGFR

Malignant transformation of HNSCC is also driven by alterations in growth factor signaling pathways. Epidermal growth factor receptor (EGFR), also known as HER1 or ErbB-1, is a tyrosine kinase receptor that is highly expressed in normal epithelial cells. EGFR is activated by several ligands, which induces receptor dimerization and autophosphorylation, resulting in activation of downstream signaling pathways [69]. These downstream pathways include *MAPK*, *PI3K/AKT*, *ERK*, and *JAK/STAT* genes that are critical for the regulation of cellular proliferation, apoptosis, angiogenesis, migration, and invasion [70]. The *EGFR* gene is overexpressed in 80–90% of HNSCC via gene amplification and transcriptional activation [71, 72]. In addition to overexpression, a mutant form of *EGFR* known as *EGFRvIII* has been implicated in resistance to anti-*EGFR*-targeted therapies [73]. This mutant form is characterized by a deletion in exons 2–7, leading to a truncated ligand-binding domain, rendering it constitutively active. Overactivation of EGFR signaling via overexpression or activating mutations enables cells to take on a malignant phenotype.

RAS

The *RAS* gene family consists of three genes that function as small GTPase molecules: *KRAS*, *HRAS*, and *NRAS*. The *RAS* genes play a critical role in cell signaling as part of the *RAS–RAF–MEK–MAPK* pathway. This pathway is involved in the regulation of cell proliferation, differentiation, morphology, and survival. The *RAS* gene family mutations have been implicated in approximately one-quarter of all human cancers, with *KRAS* being the most common and *HRAS* the least common [74]. However, in HNSCC, *KRAS* mutations are virtually absent, while *HRAS* mutations have been described at a low frequency of approximately 4–5% [29, 31, 32].

PIK3CA

The *PI3K–PTEN–AKT* pathway is another critical pathway in HNSCC carcinogenesis. The *PIK3CA* gene is located on chromosome 3q26 and functions to convert

phosphatidylinositol (4,5) biphosphate (P4,5P2) into phosphatidylinositol (3,4,5) triphosphate (PIP3), in turn activating Akt/PKB kinases. This results in the promotion of cell growth, survival, and cytoskeleton reorganization [75]. *PIK3A* is downstream of receptor tyrosine kinases such as EGFR, Met, and vascular endothelial growth factor receptor (VEGFR), which are known oncogenes in HNSCC. The prevalence of *PIK3CA* mutations was estimated to be approximately 6–21% in the above-referenced whole-exome sequencing HNSCC studies [29, 31, 32]. The overactivation of this pathway occurs through both amplification and mutations in *PIK3CA* as well as through *PTEN* loss [75]. PTEN (phosphatase and tensin homologue deleted on chromosome ten) is a key regulator of PI3K function. PTEN reverses the action of PI3K by removing the 3′ phosphate and thus preventing the activation of downstream molecules such as Akt [76]. A study conducted by Pedrero et al. found evidence for *PIK3CA* amplification in 37% of primary HNSCC tumors and in 39% of premalignant lesions, indicating that *PIK3CA* amplification could be an early event in HNSCC oncogenesis [77].

Human Papillomavirus

HPV is a non-enveloped small double-stranded, circular DNA virus that infects epithelial cells [78]. The majority of HPV subtypes cause epithelial lesions with low malignant potential, such as warts or papillomas. However, there is a subset of high-risk HPV that leads to precancerous lesions. Interestingly, only a small fraction of people infected with high-risk HPV will eventually develop cancer, often decades after the original infection.

The molecular mechanism behind HPV-driven carcinogenesis has been extensively studied in cervical cancer (Fig. 23.2). The integration of high-risk HPV DNA into the host genome results in the expression of oncogenes E6 and E7 in the host cell. The E6 oncogene binds to tumor suppressor *TP53*, which causes the degradation of *TP53* via ubiquitin-mediated processes. The degradation of *TP53* prevents the host cell from engaging in cell cycle checkpoints and enduring an apoptotic response [79]. The E7 oncogene is the most important driver of cell cycle deregulation through the binding and destabilizing of the tumor suppressor retinoblastoma (pRb). This binding of pRb results in the release of E2F transcription factors, leading to the transcription of genes involved in proliferation and cell cycle progression [80]. One of the main molecular pathways amplified through E7 is the *CDKN2A*/*p16* gene pathway, which results in the overexpression of p16 protein. Whereas in tobacco-induced HNSCC, the abrogation of TP53 and pRb pathways occurs via mutation and epigenetic alterations, in HPV-related HNSCC, wild-type TP53 and pRb are functionally inactivated by the viral oncogenes. E7 also induces cellular proliferation by disrupting the activity of cyclin-dependent kinase inhibitors p21 and p27 [80]. E5 is another viral protein that modulates the *EGFR* signaling pathway by delaying the downregulation of *EGFR* and increasing the level of EGFR [81]. In summary, HPV infection induces failures in cell cycle checkpoints, which causes genetic instability and, over time, progression of premalignant lesions to invasive squamous cell carcinoma.

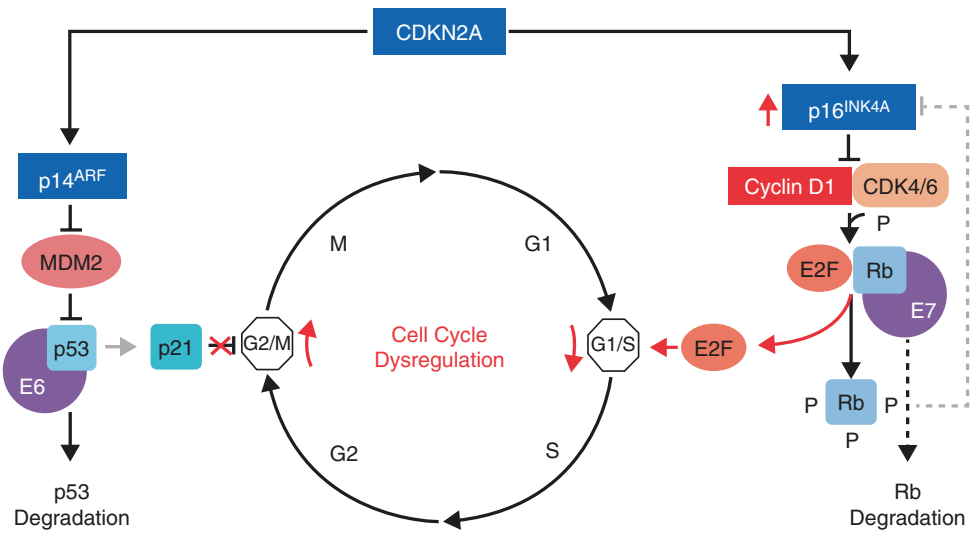


Fig. 23.2 Molecular mechanism of HPV-related carcinogenesis. The *CDKN2A* gene tumor suppressor p14 and p16 proteins. p14 leaks to disinhibition of p53, leading to activation of p21 and stopping cell cycle. HPV E6 protein binds and targets p53 for degradation, resulting in loss of cell cycle regulation. p16 inhibits cyclinD1/CDK4–6 complexes which in turn phosphorylate Rb, leading to transcription factors that halt the cell cycle. HPV E7 protein bind Rb for degradation, thereby promoting proliferation. (From Faraji et al. [82], Fig. 3, with permission of Elsevier)

Clinical Implications of Molecular Alterations in Head and Neck Cancers

The ultimate goal of understanding the molecular biology of HNSCC is to help improve patient outcomes. Improving clinical outcomes can be achieved through three sets of clinical applications: (1) applications that help achieve more accurate and earlier detection of disease, improved therapeutic monitoring, and better surveillance of recurrence; (2) accurate markers of predicting prognosis and therapeutic outcome to identify patients that will require aggressive treatment strategies; and (3) identification of novel therapeutic targets tailored to patient's tumor profile. Some of the molecular biomarkers under investigation in each of the above applications are discussed next.

Diagnostic Applications in Head and Neck Cancers

Human Papillomavirus

The primary methods of identifying HPV-related HNSCC in the clinical arena are in situ hybridization (ISH) of HPV DNA, HPV DNA polymerase chain reaction (PCR) testing, and p16 immunohistochemistry (IHC) as a surrogate marker [83]. ISH can be performed using either fluorescently labeled or chromogenic HPV type-specific probes in formalin-fixed and paraffin-embedded sections. The signals originating within the nuclei of tumor cells usually indicate HPV genome integration. Although this method has lower specificity and sensitivity than Southern blot hybridization or PCR amplification, it has the advantage of detecting genomic integration and being technically relatively less demanding [84]. With the introduction of signal amplification techniques, the sensitivity of this method has increased significantly, even to the point of detection of one viral copy per cell [85].

The use of PCR to amplify HPV DNA is an emerging approach to diagnose HPV-related HNSCC. On the one hand, this approach is highly sensitive and may detect even a single copy of the target DNA in a given sample. On the other hand, the high sensitivity can lead to false-positive results due to contamination and may detect HPV genome that is present but may not be causing the malignancy. Currently, PCR-based HPV detection is being investigated as a tool for early detection and surveillance of disease using saliva or serum [86, 87]. PCR-based methods are also being utilized in large epidemiologic studies conducted to determine the general incidence of HPV oral infection [25]. HPV16 serology can detect antibodies produced in response to HPV infection or HPV immunization. However due to the high prevalence rates in the general population, the exact role of this detection tool in the management of HNSCC is yet to be determined. Finally, NGS high-throughput technology may have a role in HPV detection. Several recent studies have demonstrated proof-of-concept results that indicate the ability of NGS to assess HPV status as well as viral load and genomic copies in tumors [32, 88, 89].

P16 overexpression may be used as a surrogate marker for HPV-positive cancers due to the inactivation of the pRb protein by the HPV protein E7 [90]. In many centers and laboratories, IHC of p16 in formalin-fixed, paraffin-embedded tissues is the main method of detecting HPV-16 positivity. In comparison to other techniques, IHC does not require specialized equipment or tissue handling. The concordance rate between HPV-16 ISH and p16 IHC has been shown to be approximately 92–93% [84, 91]. Despite the strong correlation, several studies have shown that not all p16-positive cancers are due to HPV [84]. The hypothesis is that the discordance is due to the presence of HPV subtypes other than HPV-16. Therefore, the question of whether p16 should be used as proxy for HPV-16 status in the management of HNSCC remains unanswered, and some advocate using a combination of HPV-16 ISH as well as p16 IHC for detection of HPV-16-related HNSCC [84].

For pathologists, HPV testing can play a critical role in specific clinical situations. For instance, detection of HPV in regional or distant metastatic foci can suggest the tumor origin to be likely from the oropharynx [92, 93]. This is especially important given that about 13% of patients with HNSCC present with a neck mass as their first and only clinical manifestation, and 3–9% of these patients fail to have their primary site detected upon clinical and radiologic evaluation [94]. In another scenario, detection of HPV in cystic neck lesions can provide compelling evidence of a metastatic malignant process rather than a benign process such as a branchial cleft cyst.

EGFR for Optical Imaging

The presence of positive margins following tumor resection is a known poor prognostic indicator that results from the infiltrative nature of head and neck cancers. Currently, objective intraoperative means of defining tumor margins, other than conventional crude methods of macroscopic and microscopic visual inspection and palpation, are lacking. Therefore, a novel way to reliably identify tumor margins using intraoperative real-time imaging would potentially have a significant impact on decreasing the rate of postoperative positive margins while sparing uninvolved surrounding tissues. Systemic administration of fluorescently labeled antibodies targeting cancer-specific molecules is under investigation in several European clinical trials in multiple types of cancers [72, 73]. In HNSCC, a promising intraoperative imaging system under clinical investigation is one using monoclonal anti-EGFR antibodies, such as panitumumab and cetuximab, conjugated with indocyanine green dye (IRDye800CW) [95]. Preclinical in vivo studies using orthotopic HNSCC

xenografts demonstrated that tumor tissue was clearly delineated from normal tissue on fluorescence guidance as confirmed by histology. The researchers were also able to detect subclinical microscopic residual disease as well as lymph node metastases measuring <1.0 mm [96]. A recent proof-of-concept study conducted in human subjects demonstrated successful in vivo fluorescent imaging results in nine HNSCC patients injected with cetuximab-IRDye800 3–4 days prior to planned surgical procedures [97]. Fluorescent bioconjugated anti-EGFR nanoparticles or peptides are also being investigated in various solid tumors such as esophageal cancer, glioblastoma, and epidermoid tumors [98–100]. These nanoparticles or peptides may prove to be more efficient when compared to fluorescent anti-EGFR antibodies due to a shorter half-life and superior tissue penetration and distribution [101]. Further studies are required to elucidate the potential value of these innovative optical molecular imaging techniques in improving surgical outcome and ultimately patient survival.

Prognostic Applications in Head and Neck Cancers

TP53

Mutations in tumor suppressor gene *TP53* have been associated with poor survival as well as decreased response to treatment in HNSCC. In a large multicenter prospective study, the presence of any *TP53* mutation was associated with decreased overall survival with a hazard ratio of 1.4, and the presence of *TP53* alterations that disrupt the DNA-binding domain was found to be more significantly associated with decreased survival with a hazard ratio of 1.7 [102]. Furthermore, alterations in the *TP53* gene have been implicated in poor tumor response to chemoradiation. In one study, a 95% overall incidence of *TP53* inactivation via mutation or deletion was encountered in patients with recurrent HNSCC refractory to radiotherapy [103]. One possible mechanism of radioresistance is through the inhibition of radiation-induced senescence [104]. The risk of locoregional treatment failure following primary radiation treatment or postoperative adjuvant radiation therapy is shown to be significantly greater in patients whose tumor contained mutant *TP53* genes [105, 106]. Finally, *TP53* mutation status has been found to be an independent negative predictor of response to induction and neoadjuvant chemotherapy in both retrospective and prospective studies [107, 108].

The high incidence of *TP53* alterations and the well-established prognostic value of *TP53* attest to the importance of developing a clinically robust tool to detect *TP53* mutations. One specific clinical application that has been extensively studied is *TP53* mutational status in surgical margins. It has been implied that the *TP53* mutational status at histologically tumor-free surgical margins may be critical in predicting locoregional failure, especially since the genetic alterations in *TP53* precede histologically identifiable changes at the tissue level. In fact, the detection of *TP53* mutations via molecular analysis in histologically "negative" margins has been shown to be a reliable prognostic marker of locoregional tumor recurrence [109–111]. Currently, the main method of detecting *TP53* mutations is through IHC, which cannot detect all types of mutations and has limited sensitivity in application to precancerous lesions [112]. Hence, the IHC should be complemented by genetic analysis, via PCR methods or oligonucleotide probe array technique, to increase the sensitivity and specificity of the detection of altered *TP53*.

Human Papillomavirus

Cumulative data from a large number of retrospective and prospective studies have consistently demonstrated a superior outcome in individuals with HPV-positive oropharyngeal SCC (OPSCC) compared to those with HPV-negative tumors [1, 113–115]. In multiple meta-analyses evaluating the impact of HPV infection on survival outcomes, site-specific analysis showed that patients with HPV-positive OPSCC had a 28–60% reduced risk of death for overall survival in comparison to patients with HPV-negative oropharyngeal tumors [113, 116]. Interestingly, there was no difference in the overall survival between HPV-positive and -negative non-oropharyngeal patients [113]. The authors therefore concluded that the observed improved survival benefit for HPV-positive HNSCC patients is specific to the oropharynx subsite. In the first prospective clinical trial to demonstrate survival benefit in HPV-positive HNSCC, Fakhry et al. reported that patients with HPV-positive tumors had a higher response rate after induction chemotherapy (82% versus 55%) and after chemoradiation treatment (84% versus 57%) [114]. With a median follow-up of 39.1 months, patients with HPV-positive tumors also had improved overall survival (95% versus 62%) and decreased risks of progression (with a hazard ratio of 0.27) and risk of death from any cause (hazard ratio of 0.36) than those with HPV-negative tumors [114]. It is also important to note that the positive prognostic benefit of HPV in OPSCC patients is often mitigated by the negative prognostic effects of smoking. In a large-scale retrospective study, patients with OPSCC were able to be stratified into three prognostic groups: low-, intermediate-, and high-risk groups based on HPV status, smoking, and nodal and primary tumor staging. Patients in the high-risk category had a 3-year overall survival of only 46.2% versus 93.0% for low-risk patients [1]. Therefore, when determining the best treatment option, it is important to realize that there is a subgroup of HPV-positive patients who may remain in need of more aggressive therapy.

Detoxification Enzymes

Detoxification enzymes, such as glutathione S-transferase (GST) and cytochrome P450, oxidize carcinogens into reactive metabolites that can lead to DNA damage and eventual development of cancer. Several studies have evaluated the role of this group of enzymes as prognostic markers in head and neck cancer. One study examining a subtype of GST, GSTT1, found that patients with the functional genotype were three times more likely to die from HNSCC after adjusting for age, primary therapy, and stage of disease [117]. Alternatively, these enzymes may also be a marker for chemotherapy resistance to cisplatin by inactivating reactive oxygen species induced by cisplatin to kill the offending tumor cells [118]. Several studies have demonstrated that patients with high levels of GSTpi had worse overall survival following treatment with chemotherapy [119], and the survival was worst in the group of patients who were treated with chemoradiotherapy and had elevated levels of both GSTpi and TP53 [120]. Additionally, a study in a Hungarian HNSCC cohort demonstrated that carriers of specific allelic polymorphisms of cytochrome P450 1A1 (CYP1A1) and uridine-diphosphate-glucuronosyltransferase 1A1 (UGT1A1) had the worst prognosis [121].

Therapeutic Applications in Head and Neck Cancers

De-intensification for Human Papillomavirus-Related HNSCC

The recognition of HPV-associated HNSCC in the younger, nonsmoker, and nondrinker population with improved overall prognosis has led some authors to consider revisiting the standard treatment paradigm in this group of patients [122]. The concept of de-intensification for HPV-positive OPSCC has gained attention, with the ultimate goal of achieving acceptable cure rates while minimizing long-term morbidity. Numerous clinical trials to address this question are underway, although no evidence-based de-intensification protocol is currently being utilized in the clinical setting. A recent phase II clinical trial for patients with stage III/IV resectable HPV-positive OPSCC showed promising results. Patients that achieved complete clinical responses to induction chemotherapy received dose-reduced intensity-modulated radiation therapy (IMRT), and those that did not achieve a complete response received standard-dose IMRT. Two-year overall survival rates for the reduced radiation dose group and standard radiation group were 94% and 87%, respectively [123]. Two-year overall oncologic and functional outcome data of minimally invasive surgical approaches, specifically the transoral robotic surgery (TORS), are also slowly emerging with so far promising results [124–128].

Human Papillomavirus Vaccines and Immunotherapy

There are three prophylactic vaccines currently available that have already been shown to be highly effective at preventing cervix infections by high-risk HPV subtypes [129, 130] as well as associated cervical neoplasia [131, 132]. Gardasil(®) is a quadrivalent vaccine containing virus-like particles (VLPs) of subtypes 6, 11, 16, and 18; Gardasil 9(®) is a newer vaccine that contains VLPs of the same four subtypes in addition to subtypes 31, 33, 45, 52, and 58; and Cervarix(®) is a bivalent vaccine containing VLPs of subtypes 16 and 18. These preventive HPV vaccines focus on generating neutralizing antibodies through the humoral immune response for the prevention of future infections. This involves the interaction between cell surface HPV capsid antigen (L1 and/or L2) and an antigen-specific B-cell receptor, which eventually results in the proliferation of HPV antigen-specific B cells via CD4+ helper T cells. Upon subsequent exposures to HPV, memory and plasma B cells produce HPV-specific antibodies that bind to the virus and prevent its entry into host cells. The prophylactic vaccines are anticipated to reduce the incidence of cervical cancer, although their long-term success is yet to be determined. This question will not be fully addressed for decades, until sufficient time has passed postvaccination when subjects are expected to develop cervical neoplasms. Furthermore, the vaccine's impact on HPV-associated HNSCC is unknown, as none of the studies performed thus far evaluated the effect on the incidence of oral HPV infection or oral immunity to HPV. Finally, this preventive vaccination strategy is not effective for the treatment of existing infections or established HPV-related lesions.

Treatment of established HPV disease requires cell-mediated immunity that recognizes and eliminates virus-infected cells. Therefore, therapeutic HPV vaccination strategies focus on generating cell-mediated immunity for the clearance of infected cells including HPV-associated tumor cells by using intracellular viral protein as the antigen. The HPV E6 and E7 oncoproteins are ideal tumor antigens since they are "foreign" viral proteins that are uniquely expressed by every virus-related cancer cell. The HPV E6 and E7 antigens are utilized to prime naïve T cells to become effector T cells, namely, CD8+ cytotoxic T lymphocytes and CD4+ T helper cells. These effector T cells mediate antigen-specific killing of both infected cells and tumor cells. There are numerous different phase I and II clinical trials for therapeutic HPV vaccines that are underway, mostly in cervical cancer but also in head and neck oropharyngeal cancer [133]. One of the challenges with therapeutic vaccination is generating a robust T-cell response specific to the target antigen. Currently multiple strategies are being evaluated to increase the immunogenicity of the therapeutic vaccines [134].

Immune Checkpoint Inhibitors

Deregulation of the cancer immune surveillance is a fundamental step in the carcinogenesis of HNSCC, and consequently, a robust immune system is a crucial factor in controlling neoplastic growth [135]. Under normal circumstances, T lymphocytes are responsible for discriminating between self and non-self-antigens after presentation by antigen-presenting cells, such as dendritic cells. HNSCC, like many other malignancies, displays a high rate of genetic instability and as a result presents a high number of "neoantigens" to competent immune system [136]. To counteract the host antitumor effect, neoplastic cells upregulate immunosuppressive receptors, including immune checkpoints such as CTLA-4 (cytotoxic T-lymphocyte-associated antigen 4) and PD-1/PD-L1 (programed death 1/programed death ligand 1) and effectively evade the host immune system [137]. Immunotherapy has become a promising new treatment in HNSCC by selectively targeting these checkpoints of immune function to allow tumor recognition. There are several immunotherapy agents targeting various aspects in the immune cascade, such as CTLA-4 and PD-1/PD-L1 inhibitors.

Nivolumab (Opdivo®), a monoclonal antibody directed against PD-L1, has been compared against single-agent chemotherapy in a phase III trial of refractory HNSCC. Results showed improved overall survival in patients treated with nivolumab when compared to standard therapy (7.5 months compared to 5.1 months, $P = 0.01$) [138]. Pembrolizumab (Keytruda®), another PD-1 directed antibody, was compared to standard of care treatment in the phase III clinical trial KEYNOTE-040. This randomized control trial showed a modest reduction in risk of death (19%) in the pembrolizumab arm but failed to meet the prespecified efficacy boundaries of the study [139]. Currently, pembrolizumab (Keytruda®) is FDA-approved for recurrent or metastatic head and neck SCC that has continued to progress despite standard-of-care treatment. Notably, in melanoma, favorable results have been observed in patients treated with a combination of PD-1 directed antibodies and CTLA-4 directed antibodies [140]. Studies are currently underway to evaluate similar combination strategies in advanced head and neck malignancies.

Adenovirus

Viral vector-mediated gene transfer has been investigated as a new experimental strategy to treat advanced and recurrent HNSCC. With this method, once a portion of the viral genome is replaced with the desired genetic sequence, the virus is injected into the tumor and allowed to infect the host cells. This results in propagation of the desired genetic sequence among the tumor cells. Such treatment technique is attractive for the delivery of tumor suppressor genes to restore those that have been inactivated. It is also a promising option because HNSCC tumors are often accessible for direct injection of gene therapy. As discussed, the *TP53* tumor suppressor gene is the most commonly mutated gene in HNSCC, and because wild-type TP53 protein functions to arrest aberrant cellular growth, *TP53* adenoviral gene therapy has been studied extensively. In in vitro models, transfection of wild-type *TP53* using adenoviral vectors resulted in high-efficiency expression of normal *TP53* protein as well as growth inhibition in tumor cell lines with homozygous deletion of *TP53* [141]. Additionally, preclinical in vivo studies using adenovirus containing wild-type *TP53* vector (Ad-p53) have shown successful induction of cancer cell apoptosis as well as enhanced response to chemoradiation treatment [142, 143]. A few phase II trials have demonstrated treatment with *TP53* vector adenovirus to be feasible and safe with some evidence of durable, albeit modest, activity in patients with HNSCC [103, 144]. In fact, ONYX-015, an adenovirus engineered to specifically target cells lacking *TP53* function, is approved for the management of early-stage HNSCC in China [145].

EGFR Inhibitors

The fact that *EGFR* is overexpressed in 80–90% of HNSCC and plays an important role in its pathogenesis offers a rationale for the development of *EGFR*-targeted therapy. Multiple monoclonal antibodies and small-molecule tyrosine kinase inhibitors (TKIs) have been developed and are under investigation. Cetuximab (™Erbitux) is a chimeric monoclonal antibody directed against EGFR and is the only FDA-approved targeted agent for use against HNSCC. Panitumumab (™Vectibix) is a "fully human" monoclonal antibody that is FDA-approved in colorectal cancer but not in HNSCC. Several recent clinical trials evaluating the utility of panitumumab in HNSCC showed no improvement in overall survival and only modest gains in progression free survival [140–148]. However, extensive clinical studies using cetuximab have demonstrated this agent to be particularly useful as an adjuvant to radiotherapy. In the multinational, randomized phase III trial that led to FDA approval of this agent, cetuximab combined with radiotherapy improved locoregional control and reduced mortality without increasing toxicity in patients with locoregionally advanced HNSCC [149]. More recent studies are showing increased benefit in progression-free survival as well as overall survival when adding cetuximab to platinum-based chemotherapy alone [150]. Despite the fact that *EGFR* is overexpressed in 80–90% of HNSCC, the cumulative data have only shown a marginal survival benefit with *EGFR*-targeted therapies, with treatment efficacy in a mere 20% of patients. At present, the mechanisms underlying the resistance to *EGFR*-targeted therapies are unknown and under investigation in hopes of improving clinical efficacy of this treatment strategy in HNSCC [151].

While monoclonal antibodies recognize a precise region in the extracellular ligand-binding domain of EGFR and therefore represent more specific EGFR targeting, small-molecule TKIs may cross-react with other kinases and lack specificity for EGFR. However, they have the advantage of being able to target multiple pathways involved in tumorigenesis as well as being conveniently dosed orally. Erlotinib and gefitinib are two of the most studied TKIs in HNSCC, but there is a lack of evidence to support their utility in HNSCC. In one randomized phase II study in patients with locally advanced HNSCC being treated with cisplatin and radiotherapy with or without erlotinib, the authors concluded that although erlotinib did not increase toxicity, there was no significant improvement in complete response rate or progression-free survival [152]. Two randomized, placebo-controlled phase III trials evaluating gefitinib [153] and lapatinib [154] failed to demonstrate improvement in outcomes in patients with recurrent or metastatic HNSCC. Another randomized, placebo-controlled phase II trial of gefitinib with chemoradiation therapy had similar disappointing findings [155]. Therefore, despite the convenient oral dosing of TKIs, the lack of positive phase III data limits their incorporation into the standard care for patients with HNSCC.

NOTCH Pathway Inhibitors

The appropriate therapeutic targeting of *NOTCH* will differ depending on whether the tumor contains *NOTCH* gain- or loss-of-function alterations. For tumors harboring activating *NOTCH* mutations which would lead it to function as an oncogene, a variety of GSIs are being investigated as a possible targeted strategy to inactive NOTCH signaling [61]. GSIs act by preventing NICD cleavage and nuclear translocation [156]. GSIs have shown promise in in vitro and in vivo studies of many solid tumors, including breast, lung, colorectal, and pancreatic cancers as well as melanoma and sarcoma [151–161]. Currently, there are several ongoing phase I and II clinical trials of GSIs in advanced solid tumors [61, 162, 163].

In tumor systems where *NOTCH* acts as a tumor suppressor gene and therefore is inactivated during oncogenesis, the appropriate strategy would be to activate the *NOTCH* signaling pathway. The *NOTCH* pathway is frequently silenced by epigenetic changes, and histone deacetylase (HDAC) inhibitors are under investigation to restore *NOTCH* signaling in cancers. Valproic acid, an HDAC inhibitor, is in clinical development and being studied in many ongoing phase I and II clinical trials in solid tumors [61]. Of note, it is important to recognize that administrating a systemic therapy that inhibits general *NOTCH* signaling may be complicated by loss of *NOTCH* tumor suppressor function in non-tumorous sites, thus potentially inducing secondary malignancy. Likewise, activating *NOTCH* signaling systemically to target a given tumor lineage may lead to activation of *NOTCH* signaling in a normal cell where *NOTCH* may act as an oncogene. Therefore, a better understanding of the exact *NOTCH* signaling pathway alterations within a cancer-specific context is necessary to develop appropriate *NOTCH*-targeting therapeutics.

Conclusions

In this chapter, we reviewed some of the current understanding of the molecular biology of HNSCC and discussed some of the diagnostic, prognostic, and therapeutic clinical implications. The advent of genomic technologies has greatly advanced our knowledge of the molecular changes underlying HNSCC, and the knowledge gained offers new promise for the treatment of this cancer. The hope is that the novel approaches will ultimately result in improved patient outcome through the development of new diagnostic and prognostic indicators as well as new targeted therapies for HNSCC patients.

References

1. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010;363:24–35.
2. Huang SH, Xu W, Waldron J, et al. Refining American Joint Committee on Cancer/Union for International Cancer Control TNM Stage and Prognostic Groups for human papillomavirus–related oropharyngeal carcinomas. J Clin Oncol. 2015;33:836–45.
3. Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. Mayo Clin Proc. 2008;83:489–501.
4. Pelucchi C, Gallus S, Garavello W, Bosetti C, La Vecchia C. Alcohol and tobacco use, and cancer risk for upper aerodigestive tract and liver. Eur J Cancer Prev. 2008;17:340–4.
5. Hashibe M, Brennan P, Benhamou S, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst. 2007;99:777–89.
6. Franceschi S, Talamini R, Barra S, Barón AE, Negri E, Bidoli E, Serraino D, Vecchia CL. Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in Northern Italy. Cancer Res. 1990;50:6502–7.
7. Zheng T, Boyle P, Hu H, Duan J, Jiang P, Ma D, Shui L, Niu S, MacMahon B. Tobacco smoking, alcohol consumption, and risk of oral cancer: a case-control study in Beijing, People's Republic of China. Cancer Causes Control. 1990;1:173–9.
8. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, Bernstein L, Schoenberg JB, Stemhagen A, Fraumeni JF. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res. 1988;48:3282–7.
9. Talamini R, Favero A, Franceschi S, La Vecchia C, Levi F, Conti E. Cancer of the oral cavity and pharynx in nonsmokers who drink alcohol and in nondrinkers who smoke tobacco. J Natl Cancer Inst. 1998;90:1901–3.
10. Wight AJ, Ogden GR. Possible mechanisms by which alcohol may influence the development of oral cancer—a review. Oral Oncol. 1998;34:441–7.

11. WHO. Cancer. In: WHO. http://www.who.int/mediacentre/factsheets/fs297/en/. Accessed 6 Apr 2017.
12. Forouzanfar MH, Afshin A, Alexander LT, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388:1659–724.
13. Hoffmann D, Hoffmann I, El-Bayoumy K. The less harmful cigarette: a controversial issue. A tribute to Ernst L. Wynder. Chem Res Toxicol. 2001;14:767–90.
14. Arora A, Willhite CA, Liebler DC. Interactions of β-carotene and cigarette smoke in human bronchial epithelial cells. Carcinogenesis. 2001;22:1173–8.
15. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999;91:1194–210.
16. Miller JA. Recent studies on the metabolic activation of chemical carcinogens. Cancer Res. 1994;54:1879s–81s.
17. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21:7435–51.
18. Feng J, Li L, Zhao Y-S, Tang S-Q, Yang H-B, Liu S-X. Interaction between CYP 2C19*3 polymorphism and smoking in relation to laryngeal carcinoma in the Chinese Han population. Genet Mol Res. 2011;10:3331–7.
19. XIE S, LUO C, SHAN X, ZHAO S, HE J, CAI Z. CYP1A1 MspI polymorphism and the risk of oral squamous cell carcinoma: evidence from a meta-analysis. Mol Clin Oncol. 2016;4:660–6.
20. Hashibe M, Brennan P, Chuang S, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev. 2009;18:541–50.
21. Boffetta P, Hashibe M. Alcohol and cancer. Lancet Oncol. 2006;7:149–56.
22. Matsuda T, Yabushita H, Kanaly RA, Shibutani S, Yokoyama A. Increased DNA damage in ALDH2-deficient alcoholics. Chem Res Toxicol. 2006;19:1374–8.
23. Druesne-Pecollo N, Tehard B, Mallet Y, Gerber M, Norat T, Hercberg S, Latino-Martel P. Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. Lancet Oncol. 2009;10:173–80.
24. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJF, Meijer CJLM. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003;348:518–27.
25. Gillison ML, Broutian T, Pickard RKL, Tong Z, Xiao W, Kahle L, Graubard BI, Chaturvedi AK. Prevalence of oral HPV infection in the United States, 2009-2010. JAMA. 2012;307:693–703.
26. Mork J, Lie AK, Glattre E, et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. N Engl J Med. 2001;344:1125–31.
27. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011;29:4294–301.
28. Cancer Facts & Figures 2016. American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html. Accessed 7 Apr 2017.
29. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517:576–82.
30. Brennan JA, Boyle JO, Koch WM, Goodman SN, Hruban RH, Eby YJ, Couch MJ, Forastiere AA, Sidransky D. Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck. N Engl J Med. 1995;332:712–7.
31. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011;333:1154–7.
32. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011;333:1157–60.
33. Haupt S, Berger M, Goldberg Z, Haupt Y. Apoptosis – the p53 network. J Cell Sci. 2003;116:4077–85.
34. Muller PAJ, Vousden KH. p53 mutations in cancer. Nat Cell Biol. 2013;15:2–8.
35. El-Naggar AK, Lai S, Luna MA, Zhou X-D, Weber RS, Goepfert H, Batsakis JG. Sequential p53 mutation analysis of pre-invasive and invasive head and neck squamous carcinoma. Int J Cancer. 1995;64:196–201.
36. Ogmundsdóttir HM, Hilmarsdóttir H, Astvaldsdóttir A, Jóhannsson JH, Holbrook WP. Oral lichen planus has a high rate of TP53 mutations. A study of oral mucosa in icelanD. Eur J Oral Sci. 2002;110:192–8.
37. Boyle JO, Hakim J, Koch W, van der Riet P, Hruban RH, Roa RA, Correo R, Eby YJ, Ruppert JM, Sidransky D. The incidence of p53 mutations increases with progression of head and neck cancer. Cancer Res. 1993;53:4477–80.
38. Field JK, Zoumpourlis V, Spandidos DA, Jones AS (1994) p53 expression and mutations in squamous cell carcinoma of the head and neck: expression correlates with the patients' use of tobacco and alcohol. Cancer Detect Prev 18:197–208.
39. Hsieh L-L, Wang P-F, Chen I-H, Liao C-T, Wang H-M, Chen M-C, Chang JT-C, Cheng A-J. Characteristics of mutations in the p53 gene in oral squamous cell carcinoma associated with betel quid chewing and cigarette smoking in Taiwanese. Carcinogenesis. 2001;22:1497–503.
40. Tanoue T, Takeichi M. New insights into Fat cadherins. J Cell Sci. 2005;118:2347–53.
41. Tanoue T, Takeichi M. Mammalian Fat1 cadherin regulates actin dynamics and cell–cell contact. J Cell Biol. 2004;165:517–28.
42. Bryant PJ, Huettner B, Held LI, Ryerse J, Szidonya J. Mutations at the fat locus interfere with cell proliferation control and epithelial morphogenesis in Drosophila. Dev Biol. 1988;129:541–54.
43. Mahoney PA, Weber U, Onofrechuk P, Biessmann H, Bryant PJ, Goodman CS. The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily. Cell. 1991;67:853–68.
44. Mao Y, Mulvaney J, Zakaria S, Yu T, Morgan KM, Allen S, Basson MA, Francis-West P, Irvine KD. Characterization of a Dchs1 mutant mouse reveals requirements for Dchs1-Fat4 signaling during mammalian development. Development. 2011;138:947–57.
45. Saburi S, Hester I, Goodrich L, McNeill H. Functional interactions between Fat family cadherins in tissue morphogenesis and planar polarity. Development. 2012;139:1806–20.
46. Ishiuchi T, Misaki K, Yonemura S, Takeichi M, Tanoue T. Mammalian Fat and Dachsous cadherins regulate apical membrane organization in the embryonic cerebral cortex. J Cell Biol. 2009;185:959–67.
47. Morris LGT, Kaufman AM, Gong Y, et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat Genet. 2013;45:253–61.
48. Behrens J, von Kries JP, Kühl M, Bruhn L, Wedlich D, Grosschedl R, Birchmeier W. Functional interaction of beta-catenin with the transcription factor LEF-1. Nature. 1996;382:638–42.
49. Clevers H. Wnt/beta-catenin signaling in development and disease. Cell. 2006;127:469–80.
50. Lustig B, Behrens J. The Wnt signaling pathway and its role in tumor development. J Cancer Res Clin Oncol. 2003;129:199–221.
51. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature. 2005;434:843–50.

52. Kim KT, Kim B, Kim JH. Association between FAT1 mutation and overall survival in patients with human papillomavirus–negative head and neck squamous cell carcinoma. Head Neck. 2016;38:E2021–9.
53. Bowles DW, Diamond JR, Lam ET, et al. Phase I study of oral rigosertib (ON 01910.Na), a dual inhibitor of the PI3K and Plk1 pathways, in adult patients with advanced solid malignancies. Clin Cancer Res. 2014;20:1656–65.
54. Califano J, van der Riet P, Westra W, et al. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res. 1996;56:2488–92.
55. van der Riet P, Nawroz H, Hruban RH, Corio R, Tokino K, Koch W, Sidransky D. Frequent loss of chromosome 9p21-22 early in head and neck cancer progression. Cancer Res. 1994;54:1156–8.
56. Miracca EC, Kowalski LP, Nagai MA. High prevalence of p16 genetic alterations in head and neck tumours. Br J Cancer. 1999;81:677–83.
57. Chen P-L, Scully P, Shew J-Y, Wang JYJ, Lee W-H. Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. Cell. 1989;58:1193–8.
58. Lukas J, Parry D, Aagaard L, Mann DJ, Bartkova J, Strauss M, Peters G, Bartek J. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. Nature. 1995;375:503–6.
59. Nicolas M, Wolfer A, Raj K, Kummer JA, Mill P, van Noort M, Hui C, Clevers H, Dotto GP, Radtke F. Notch1 functions as a tumor suppressor in mouse skin. Nat Genet. 2003;33:416–21.
60. Proweller A, Tu L, Lepore JJ, Cheng L, Lu MM, Seykora J, Millar SE, Pear WS, Parmacek MS. Impaired Notch signaling promotes de novo squamous cell carcinoma formation. Cancer Res. 2006;66:7438–44.
61. Egloff AM, Grandis JR. Molecular pathways: context-dependent approaches to Notch targeting as cancer therapy. Clin Cancer Res. 2012;18:5188–95.
62. Weng AP, Ferrando AA, Lee W, Morris JP, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, Aster JC. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science. 2004;306:269–71.
63. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475:101–5.
64. Zhang Z-P, Sun Y-L, Fu L, Gu F, Zhang L, Hao X-S. Correlation of Notch1 expression and activation to cisplatin-sensitivity of head and neck squamous cell carcinoma. Ai Zheng. 2009;28:100–3.
65. Lin J-T, Chen M-K, Yeh K-T, Chang C-S, Chang T-H, Lin C-Y, Wu Y-C, Su B-W, Lee K-D, Chang P-J. Association of high levels of Jagged-1 and Notch-1 expression with poor prognosis in head and neck cancer. Ann Surg Oncol. 2010;17:2976–83.
66. Gu F, Ma Y, Zhang Z, Zhao J, Kobayashi H, Zhang L, Fu L. Expression of Stat3 and Notch1 is associated with cisplatin resistance in head and neck squamous cell carcinoma. Oncol Rep. 2010;23:671–6.
67. Zhang T-H, Liu H-C, Zhu L-J, Chu M, Liang Y-J, Liang L-Z, Liao G-Q. Activation of Notch signaling in human tongue carcinoma. J Oral Pathol Med. 2011;40:37–45.
68. Hijioka H, Setoguchi T, Miyawaki A, Gao H, Ishida T, Komiya S, Nakamura N. Upregulation of Notch pathway molecules in oral squamous cell carcinoma. Int J Oncol. 2010;36:817–22.
69. Kalyankrishna S, Grandis JR. Epidermal growth factor receptor biology in head and neck Cancer. J Clin Oncol. 2006;24:2666–72.
70. Klein JD, Grandis JR. The molecular pathogenesis of head and neck cancer. Cancer Biol Ther. 2010;9:1.
71. Rodrigo JP, Ramos S, Lazo PS, Alvarez I, Suárez C (1996) Amplification of ERBB oncogenes in squamous cell carcinomas of the head and neck. Eur J Cancer 1990 32A:2004–2010.
72. Ibrahim SO, Vasstrand EN, Liavaag PG, Johannessen AC, Lillehaug JR. Expression of c-erbB proto-oncogene family members in squamous cell carcinoma of the head and neck. Anticancer Res. 1997;17:4539–46.
73. Sok JC, Coppelli FM, Thomas SM, et al. Mutant epidermal growth factor receptor (EGFRvIII) contributes to head and neck cancer growth and resistance to EGFR targeting. Clin Cancer Res. 2006;12:5064–73.
74. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. J Cell Sci. 2016;129:1287–92.
75. Rogers SJ, Harrington KJ, Rhys-Evans P, O-Charoenrat P, Eccles SA. Biological significance of c-erbB family oncogenes in head and neck cancer. Cancer Metastasis Rev. 2005;24:47–69.
76. Leslie NR, Downes CP. PTEN function: how normal cells control it and tumour cells lose it. Biochem J. 2004;382:1–11.
77. Pedrero JMG, Carracedo DG, Pinto CM, Zapatero AH, Rodrigo JP, Nieto CS, Gonzalez MV. Frequent genetic and biochemical alterations of the PI 3-K/AKT/PTEN pathway in head and neck squamous cell carcinoma. Int J Cancer. 2005;114:242–8.
78. Sano D, Oridate N. The molecular mechanism of human papillomavirus-induced carcinogenesis in head and neck squamous cell carcinoma. Int J Clin Oncol. 2016;21:819–26.
79. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell. 1990;63:1129–36.
80. Münger K, Baldwin A, Edwards KM, Hayakawa H, Nguyen CL, Owens M, Grace M, Huh K. Mechanisms of human papillomavirus-induced oncogenesis. J Virol. 2004;78:11451–60.
81. Straight SW, Hinkle PM, Jewers RJ, McCance DJ. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. J Virol. 1993;67:4521–32.
82. Faraji F, Zaidi M, Fakhry C, Gaykalova DA. Molecular mechanisms of human papillomavirus-related carcinogenesis in head and neck cancer. Microbes Infect. 2017;19:464–75.
83. Venuti A, Paolini F. HPV detection methods in head and neck cancer. Head Neck Pathol. 2012;6:63–74.
84. Singhi AD, Westra WH. Comparison of human papillomavirus in situ hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. Cancer. 2010;116:2166–73.
85. Huang CC, Qiu JT, Kashima ML, Kurman RJ, Wu TC. Generation of type-specific probes for the detection of single-copy human papillomavirus by a novel in situ hybridization method. Mod Pathol. 1998;11:971–7.
86. Capone RB, Pai SI, Koch WM, Gillison ML, Danish HN, Westra WH, Daniel R, Shah KV, Sidransky D. Detection and quantitation of human papillomavirus (HPV) DNA in the sera of patients with HPV-associated head and neck squamous cell carcinoma. Clin Cancer Res. 2000;6:4171–5.
87. Chuang AY, Chuang TC, Chang S, Zhou S, Begum S, Westra WH, Ha PK, Koch WM, Califano JA. Presence of HPV DNA in convalescent salivary rinses is an adverse prognostic marker in head and neck squamous cell carcinoma. Oral Oncol. 2008;44:915–9.
88. Conway C, Chalkley R, High A, et al. Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. J Mol Diagn. 2012;14:104–11.
89. Barzon L, Militello V, Lavezzo E, et al. Human papillomavirus genotyping by 454 next generation sequencing technology. J Clin Virol. 2011;52:93–7.
90. Wittekindt C, Gültekin E, Weissenborn SJ, Dienes HP, Pfister HJ, Klussmann JP. Expression of p16 protein is associated with human papillomavirus status in tonsillar carcinomas and has implications on survival. Adv Otorhinolaryngol. 2005;62:72–80.
91. Shi W, Kato H, Perez-Ordonez B, et al. Comparative prognostic value of HPV16 E6 mRNA compared with in situ hybridization

for human oropharyngeal squamous carcinoma. J Clin Oncol. 2009;27:6213–21.
92. Begum S, Gillison ML, Nicol TL, Westra WH. Detection of human Papillomavirus-16 in fine-needle aspirates to determine tumor origin in patients with metastatic squamous cell carcinoma of the head and neck. Clin Cancer Res. 2007;13:1186–91.
93. Umudum H, Rezanko T, Dag F, Dogruluk T. Human papillomavirus genome detection by in situ hybridization in fine-needle aspirates of metastatic lesions from head and neck squamous cell carcinomas. Cancer Cytopathol. 2005;105:171–7.
94. de Braud F, al-Sarraf M. Diagnosis and management of squamous cell carcinoma of unknown primary tumor site of the neck. Semin Oncol. 1993;20:273–8.
95. Day KE, Sweeny L, Kulbersh B, Zinn KR, Rosenthal EL. Preclinical comparison of near-infrared-labeled cetuximab and panitumumab for optical imaging of head and neck squamous cell carcinoma. Mol Imaging Biol. 2013;15:722–9.
96. Heath CH, Deep NL, Sweeny L, Zinn KR, Rosenthal EL. Use of Panitumumab-IRDye800 to image microscopic head and neck Cancer in an Orthotopic surgical model. Ann Surg Oncol. 2012;19:3879–87.
97. de Boer E, Warram JM, Tucker MD, et al. In vivo fluorescence immunohistochemistry: localization of fluorescently labeled cetuximab in squamous cell carcinomas. Sci Rep. 2015;5:10169. https://doi.org/10.1038/srep10169.
98. Qi S, Miao Z, Liu H, Xu Y, Feng Y, Cheng Z. Evaluation of four Affibody-based near-infrared fluorescent probes for optical imaging of epidermal growth factor receptor positive tumors. Bioconjug Chem. 2012;23:1149–56.
99. Agnes RS, Broome A-M, Wang J, Verma A, Lavik K, Basilion JP. An optical probe for noninvasive molecular imaging of orthotopic brain tumors overexpressing epidermal growth factor receptor. Mol Cancer Ther. 2012;11:2202. https://doi.org/10.1158/1535-7163.MCT-12-0211.
100. Chan LW, Wang Y-N, Lin LY, Upton MP, Hwang JH, Pun SH. Synthesis and characterization of anti-EGFR fluorescent nanoparticles for optical molecular imaging. Bioconjug Chem. 2013;24:167–75.
101. Oliveira S, van Dongen GAMS, Stigter-van Walsum M, Roovers RC, Stam JC, Mali W, van Diest PJ, van Bergen en Henegouwen PMP. Rapid visualization of human tumor xenografts through optical imaging with a near-infrared fluorescent anti-epidermal growth factor receptor nanobody. Mol Imaging. 2012;11:33–46.
102. Poeta ML, Manola J, Goldwasser MA, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. N Engl J Med. 2007;357:2552–61.
103. Ganly I, Kirn D, Eckhardt SG, et al. A phase I study of Onyx-015, an E1B attenuated adenovirus, administered intratumorally to patients with recurrent head and neck cancer. Clin Cancer Res. 2000;6:798–806.
104. Skinner HD, Sandulache VC, Ow TJ, Meyn RE, Yordy JS, Beadle BM, Fitzgerald AL, Giri U, Ang KK, Myers JN. TP53 disruptive mutations lead to head and neck cancer treatment failure through inhibition of radiation-induced senescence. Clin Cancer Res. 2012;18:290–300.
105. Koch WM, Brennan JA, Zahurak M, Goodman SN, Westra WH, Schwab D, Yoo GH, Lee DJ, Forastiere AA, Sidransky D. p53 mutation and locoregional treatment failure in head and neck squamous cell carcinoma. J Natl Cancer Inst. 1996;88:1580–6.
106. Alsner J, Sørensen SB, Overgaard J. TP53 mutation is related to poor prognosis after radiotherapy, but not surgery, in squamous cell carcinoma of the head and neck. Radiother Oncol. 2001;59:179–85.
107. Temam S, Flahault A, Périé S, Monceaux G, Coulet F, Callard P, Bernaudin J-F, St Guily JL, Fouret P. p53 gene status as a predictor of tumor response to induction chemotherapy of patients with locoregionally advanced squamous cell carcinomas of the head and neck. J Clin Oncol. 2000;18:385.
108. Cabelguenne A, Blons H, de Waziers I, Carnot F, Houllier A-M, Soussi T, Brasnu D, Beaune P, Laccourreye O, Laurent-Puig P. p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma: a prospective series. J Clin Oncol. 2000;18:1465–73.
109. Brennan JA, Mao L, Hruban RH, Boyle JO, Eby YJ, Koch WM, Goodman SN, Sidransky D. Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. N Engl J Med. 1995;332:429–35.
110. van HVMM, Leemans CR, Kummer JA, Dijkstra J, Kuik DJ, van den BMWM, Snow GB, Brakenhoff RH. Molecular diagnosis of surgical margins and local recurrence in head and neck cancer patients. Clin Cancer Res. 2004;10:3614–20.
111. Pena Murillo C, Huang X, Hills A, et al. The utility of molecular diagnostics to predict recurrence of head and neck carcinoma. Br J Cancer. 2012;107:1138–43.
112. Lopez-Martinez M, Anzola M, Cuevas N, Aguirre JM, De-Pancorbo M. Clinical applications of the diagnosis of p53 alterations in squamous cell carcinoma of the head and neck. Med Oral. 2002;7:108–20.
113. Ragin CCR, Taioli E. Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: review and meta-analysis. Int J Cancer. 2007;121:1813–20.
114. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, Forastiere A, Gillison ML. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst. 2008;100:261–9.
115. Sedaghat AR, Zhang Z, Begum S, et al. Prognostic significance of human papillomavirus in oropharyngeal squamous cell carcinomas. Laryngoscope. 2009;119:1542–9.
116. Dayyani F, Etzel CJ, Liu M, Ho C-H, Lippman SM, Tsao AS. Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). Head Neck Oncol. 2010;2:15.
117. Geisler SA, Olshan AF, Cai J, Weissler M, Smith J, Bell D. Glutathione S-transferase polymorphisms and survival from head and neck cancer. Head Neck. 2005;27:232–42.
118. Nishimura T, Newkirk K, Sessions RB, Andrews PA, Trock BJ, Rasmussen AA, Montgomery EA, Bischoff EK, Cullen KJ. Immunohistochemical staining for glutathione S-transferase predicts response to platinum-based chemotherapy in head and neck cancer. Clin Cancer Res. 1996;2:1859–65.
119. Shiga H, Heath EI, Rasmussen AA, Trock B, Johnston PG, Forastiere AA, Langmacher M, Baylor A, Lee M, Cullen KJ. Prognostic value of p53, glutathione S-transferase π, and thymidylate synthase for neoadjuvant cisplatin-based chemotherapy in head and neck cancer. Clin Cancer Res. 1999;5:4097–104.
120. Schumaker L, Nikitakis N, Goloubeva O, Tan M, Taylor R, Cullen KJ. Elevated expression of glutathione S-transferase π and p53 confers poor prognosis in head and neck cancer patients treated with chemoradiotherapy but not radiotherapy alone. Clin Cancer Res. 2008;14:5877–83.
121. Szanyi I, Ráth G, Móricz P, Somogyvári K, Révész P, Gerlinger I, Orsós Z, Ember I, Kiss I. Effects of cytochrome P450 1A1 and uridine-diphosphate-glucuronosyltransferase 1A1 allelic polymorphisms on the risk of development and the prognosis of head and neck cancers. Eur J Cancer. 2012;21:560–8.
122. Sturgis EM, Ang KK. The epidemic of HPV-associated oropharyngeal cancer is here: is it time to change our treatment paradigms? J Natl Compr Cancer Netw. 2011;9:665–73.
123. Marur S, Li S, Cmelak AJ, et al. E1308: phase II trial of induction chemotherapy followed by reduced-dose radiation and weekly cetuximab in patients with HPV-associated resectable squamous

cell carcinoma of the oropharynx— ECOG-ACRIN Cancer Research Group. J Clin Oncol. 2016;35:490–7.
124. Moore EJ, Olsen KD, Kasperbauer JL. Transoral robotic surgery for oropharyngeal squamous cell carcinoma: a prospective study of feasibility and functional outcomes. Laryngoscope. 2009;119:2156–64.
125. Weinstein GS, Quon H, O'Malley BW, Kim GG, Cohen MA. Selective neck dissection and deintensified postoperative radiation and chemotherapy for oropharyngeal cancer: a subset analysis of the university of Pennsylvania transoral robotic surgery trial. Laryngoscope. 2010;120:1749–55.
126. White HN, Moore EJ, Rosenthal EL, Carroll WR, Olsen KD, Desmond RA, Magnuson JS. Transoral robotic-assisted surgery for head and neck squamous cell carcinoma: one- and 2-year survival analysis. Arch Otolaryngol Head Neck Surg. 2010;136:1248–52.
127. Cohen MA, Weinstein GS, O'Malley BW, Feldman M, Quon H. Transoral robotic surgery and human papillomavirus status: oncologic results. Head Neck. 2011;33:573–80.
128. Moore EJ, Olsen SM, Laborde RR, García JJ, Walsh FJ, Price DL, Janus JR, Kasperbauer JL, Olsen KD. Long-term functional and oncologic results of Transoral robotic surgery for oropharyngeal squamous cell carcinoma. Mayo Clin Proc. 2012;87:219–25.
129. Schiller JT, Castellsagué X, Garland SM. A review of clinical trials of human papillomavirus prophylactic vaccines. Vaccine. 2012;30:F123–38.
130. Joura EA, Giuliano AR, Iversen O-E, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. N Engl J Med. 2015;372:711–23.
131. Joura EA, Garland SM, Paavonen J, Ferris DG, Perez G, Ault KA, Huh WK, Sings HL, James MK, Haupt RM. Effect of the human papillomavirus (HPV) quadrivalent vaccine in a subgroup of women with cervical and vulvar disease: retrospective pooled analysis of trial data. BMJ. 2012;344:e1401. https://doi.org/10.1136/bmj.e1401.
132. FUTURE I/II Study Group, Dillner J, Kjaer SK, Wheeler CM, Sigurdsson K, Iversen OE, Hernandez-Avila M, Perez G, Brown DR, Koutsky LA, Tay EH, García P, Ault KA, Garland SM, Leodolter S, Olsson SE, Tang GW, Ferris DG, Paavonen J, Lehtinen M, Steben M, Bosch FX, Joura EA, Majewski S, Muñoz N, Myers ER, Villa LL, Taddeo FJ, Roberts C, Tadesse A, Bryan JT, Maansson R, Lu S, Vuocolo S, Hesley TM, Barr E, Haupt R. Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. BMJ. 2010;341:c3493. https://doi.org/10.1136/bmj.c3493.
133. Yang A, Farmer E, Lin J, Wu T-C, Hung C-F. The current state of therapeutic and T cell-based vaccines against human papillomaviruses. Virus Res. 2017;231:148–65.
134. Best SR, Niparko KJ, Pai SI. Biology of HPV infection and immune therapy for HPV-related head and neck cancers. Otolaryngol Clin N Am. 2012;45:807–22.
135. Duray A, Demoulin S, Hubert P, Delvenne P, Saussez S. Immune suppression in head and neck cancers: a review. Clin Dev Immunol. 2010;2010:1. https://doi.org/10.1155/2010/701657.
136. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science. 2015;348:69–74.
137. Mandal R, Şenbabaoğlu Y, Desrichard A, et al. The head and neck cancer immune landscape and its immunotherapeutic implications. JCI Insight. 1:e89829. https://doi.org/10.1172/jci.insight.89829.
138. Ferris RL, Blumenschein GJ, Fayette J, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med. 2016;375:1856–67.
139. Cohen EE, Harrington KJ, Le Tourneau C, et al. LBA45_PRPembrolizumab (pembro) vs standard of care (SOC) for recurrent or metastatic head and neck squamous cell carcinoma (R/M HNSCC): Phase 3 KEYNOTE-040 trial. Ann Oncol. 2017; https://doi.org/10.1093/annonc/mdx440.040.
140. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med. 2015;373:23–34.
141. Zhang WW, Fang X, Mazur W, French BA, Georges RN, Roth JA. High-efficiency gene transfer and high-level expression of wild-type p53 in human lung cancer cells mediated by recombinant adenovirus. Cancer Gene Ther. 1994;1:5–13.
142. Liu T-J, Zhang W-W, Taylor DL, Roth JA, Goepfert H, Clayman GL. Growth suppression of human head and neck Cancer cells by the introduction of a wild-type p53 gene via a recombinant adenovirus. Cancer Res. 1994;54:3662–7.
143. Pirollo KF, Hao Z, Rait A, Jang YJ, Fee WE, Ryan P, Chiang Y, Chang EH. p53 mediated sensitization of squamous cell carcinoma of the head and neck to radiotherapy. Oncogene. 1997;14:1735–46.
144. Nemunaitis J, Clayman G, Agarwala SS, et al. Biomarkers predict p53 gene therapy efficacy in recurrent squamous cell carcinoma of the head and neck. Clin Cancer Res. 2009;15:7719–25.
145. Nemunaitis J, Nemunaitis J. Head and neck cancer: response to p53-based therapeutics. Head Neck. 2011;33:131–4.
146. Vermorken JB, Stöhlmacher-Williams J, Davidenko I, et al. Cisplatin and fluorouracil with or without panitumumab in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck (SPECTRUM): an open-label phase 3 randomised trial. Lancet Oncol. 2013;14:697–710.
147. Mesía R, Henke M, Fortin A, et al. Chemoradiotherapy with or without panitumumab in patients with unresected, locally advanced squamous-cell carcinoma of the head and neck (CONCERT-1): a randomised, controlled, open-label phase 2 trial. Lancet Oncol. 2015;16:208–20.
148. Giralt J, Trigo J, Nuyts S, et al. Panitumumab plus radiotherapy versus chemoradiotherapy in patients with unresected, locally advanced squamous-cell carcinoma of the head and neck (CONCERT-2): a randomised, controlled, open-label phase 2 trial. Lancet Oncol. 2015;16:221–32.
149. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med. 2006;354:567–78.
150. Vermorken JB, Mesia R, Rivera F, et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med. 2008;359:1116–27.
151. Cohen RB. Current challenges and clinical investigations of epidermal growth factor receptor (EGFR)- and ErbB family-targeted agents in the treatment of head and neck squamous cell carcinoma (HNSCC). Cancer Treat Rev. 2014;40:567–77.
152. Martins RG, Parvathaneni U, Bauman JE, et al. Cisplatin and radiotherapy with or without erlotinib in locally advanced squamous cell carcinoma of the head and neck: a randomized phase II trial. J Clin Oncol. 2013;31:1415–21.
153. Argiris A, Ghebremichael M, Gilbert J, Lee J-W, Sachidanandam K, Kolesar JM, Burtness B, Forastiere AA. Phase III randomized, placebo-controlled trial of docetaxel with or without Gefitinib in recurrent or metastatic head and neck Cancer: an eastern cooperative oncology group trial. J Clin Oncol. 2013;31:1405–14.
154. Harrington K, Temam S, Mehanna H, et al. Postoperative adjuvant Lapatinib and concurrent chemoradiotherapy followed by maintenance lapatinib monotherapy in high-risk patients with resected squamous cell carcinoma of the head and neck: a phase III, randomized, double-blind, placebo-controlled study. J Clin Oncol. 2015;33:4202–9.
155. Gregoire V, Hamoir M, Chen C, et al. Gefitinib plus cisplatin and radiotherapy in previously untreated head and neck squamous cell carcinoma: a phase II, randomized, double-blind, placebo-controlled study. Radiother Oncol. 2011;100:62–9.

156. Fortini ME. Gamma-secretase-mediated proteolysis in cell-surface-receptor signalling. Nat Rev Mol Cell Biol. 2002;3:673–84.
157. Curry CL, Reed LL, Golde TE, Miele L, Nickoloff BJ, Foreman KE. Gamma secretase inhibitor blocks Notch activation and induces apoptosis in Kaposi's sarcoma tumor cells. Oncogene. 2005;24:6333–44.
158. Huynh C, Poliseno L, Segura MF, et al. The novel gamma secretase inhibitor RO4929097 reduces the tumor initiating potential of melanoma. PLoS One. 2011;6:e25264. https://doi.org/10.1371/journal.pone.0025264.
159. Maraver A, Fernández-Marcos PJ, Herranz D, et al. Therapeutic effect of γ-secretase inhibition in KrasG12V-driven non-small cell lung carcinoma through derepression of DUSP1 phosphatase and inhibition of ERK. Cancer Cell. 2012;22:222. https://doi.org/10.1016/j.ccr.2012.06.014.
160. Portanova P, Notaro A, Pellerito O, Sabella S, Giuliano M, Calvaruso G. Notch inhibition restores TRAIL-mediated apoptosis via AP1-dependent upregulation of DR4 and DR5 TRAIL receptors in MDA-MB-231 breast cancer cells. Int J Oncol. 2013;43:121–30.
161. Yabuuchi S, Pai SG, Campbell NR, et al. Notch signaling pathway targeted therapy suppresses tumor progression and metastatic spread in pancreatic cancer. Cancer Lett. 2013;335:41–51.
162. Tolcher AW, Messersmith WA, Mikulski SM, et al. Phase I study of RO4929097, a gamma secretase inhibitor of Notch signaling, in patients with refractory metastatic or locally advanced solid tumors. J Clin Oncol. 2012;30:2348–53.
163. Richter S, Bedard PL, Chen EX, et al. A phase I study of the oral gamma secretase inhibitor R04929097 in combination with gemcitabine in patients with advanced solid tumors (PHL-078/CTEP 8575). Investig New Drugs. 2014;32:243–9.

Genomic Applications in Thyroid Cancer

24

Thomas J. Giordano

Introduction

All cancers arise from the same pathogenetic processes involving ongoing acquisition of genetic variation, mostly through random mutations in DNA and natural selection that acts on the ensuing diversity in phenotype. Despite the body's various defenses against transformed cancer-causing cells, rare cells accumulate enough advantageous mutations that permit their survival and increased cellular proliferation, resulting clinically in a tumor. These and other properties bestow upon these cells the ability to invade adjacent tissues and metastasize distantly, the pathologic hallmarks of cancer [1, 2]. Accordingly, much effort in cancer research over the last few decades has been directed at identifying these genetic alterations and understanding their consequences on cellular function. These endeavors have been greatly accelerated due to phenomenal advances in genomic technologies such as next-generation DNA sequencing [3]. As a consequence and through the coordinated efforts of networks of investigators such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), the last decade has witnessed a rapid characterization of the genome-wide alterations that occur in the most common types of cancers [4], including cancers derived from follicular thyroid cells.

The thyroid gland contains two endocrine cell types, and thyroid cancers are appropriately classified accordingly. Tumors derived from thyroid follicular cells represent the vast majority of thyroid neoplasms, which can be benign (i.e., follicular adenomas) or possess malignant potential (i.e., carcinomas). Conversely, tumors derived from thyroid neuroendocrine cells, termed parafollicular or C cells, are neuroendocrine tumors termed medullary carcinoma. Given the relatively rare nature of medullary carcinoma, almost all of the genome characterization efforts to data efforts have been directed at follicular cell thyroid cancers [5].

Follicular thyroid neoplasms are classified according to a simple taxonomic scheme based on traditional histopathological parameters. Tumors composed of follicular cells that lack malignant potential are diagnosed as follicular thyroid adenoma (FTA). Given their benign behavior, FTAs have not yet been genomically characterized to any significant degree. Papillary thyroid carcinoma (PTC) represents the most common thyroid cancer and accordingly was the tumor type selected by TCGA for their project on thyroid cancer (see below). Beyond PTC, other highly differentiated thyroid cancers include follicular thyroid carcinoma (FTC) and Hurthle cell carcinoma (HCC). Some genetic information is available for these tumors, but multidimensional genomic characterization efforts of FTC and HCC are ongoing. Collectively, PTC, FTC, and HCC have been grouped together as differentiated thyroid cancers because they retain significant follicular cell differentiation. Beyond these differentiated thyroid cancers, anaplastic thyroid cancer (ATC) represents a clinically aggressive form of thyroid cancer in which the tumor cells have lost their thyroid differentiation. Accordingly, ATCs are synonymously termed undifferentiated thyroid cancer. Poorly differentiated thyroid carcinoma (PDCA) refers to intermediate tumors with reduced thyroid differentiation and clinically and pathologically fall between differentiated and undifferentiated carcinomas. Given their aggressive nature, PCDA and ATC have been the subject of significant genomic investigation, although additional studies are ongoing. In this chapter, the essential information about the genetics and genomics of these thyroid cancers is presented in detail and summarized in Table 24.1.

Papillary Carcinoma

Papillary thyroid carcinoma (PTC) was genomically characterized by the TCGA Research Network as 1 of their 23 common cancer projects, in addition to their 10 rare cancer

T. J. Giordano (✉)
Divisions of Anatomic Pathology and Molecular & Genomic Pathology, Departments of Pathology and Internal Medicine, Michigan Medicine, University of Michigan, Ann Arbor, MI, USA
e-mail: giordano@umich.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_24

Table 24.1 Summary of genetic and genomic features of thyroid cancers

	Somatic alterations				
Type of thyroid carcinoma	SSNVs	Gene fusions	SCNAs	*BRAF*^{V600E}-like expression signature	Tumor mutational burden
Papillary carcinoma	*BRAF*^{V600E}, *RAS*	*RET, NTRK1/3, BRAF, ALK*	Few; 22q loss, 1q gain	Present	Low
Follicular carcinoma	*RAS, PTEN, TERT* promoter, *RB1*	*PAX8-PPARG*	Many; 22q loss, 1q gain	NA	Low
Hurthle cell carcinoma	*RAS, TP53, TERT* promoter, mitochondrial DNA	Unknown	Many	NA	Low
Poorly differentiated carcinoma	*BRAF*^{V600E}, *RAS, TERT* promoter, *TP53, EIF1AX*	Unknown	Unknown	Present	Moderate
Anaplastic carcinoma	*BRAF*^{V600E}, *RAS, TP53, TERT promoter, EIF1AX, heterogeneous mutations*	RET	Many (aneuploidy)	Absent	High

projects [6]. The study cohort consisted of 496 primary and 8 metastatic tumors involving regional lymph nodes. Associated clinicopathological data was obtained for each tumor, and PTCs were further classified as one of three predominant pathological subtypes, i.e., classical or usual PTC (n = 324), tall cell variant PTC (n = 35), or follicular variant PTC (n = 99). Tumors were evaluated by the standard TCGA molecular platforms, which included whole-exome DNA sequencing, mRNA sequencing, miRNA sequencing, copy number profiling, and DNA methylation profiling. Analysis of the resulting molecular data fell along three lines. First, somatic mutations, including single nucleotide variants (SSNVs), small insertions and deletions (indels), gene fusions, and copy number alterations (SCNAs), were identified to characterize the genomic landscape of PTC, with a special emphasis on those cases without previously known driver mutations. Second, the biological consequences of the mutations were explored using multidimensional molecular data. Finally, molecular data was used to derive novel insights into the classification of PTC by incorporating data on genotype, cell signaling, differentiation, and clinical risk of aggressive behavior.

Using whole-exome DNA sequencing data, the TCGA study demonstrated a relatively low tumor mutational burden (0.41 non-synonymous mutations per Mb) compared with other common cancers, especially those associated with carcinogens such as cigarette smoke (lung) and sunlight (melanoma) [7]. Tumor mutational burden increased with patient age.

The TCGA study confirmed the dominant role of *BRAF* mutations in PTC. *BRAF*^{V600E} was by far the most common mutation, although the study illustrated the many ways that *BRAF* can transform to an oncogene. For instance, *BRAF* indels as well as diverse gene fusions were identified and found to be mutually exclusive with the more common alterations, suggesting that these rare mutations are oncogenic. A single case of a follicular variant PTC with *BRAF*^{K601E} was identified. Collectively, some form of *BRAF* mutation was present in >60% of tumors.

Most other SSNVs beyond *BRAF*^{V600E} occurred within one of the *RAS* genes, with the following relative frequency (NRAS, 8.5%; HRAS, 3.5%; KRAS, 1.2%). As expected, *RAS* mutations occurred as hotspot mutations in codons 12 and 61.

Somatic mutations of *EIF1AX* were first identified in thyroid cancer by the TCGA study, with 1.5% of PTC cases carrying point mutations of this X-linked translation initiation factor. Since the TCGA report, mutations of *EIF1AX* have been confirmed in PTC and found in other types of thyroid cancer [8–11]. The oncogenic mechanism of mutated *EIF1AX* may involve aberrant gene expression via stabilization of translation pre-initiation sites [12].

Other significantly mutated genes accordingly to the MutSig algorithm [13] included *PPM1D* and *CHEK2*. Known cancer genes with mutations included *TP53*, *ARIDB1*, *MLL*, *PTEN*, *ATM*, *RB1*, *EZH1*, *MEN1*, *MLL3*, *APC*, and *NF1*. In addition, *TERT* promoter mutations were identified in 9% of cases and associated with aggressive features.

The TCGA study reaffirmed and expanded the role of gene rearrangements and translocations in PTC. Such gene fusions were found in 15.3% of tumors, and no tumor contained more than one fusion. In addition to being mutually exclusive with each other, gene fusions did not co-occur with *BRAF*, *RAS*, or *EIF1AX* mutations. This relationship of mutually exclusive driving mutations represents one the most striking examples across all cancer types and illustrates the dominant role these mutations play in PTC pathogenesis. In addition to identifying novel partners of common gene fusions (e.g., *FKBP15-RET*), entirely new fusions were discovered (e.g., *UCAC-LTK*). The larger majority of these tumors demonstrated preservation and overexpression of kinase domains in the resulting fusion, providing evidence for their function as constitutively activated kinases.

In addition to point mutations and gene fusions, a significant minority (27.2%) of PTCs also contained SCNAs, which were preferentially present in cases without the common driver mutations described above. This relationship

suggests a role for SCNAs as driver events in PTC. Tumors with SCNAs were enriched for follicular variant PTC, consistent with prior observations and providing support for a close relationship between follicular variant PTC and true follicular neoplasms [14]. Recent comparative pan-cancer studies across many cancer types highlighted the exceptional low rates of aneuploidy observed in PTC [15].

Somatically altered pathways were categorized using sophisticated computational approaches that combine data on SSNVs, gene fusions, and SCNAs. The HotNet2 algorithm [16] identified numerous relevant subnetworks, including MAPK pathway, ECM-receptor interactions, and FANCA-associated protein complex. Using this approach, *RAP1GAP* was identified as a rare potential driver gene in the MAPK pathway, and the COMSIC database confirms this gene is rarely mutated in thyroid cancers [17].

By combining single nucleotide variants, indels, gene fusions, and copy number alterations, driving mutations were found in the large majority of PTC (98.8%). This collective result represented a major advance with significant clinical implications for genotyping assays of thyroid nodules [18–20].

The TCGA study also yielded insights into the consequence of the above-described driving genetic defects by leveraging the availability of rich multidimensional genomic data and exploiting the mutual exclusively of *BRAF*^{V600E} and *RAS* mutations. A gene expression-based score was developed that measured whether the gene expression profile of a given tumor resembled either *BRAF*^{V600E} or *RAS* mutant profiles. Using a large cohort of tumors (n = 391) that had both exome and RNA sequencing data, a 71-gene expression signature was derived by comparing *BRAF*^{V600E}-mutated and *RAS*-mutated tumors. Data from these 71 genes was then used to derive a continuous score, termed the *BRAF*^{V600E}*-RAS score (BRS)*, with negative scores being *BRAF*^{V600E}-like and positive *RAS*-like.

The BRS was used as a reference scale to interrogate the signaling consequences of the observed mutations across the cohort (Fig. 24.1). The resulting view yielded several interesting insights. All *BRAF*^{V600E} mutations were *BRAF*^{V600E}-like, and all *RAS* mutations were *RAS*-like, with one triple mutant exception. The rare *BRAF* mutations (indels and *BRAF*^{K601E}) were *RAS*-like, a result that illustrates the complexity of MAPK signaling with differences in ERK feedback

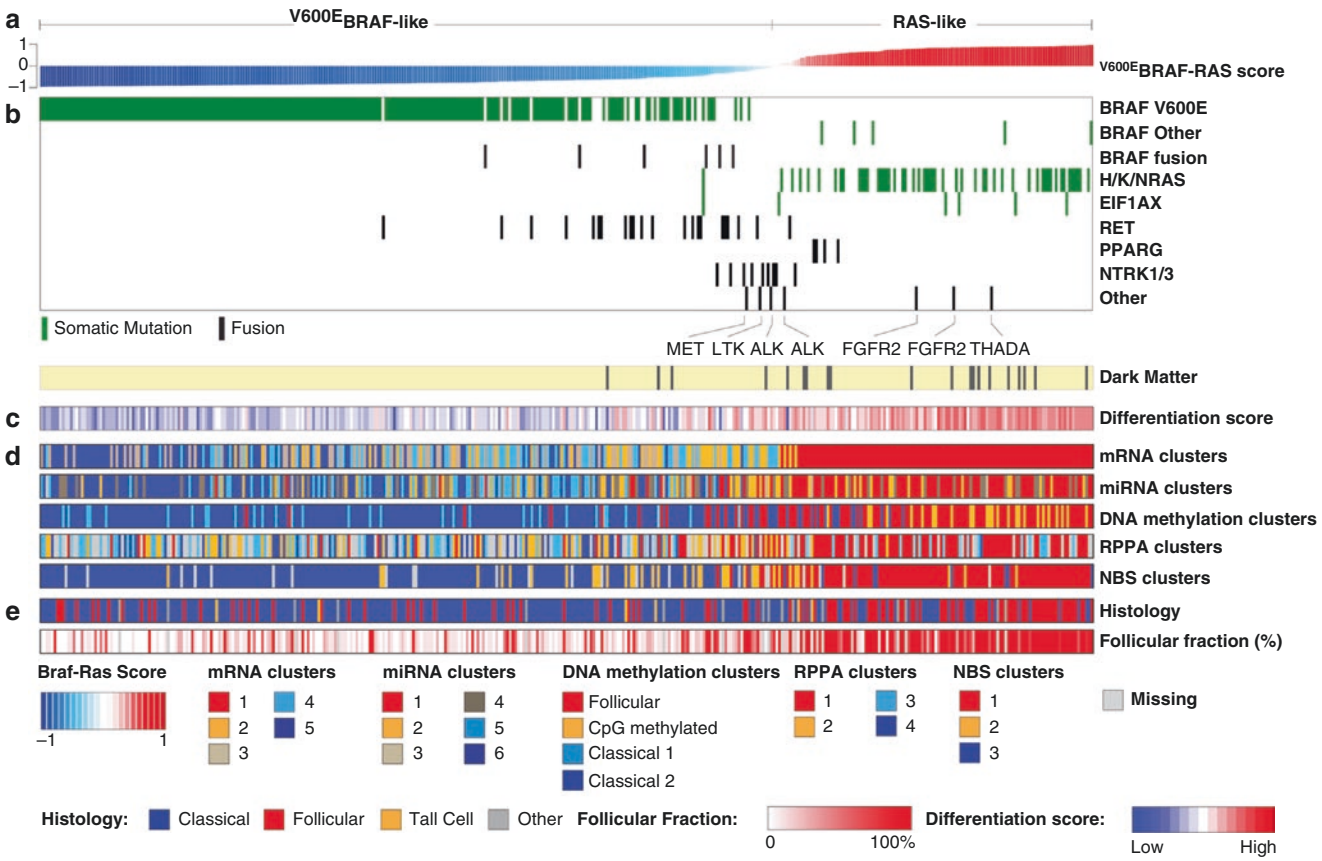


Fig. 24.1 The *BRAF*^{V600E}*-RAS* Score (BRS). (**a**) 391 PTCS were ranked according to their *BRAF*^{V600E}*-RAS* scores, with negative being *BRAF*^{V600E}-like and positive being *RAS*-like. The BRS was strongly associated with (**b**) mutational status, (c) thyroid differentiation score, (**d**) genomic clusters, and (**e**) histologic type and follicular fraction. (From Cancer Genome Atlas Research N [6] with permission of Elsevier)

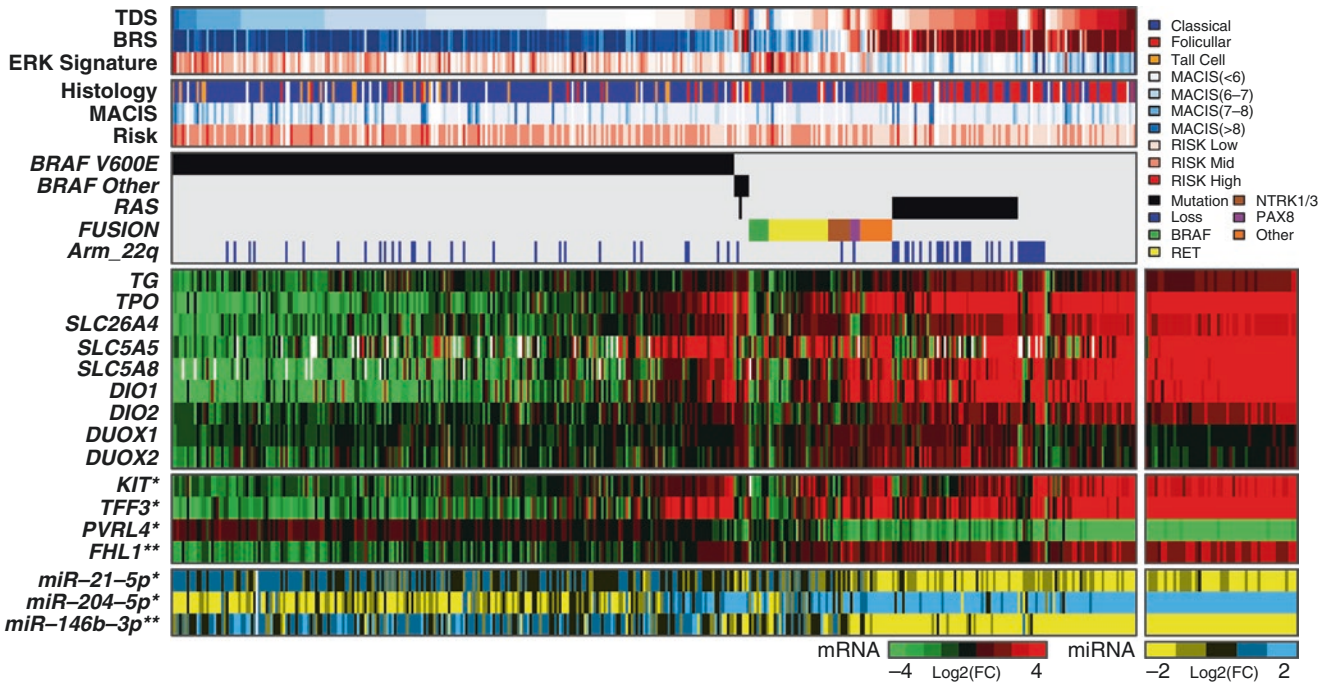


Fig. 24.2 Role of thyroid differentiation. The thyroid differentiation score (TDS) across the TCGA cohort with tumors sorted by mutational status. Below the TDS is the BRS, ERK signature. See [6], histology, MACIS score, recurrence risk, mutations, and gene expression data for thyroid-related and other genes. (From Cancer Genome Atlas Research N [6] with permission of Elsevier)

[21, 22]. Conversely, *BRAF* and *RET* fusions were weakly *BRAF*^{V600E}-like, and other fusions were neutral in the BRS scheme. These results highlight the differences in the functional consequences of these driving mutations and are reinforced by genotype-phenotype correlations, including the observation that *BRAF*^{K601E}-mutated cancers are highly enriched for follicular variant PTC [23].

Given the essential role of follicular cell differentiation in thyroid cancer, the TCGA study explored the impact of differentiation across the cohort by extracting the gene expression data for 16 thyroid metabolism and function genes. The resulting score, termed the *Thyroid Differentiation Score (TDS),* was again used as a reference scale to interrogate the role of differentiation across the various mutation groups (Fig. 24.2). This analysis confirmed higher follicular cell differentiation in *RAS*-mutated tumors compared to tumors with *RET* or *BRAF* fusions and *BRAF*^{V600E}. However, differentiation was highly variable within the *BRAF*^{V600E} cohort, an unexpected finding that has implications for the prognostic power of *BRAF*^{V600E} and response to RAI therapy.

Genomic data was also used to derive various molecular classifications of PTC. The BRS and TDS scores were incorporated in analyses to illustrate the biological differences between the resulting molecular clusters and to inform the any relationships between tumor cluster, tumor histology, genotype, signaling, and differentiation. One such example based on miRNA expression identified different clusters with differential oncogenic miR expression (e.g., miR-146b) and low BRS and TDS measures, suggesting that expression of such oncogenic miRs may modulate the PTC phenotype (Fig. 24.3).

At the outset of the TCGA project, there was concern that the results might be pedestrian due to the indolent nature of PTC. On the contrary, the overall quiet genome of PTC with low tumor mutational burden and stable genome with few copy number alterations permitted derivation of critical insights into many aspects of PTC biology including the consequences of common driver mutations. Such insights are difficult to extract in more complicated tumor types with unstable genomes, higher degrees of mutational burden, and multiple activated pathways. Therefore, in the end the TCGA study succeeded by identifying novel mutations, revealing insights into the pathogenesis of PTC, and providing a foundation for comparison of additional studies of other types of thyroid cancer.

Routine genomic profiling of tumors from patients with advanced cancer has rapidly become the standard of oncology care. Accordingly, data from these clinical sequencing programs, from both academic and commercial laboratories, have been accumulating to the point where in silico analysis of large cohorts is now feasible and informative. Such a report [24] on the genetic alterations present in a large cohort of advanced-stage PTCS using two similar genomic assays

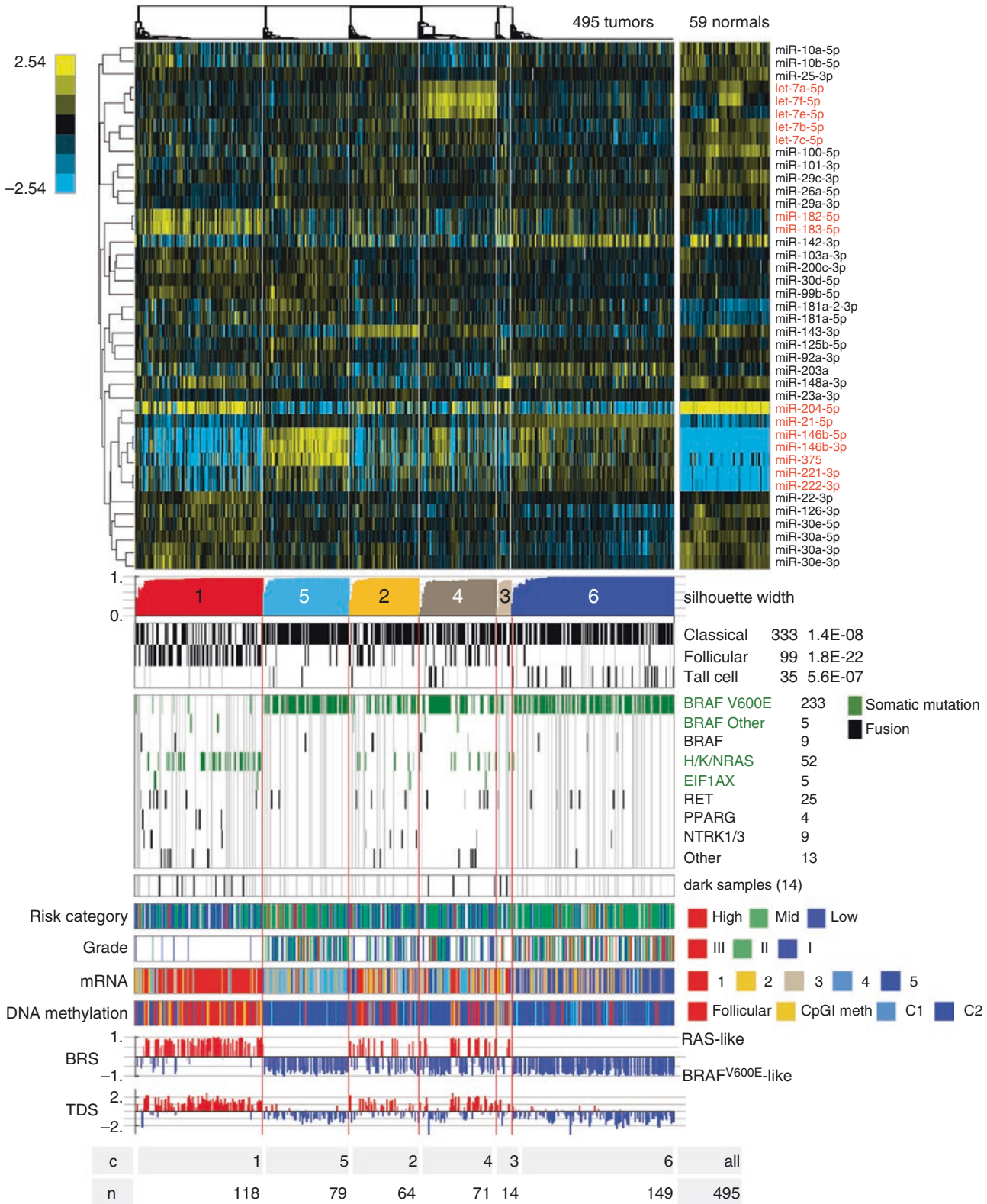


Fig. 24.3 miRNA-seq hierarchical cluster analysis. Hierarchical clustering of miRNA expression data along with various histological, mutational, clinical, and genomic data. See [6] for additional details. (From Cancer Genome Atlas Research N [6] with permission of Elsevier)

(MSK-IMPACT [25] and FoundationOne [26]) reinforced TCGA findings but also highlighted the differences in advanced-stage disease compared to the TCGA cohort. Similar to TCGA findings, a low overall mutational tumor burden was also observed. Mutational burden again increased with patient age, with pediatric PTCs exhibiting the lowest number of mutations. Mutations involving *BRAF* dominated the somatic mutational landscape, with 74% of cases possessing one form of mutation. *RAS* mutations and *RET* fusions were found in 9% and 7% of cases, respectively. Similar to TCGA results, driver mutations involving *BRAF*, *RAS,* and *RET* were strongly mutual exclusive. Collectively, 84% and 18% of cases had alterations of MAPK and PI3K/AKT/mTOR pathway genes, respectively. In keeping with the emerging model of *TERT* promoter mutations being associated with clinically aggressive tumors, these mutations were present in 61% of cases and were enriched in older patients. Similarly, mutations of *TP53* were much more frequent (10%) compared to the TCGA cohort (0.7%). Mutations of *CDKN2A* and *RBM10*, not observed in TCGA, were present in 8% and 7% of cases, respectively. Collectively, the observed mutated genes largely overlapped with those in TCGA but are mutated at markedly higher rates, a finding which almost certainly reflects true biological differences present in advanced-stage disease.

Pediatric PTCs possess a somewhat distinct genetic landscape, characterized by a lower tumoral mutational burden; more frequent kinase gene fusions involving *RET*, *ALK*, and *NTRK*; and less frequent *BRAF* mutations [24, 27–29].

Additional insights into the PTC pathogenesis will be derived from the recently published studies of the TCGA Pan-Cancer Atlas project (see https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html?code=cell-site) that focus on different aspects of cancer genomics [30–33].

Follicular Carcinoma

Follicular thyroid carcinoma (FTC) is much less common than PTC, and, accordingly, their genomic characteristics have been less thoroughly studied and defined. FTC is characterized by *RAS* point mutations and *PPARG* rearrangements [34], as well as other somatic alterations that involve the PI3K/AKT/mTOR pathway. Patients with Cowden syndrome have germline *PTEN* inactivation and develop both FTA and FTC in a multifocal distribution, supporting a model in which FTC arises from pre-existing FTA [35].

Targeted genome-wide DNA sequencing of FTCs from 48 patients revealed a heterogeneous group of somatic driving alterations, many of which were novel, including mutations of oncogenes (*MDM2*, *FLI1*), transcriptional regulators (*MITF*, *FLI1*, *ZNF331*), epigenetic enzymes (*KMT2A*, *NSD1*, *NCOA1*, *NCOA2*), and kinases (*JAK3*, *CHEK2*, *ALK*) [36]. Interestingly, this Polish population had a very low rate (4%) of *RAS* mutation.

A recent genomic study of follicular thyroid adenoma (FTA) and FTC was performed to characterize their genomes and explore the etiologic relationship between FTA and FTC [37]. Using whole-exome sequencing, no significant differences were observed in the frequency of non-synonymous mutations in FTA and FTC. Moreover, the frequency of known driving mutations was similar between the tumor types, and there were no observed correlations between genomic and clinicopathological features. However, there were significant differences in SCNAs with FTC possessing a sixfold higher rate of SCNAs, supporting a role for their involvement in the progression from FTA to FTC [38, 39]. Recurrent SCNAs included gain of 1q and loss of 22q, both of which were observed in the TCGA study. Elegant work has shown that *NF2* loss, located on 22q, promotes oncogenesis via increased MAPK signaling specifically in *RAS*-mutated thyroid cancers [40].

TERT promoter mutations have also been reported in a significant minority of FTC and are associated with aggressive disease [41, 42]. Data from clinical sequencing of 65 cases of advanced-stage FTCs provided the opportunity to examine its genetics [24]. *RAS* mutations were most frequent (66%), and *BRAF* mutations were less frequent (7.6%). Rare *BRAF* mutations known to be associated with follicular-patterned tumors (e.g., $BRAF^{K601E}$ in follicular variant PTC) were overrepresented, and only one case had $BRAF^{V600E}$. In this advanced cohort, *TERT* promoter mutations were markedly more frequent (71%), providing more support for its role in aggressive disease [43, 44]. Mutations in *PTEN* (9%) and *RB1* (9%) were more frequent compared to PTC.

Circulating cell free DNA genomic approaches are just beginning to be applied to FTC, with promising results [45].

Hurthle Cell Carcinoma

Hurthle cell carcinoma (HCC) is a distinctive type of thyroid cancer that is composed of large cells with abundant granular cytoplasm and prominent nucleoli, termed oncocytes [46]. The oncocytic phenotype is related to the cellular accumulation of mitochondria [47]. The collective data suggest that HCC has distinctive cancer genome characterized by a higher frequency of mitochondrial mutations [48, 49], different driver mutational profiles [50], and high rates of aneuploidy [39, 50, 51]. Transcriptome studies have revealed three tumor classes that roughly correspond to adenoma, minimally invasive HCC, and widely invasive HCC [50].

Data from clinical sequencing of 35 advanced cancer patients with HCC affirm the view that HCC is a distinct type of thyroid cancer [24]. *RAS* mutations were less frequent

(15%), and there were no *BRAF* or *RET* mutations identified. *TP53* was present in 20% of tumors. *TERT* promoter mutations were frequent (59%), again suggesting a universal role for *TERT* promoter mutation in the development of aggressive thyroid cancers of all types.

As noted above, multidimensional genomic studies of HCC are underway and are expected to provide additional insights into the cancer genome of this enigmatic type of thyroid cancer.

Poorly Differentiated Carcinoma

PDCA is a relatively rare thyroid cancer that represents an intermediate type that taxonomically lies between differentiated thyroid cancers and anaplastic carcinoma. Recent genomic studies of PDCA confirm this pathology-informed viewpoint. A study of a large cohort of PDCA ($n = 84$) using MSK-IMAPCT suggested that PDCA contained increased numbers of driving alterations that concentrated in a few pathways or functional gene groups, including PI3K/AKT/mTOR, SWI/SNF chromatin remodeling complex, histone methyltransferases, and DNA mismatch repair [10]. Moreover, *TP53* (8%) and *TERT* promoter (40%) mutations were mutated at higher frequency compared to PTC. Moreover, *TERT* promoter mutations were clonal, in contrast to PTC in which they were subclonal according to TCGA data.

Similar to PTC, genomic data support the view that PDCA can be divided into *BRAF*^{V600E}-like and *RAS*-like subgroups [52]. This distinction exhibits high correlation with histopathological classification. The original form of PDCA, termed insular carcinoma, is now known to represent a *RAS*-like form of PDCA with a high frequency of *RAS* mutations [10, 53, 54]. Conversely, a non-insular form of PDCA is now recognized based on mitotic rate, necrosis, and high nuclear grade [55]. Unlike *RAS*-like PDCA, such PDCAs frequently contain *BRAF*^{V600E} mutations and have a *BRAF*^{V600E}-like phenotype [10].

PDCA also contain frequent *EIF1AX* mutations (11%) compared to PTC (1.5%). In PTC, *EIF1AX* mutations were mutually exclusive with the common driver mutation; however, in PDCA, they were coexistent with *RAS* mutations. Moreover, tumors with *EIF1AX* mutations were associated with larger tumors and shorter survival [10].

Anaplastic Carcinoma

ATC is the most aggressive and lethal form of thyroid cancer and, by definition, is composed of undifferentiated follicular cells. Given its undifferentiated state, ATC represents the ultimate stage in the scheme of thyroid cancer progression, which fortunately only occurs in a very small number of patients (incidence of 1 to 2 cases per million). However, ATC is responsible for a significant proportion of thyroid cancer mortality. Accordingly, ATC has been the subject of significant genetic and genomic investigation in an attempt to better understand its pathogenesis and develop novel therapeutic approaches.

ATC is thought to arise from pre-existing thyroid cancers, a view supported by its shared genetic drivers with differentiated thyroid cancers (e.g., *BRAF*^{V600E} and *RAS*) and several pathologic observations. Many ATC tumors contain differentiated components, most commonly high-risk PTCs such as tall cell variants, and small ATCs without coexisting differentiated tumors are essentially nonexistent.

The cancer genome of ATC displays a markedly high degree of genetic heterogeneity according to numerous studies, a finding that largely explains the challenges in effectively treating ATC. Among all thyroid cancers, ATC possesses the highest frequency of *TP53* mutation (54–73%) [10, 24, 56]. *TP53* is one of the most commonly mutated genes in all cancers (50%) and has been referred to as the "guardian of the genome" because of its role in triggering an antiproliferative transcriptional program in response to stress stimuli [57]. Accordingly, tumors with loss of p53 function, including most ATCs, accumulate a wide variety of genome changes compared to tumors without p53 loss [58, 59].

An initial NGS study of 22 ATCs defined its mutational landscape, which included the common drivers *BRAF*^{V600E} and *RAS*, as well as cancer genes not previously associated with thyroid cancer such as mismatch repair genes. Two such tumors displayed a hypermutator phenotype with a high tumor mutational burden. A similar genetic investigation of a large French cohort ($n = 144$ from ten centers) revealed significant genetic heterogeneity with the following mutational profile: *TP53* (54%), *RAS* (43%), *BRAF* (13.8%), and PI3K/AKT/mTOR pathway (17%) [56].

Using the MSK-IMPACT genomic assay, Landa et al. derived significant insights into ATC using a cohort of 33 tumors [10]. High mutation frequencies of *TP53* (73%) and *TERT* promoter (73%) mutations were observed. Similar to other studies, *BRAF*^{V600E} and *RAS* mutations were mutually exclusive. Using transcriptome data, the *BRAF*^{V600E}-like and *RAS*-like distinction that was preserved in PDCA was lost in ATC, presumably due to the higher tumor mutational burden and aneuploidy in ATC [60, 61]. The mutational frequency of *EIF1AX* (9%) was similar to that seen in PDCA. The same four pathways and gene groups identified in PDCA were also mutated in ATC but with higher mutational frequencies.

A large report of 196 advanced ATCs provided further insights into its genetics and tumor classification [24]. Using hierarchical clustering and machine learning analysis of the mutational data, four molecular clusters were identified. Cluster 1 was characterized by *BRAF* mutations and are

thought to represent those ATCs that evolved from pre-existing PTCs. Likewise, cluster 3 contained *RAS* mutations and are believed to be derived from *RAS*-mutated FTCs. Cluster 4 ATCs have a diversity of mutations (*RAS*, *PTEN*, *NF1*, *RB1*) with higher mutational burdens related to mismatch repair gene defects. These ATCs are speculated to be originating from HCCs, but this association requires validation as the cancer genome of HCC is not fully defined. Finally, cluster 2 ATCs are characterized by loss of function mutations of cell cycle regulators, *CDKN2A* and *CDKN2B*, and do not have a well-defined precursor lesion.

Clinical Applications of Genomic Information

The availability of genome-wide molecular data for thyroid cancer has catalyzed the application of selected molecular information to aid in the decision to undergo surgical removal of a thyroid nodule. Clinical evaluation of thyroid nodules involves clinical exam, ultrasound imaging, and fine-needle aspiration (FNA) cytology [62, 63]. Despite the overall efficacy of this approach, a significant number of patients have indeterminate FNA cytology results and subsequently undergo surgery for what turns out on final pathology examination to be benign disease, either nodular hyperplasia or adenomas. It was envisioned a decade ago that genome-wide molecular information would produce significant advances in thyroid tumor classification [18] and molecular tests have been developed specifically to address these issues. Because there is a strong genotype-phenotype correlation in differentiated thyroid cancers, a variety of testing approaches are feasible and have been developed (reviewed in [64, 65]). These include direct identification of common driving mutations (i.e., genotyping) [19, 66–68] and/or identification of gene expression profiles associated with benign or malignant nodules [69–71]. The clinical and economic impact of molecular testing is a subject of active investigation [67, 72–75], especially as these assays continue to undergo continuous development with resulting refinements [68, 69]. Along similar lines, genome-wide DNA methylation profiles have also been proposed to have prognostic significance in differentiated thyroid cancers [76].

Conclusion

The last decade has witnessed an explosion in genome-wide data for many cancer types, including follicular cell tumors of the thyroid gland. This information is yielding advances in identification and understanding of the genetic drivers of thyroid cancer, which has significance for molecular diagnostics and cancer therapy. Molecular information derived from genomic studies, in various forms, has been deployed for a variety of clinical applications. The coming years will see further gains as the cancer genomes of the less common types of thyroid cancer become available.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
2. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57–70.
3. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418–26.
4. Blum A, Wang P, Zenklusen JC. SnapShot: TCGA-analyzed tumors. Cell. 2018;173(2):530.
5. Giordano TJ. Genomic hallmarks of thyroid neoplasia. Annu Rev Pathol. 2018;13:141–62.
6. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014;159(3):676–90.
7. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173(2):371–85.e18.
8. Kunstman JW, Juhlin CC, Goh G, Brown TC, Stenman A, Healy JM, et al. Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. Hum Mol Genet. 2015;24(8):2318–29.
9. Karunamurthy A, Panebianco F, Hsiao S J, Vorhauer J, Nikiforova MN, Chiosea S, et al. Prevalence and phenotypic correlations of EIF1AX mutations in thyroid nodules. Endocr Relat Cancer. 2016;23(4):295–301.
10. Landa I, Ibrahimpasic T, Boucai L, Sinha R, Knauf JA, Shah RH, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. J Clin Invest. 2016;126(3):1052–66.
11. Yoo SK, Lee S, Kim SJ, Jee HG, Kim BA, Cho H, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. PLoS Genet. 2016;12(8):e1006239.
12. Martin-Marcos P, Zhou F, Karunasiri C, Zhang F, Dong J, Nanda J, et al. eIF1A residues implicated in cancer stabilize translation preinitiation complexes and favor suboptimal initiation sites in yeast. elife. 2017;6
13. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505(7484):495–501.
14. Wreesmann VB, Ghossein RA, Hezel M, Banerjee D, Shaha AR, Tuttle RM, et al. Follicular variant of papillary thyroid carcinoma: genome-wide appraisal of a controversial entity. Genes Chromosomes Cancer. 2004;40(4):355–64.
15. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell. 2018;33(4):676–89. e3
16. Vandin F, Clay P, Upfal E, Raphael BJ. Discovery of mutated subnetworks associated with clinical data in cancer. Pac Symp Biocomput. 2012:55–66.
17. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777–D83.
18. Eszlinger M, Krohn K, Hauptmann S, Dralle H, Giordano TJ, Paschke R. Perspectives for improved and more accurate classification of thyroid epithelial tumors. J Clin Endocrinol Metab. 2008;93(9):3286–94.

19. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, et al. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. Cancer. 2014;120(23):3627–34.
20. Wylie D, Beaudenon-Huibregtse S, Haynes BC, Giordano TJ, Labourier E. Molecular classification of thyroid lesions by combined testing for miRNA gene expression and somatic gene alterations. J Pathol Clin Res. 2016;2(2):93–103.
21. Fagin JA, Wells SA Jr. Biologic and clinical perspectives on thyroid cancer. N Engl J Med. 2016;375(11):1054–67.
22. Kidger AM, Keyse SM. The regulation of oncogenic Ras/ERK signalling by dual-specificity mitogen activated protein kinase phosphatases (MKPs). Semin Cell Dev Biol. 2016;50:125–32.
23. Afkhami M, Karunamurthy A, Chiosea S, Nikiforova MN, Seethala R, Nikiforov YE, et al. Histopathologic and clinical characterization of thyroid tumors carrying the BRAF(K601E) mutation. Thyroid. 2016;26(2):242–7.
24. Pozdeyev N, Gay L, Sokol ES, Hartmaier RJ, Deaver KE, Davis SN, et al. Genetic analysis of 779 advanced differentiated and anaplastic thyroid cancers. Clin Cancer Res. 2018;24:3059.
25. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn. 2015;17(3):251–64.
26. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31(11):1023–31.
27. Prasad ML, Vyas M, Horne MJ, Virk RK, Morotti R, Liu Z, et al. NTRK fusion oncogenes in pediatric papillary thyroid carcinoma in Northeast United States. Cancer. 2016;122(7):1097–107.
28. Hardee S, Prasad ML, Hui P, Dinauer CA, Morotti RA. Pathologic characteristics, natural history, and prognostic implications of BRAF(V600E) mutation in pediatric papillary thyroid carcinoma. Pediatr Dev Pathol. 2017;20(3):206–12.
29. Vanden Borre P, Schrock AB, Anderson PM, Morris JC 3rd, Heilmann AM, Holmes O, et al. Pediatric, adolescent, and young adult thyroid carcinoma harbors frequent and diverse targetable genomic alterations, including kinase fusions. Oncologist. 2017;22(3):255–63.
30. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell. 2018;173(2):291–304. e6
31. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. Cell. 2018;173(2):305–20. e10
32. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. Immunity. 2018;48(4):812–30. e14
33. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. Cell. 2018;173(2):321–37. e10
34. Nikiforova MN, Lynch RA, Biddinger PW, Alexander EK, Dorn GW 2nd, Tallini G, et al. RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. J Clin Endocrinol Metab. 2003;88(5):2318–26.
35. Harach HR, Soubeyran I, Brown A, Bonneau D, Longy M. Thyroid pathologic findings in patients with Cowden disease. Ann Diagn Pathol. 1999;3(6):331–40.
36. Swierniak M, Pfeifer A, Stokowy T, Rusinek D, Chekan M, Lange D, et al. Somatic mutation profiling of follicular thyroid cancer by next generation sequencing. Mol Cell Endocrinol. 2016;433:130–7.
37. Jung SH, Kim MS, Jung CK, Park HC, Kim SY, Liu J, et al. Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma. Oncotarget. 2016;7(43):69638–48.
38. Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K. Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. Br J Cancer. 1998;78(8):1012–7.
39. Roque L, Rodrigues R, Pinto A, Moura-Nunes V, Soares J. Chromosome imbalances in thyroid follicular neoplasms: a comparison between follicular adenomas and carcinomas. Genes Chromosomes Cancer. 2003;36(3):292–302.
40. Garcia-Rendueles ME, Ricarte-Filho JC, Untch BR, Landa I, Knauf JA, Voza F, et al. NF2 loss promotes oncogenic RAS-induced thyroid cancers via YAP-dependent transactivation of RAS proteins and sensitizes them to MEK inhibition. Cancer Discov. 2015;5(11):1178–93.
41. Liu T, Wang N, Cao J, Sofiadis A, Dinets A, Zedenius J, et al. The age- and shorter telomere-dependent TERT promoter mutation in follicular thyroid cell-derived carcinomas. Oncogene. 2014;33(42):4978–84.
42. Melo M, da Rocha AG, Vinagre J, Batista R, Peixoto J, Tavares C, et al. TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. J Clin Endocrinol Metab. 2014;99(5):E754–65.
43. Kim TH, Kim YE, Ahn S, Kim JY, Ki CS, Oh YL, et al. TERT promoter mutations and long-term survival in patients with thyroid cancer. Endocr Relat Cancer. 2016;23(10):813–23.
44. Liu R, Xing M. TERT promoter mutations in thyroid cancer. Endocr Relat Cancer. 2016;23(3):R143–55.
45. Song J, Yang Z. Case report: whole exome sequencing of circulating cell-free tumor DNA in a follicular thyroid carcinoma patient with lung and bone metastases. J Circ Biomark. 2018;7:1849454418763725.
46. Carcangiu ML. Hurthle cell carcinoma: clinic-pathological and biological aspects. Tumori. 2003;89(5):529–32.
47. Satoh M, Yagawa K. Electron microscopic study on mitochondria in Hurthle cell adenoma of thyroid. Acta Pathol Jpn. 1981;31(6):1079–87.
48. Gasparre G, Porcelli AM, Bonora E, Pennisi LF, Toller M, Iommarini L, et al. Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. Proc Natl Acad Sci U S A. 2007;104(21):9001–6.
49. Evangelisti C, de Biase D, Kurelac I, Ceccarelli C, Prokisch H, Meitinger T, et al. A mutation screening of oncogenes, tumor suppressor gene TP53 and nuclear encoded mitochondrial complex I genes in oncocytic thyroid tumors. BMC Cancer. 2015;15:157.
50. Ganly I, Ricarte Filho J, Eng S, Ghossein R, Morris LG, Liang Y, et al. Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid malignancy. J Clin Endocrinol Metab. 2013;98(5):E962–72.
51. Corver WE, Ruano D, Weijers K, den Hartog WC, van Nieuwenhuizen MP, de Miranda N, et al. Genome haploidisation with chromosome 7 retention in oncocytic follicular thyroid carcinoma. PLoS One. 2012;7(6):e38287.
52. Xu B, Ghossein R. Genomic landscape of poorly differentiated and anaplastic thyroid carcinoma. Endocr Pathol. 2016;27(3):205–12.
53. Volante M, Rapa I, Gandhi M, Bussolati G, Giachino D, Papotti M, et al. RAS mutations are the predominant molecular alteration in poorly differentiated thyroid carcinomas and bear prognostic impact. J Clin Endocrinol Metab. 2009;94(12):4735–41.
54. Pita JM, Figueiredo IF, Moura MM, Leite V, Cavaco BM. Cell cycle deregulation and TP53 and RAS mutations are major events in poorly differentiated and undifferentiated thyroid carcinomas. J Clin Endocrinol Metab. 2014;99(3):E497–507.

55. Hiltzik D, Carlson DL, Tuttle RM, Chuai S, Ishill N, Shaha A, et al. Poorly differentiated thyroid carcinomas defined on the basis of mitosis and necrosis: a clinicopathologic study of 58 patients. Cancer. 2006;106(6):1286–95.
56. Bonhomme B, Godbert Y, Perot G, Al Ghuzlan A, Bardet S, Belleannee G, et al. Molecular pathology of anaplastic thyroid carcinomas: a retrospective study of 144 cases. Thyroid. 2017;27(5):682–92.
57. Kastenhuber ER, Lowe SW. Putting p53 in context. Cell. 2017;170(6):1062–78.
58. Eischen CM. Genome stability requires p53. Cold Spring Harb Perspect Med. 2016;6(6):pii: a026096.
59. Tomasini R, Mak TW, Melino G. The impact of p53 and p73 on aneuploidy and cancer. Trends Cell Biol. 2008;18(5):244–52.
60. Klemi PJ, Joensuu H, Eerola E. DNA aneuploidy in anaplastic carcinoma of the thyroid gland. Am J Clin Pathol. 1988;89(2):154–9.
61. Pinto AE, Silva G, Banito A, Leite V, Soares J. Aneuploidy and high S-phase as biomarkers of poor clinical outcome in poorly differentiated and anaplastic thyroid carcinoma. Oncol Rep. 2008;20(4):913–9.
62. Baldini E, Sorrenti S, Tartaglia F, Catania A, Palmieri A, Pironi D, et al. New perspectives in the diagnosis of thyroid follicular lesions. Int J Surg. 2017;41(Suppl 1):S7–S12.
63. Cibas ES. Fine-needle aspiration in the work-up of thyroid nodules. Otolaryngol Clin N Am. 2010;43(2):257–71. vii–viii.
64. Nishino M, Nikiforova M. Update on molecular testing for cytologically indeterminate thyroid nodules. Arch Pathol Lab Med. 2018;142(4):446–57.
65. Vargas-Salas S, Martinez JR, Urra S, Dominguez JM, Mena N, Uslar T, et al. Genetic testing for indeterminate thyroid cytology: review and meta-analysis. Endocr Relat Cancer. 2018;25(3):R163–R77.
66. Nikiforova MN, Wald AI, Roy S, Durso MB, Nikiforov YE. Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. J Clin Endocrinol Metab. 2013;98(11):E1852–60.
67. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, et al. Impact of the multi-gene ThyroSeq next-generation sequencing assay on cancer diagnosis in thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance cytology. Thyroid. 2015;25(11):1217–23.
68. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, et al. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. Cancer. 2018;124(8):1682–90.
69. Jug RC, Datto MB, Jiang XS. Molecular testing for indeterminate thyroid nodules: performance of the Afirma gene expression classifier and ThyroSeq panel. Cancer Cytopathol. 2018; https://doi.org/10.1002/cncy.21993.
70. Kloos RT. Molecular profiling of thyroid nodules: current role for the Afirma gene expression classifier on clinical decision making. Mol Imaging Radionucl Ther. 2017;26(Suppl 1):36–49.
71. Duh QY, Busaidy NL, Rahilly-Tierney C, Gharib H, Randolph G. A systematic review of the methods of diagnostic accuracy studies of the Afirma gene expression classifier. Thyroid. 2017;27(10):1215–22.
72. Aragon Han P, Olson MT, Fazeli R, Prescott JD, Pai SI, Schneider EB, et al. The impact of molecular testing on the surgical management of patients with thyroid nodules. Ann Surg Oncol. 2014;21(6):1862–9.
73. Sacks WL, Bose S, Zumsteg ZS, Wong R, Shiao SL, Braunstein GD, et al. Impact of Afirma gene expression classifier on cytopathology diagnosis and rate of thyroidectomy. Cancer Cytopathol. 2016;124(10):722–8.
74. Guo HQ, Zhao H, Zhang ZH, Zhu YL, Xiao T, Pan QJ. Impact of molecular testing in the diagnosis of thyroid fine needle aspiration cytology: data from mainland China. Dis Markers. 2014;2014:912182.
75. Nicholson KJ, Yip L. An update on the status of molecular testing for the indeterminate thyroid nodule and risk stratification of differentiated thyroid cancer. Curr Opin Oncol. 2018;30(1):8–15.
76. Bisarro Dos Reis M, Barros-Filho MC, Marchi FA, Beltrami CM, Kuasne H, Pinto CAL, et al. Prognostic classifier based on genome-wide DNA methylation profiling in well-differentiated thyroid tumors. J Clin Endocrinol Metab. 2017;102(11):4089–99.

Genomic Applications in Salivary Gland Tumors

25

Todd M. Stevens and Justin A. Bishop

Introduction

Recently, the field of salivary gland pathology has helped debunk the once-held adage that translocations are rare in epithelial malignancies [1]. In fact, one could argue that outside of hematologic and soft tissue pathology, the field of diagnostic salivary gland pathology is one of the subspecialties of medicine most defined by genetics. For example, many entities such as adenoid cystic carcinoma, mucoepidermoid carcinoma, and secretory carcinoma are now largely defined by their specific, recurrent genetic alterations, making molecular techniques indispensable to the salivary gland pathologist and treating clinician. In addition, molecular advances have led to the identification of previously unidentified entities or variants in salivary gland pathology (e.g., secretory carcinoma) and are also used to help guide personalized therapy. In this chapter, the salient molecular signatures of salivary gland neoplasms will be presented as they relate to diagnosis and therapy.

Salivary Gland Neoplasms

Pleomorphic Adenoma

Pleomorphic adenoma (PA) is the most common salivary gland tumor and can arise in both major and minor salivary gland sites. Histologically similar tumors occur in the skin and are called "cutaneous mixed tumor." PA is a circumscribed tumor with variable peripheral pseudopod-like extensions, but true invasion is not seen. Histologically, the core features of PA include ductal differentiation with luminal epithelial cells surrounded by myoepithelial cells. In any given example, the luminal or the myoepithelial cells can predominate, but focally this luminal-abluminal relationship will be found. The myoepithelial cells tend to "swirl" off the luminal cells and produce a chondromyxoid matrix which can differentiate into cartilage, adipose tissue, and even bone. Examples can consist predominantly of myxoid stroma with scant embedded myoepithelial cells, and some cases can be markedly hypercellular with minimal stroma. Variable mixtures of epithelial to spindled cells can be present. As such PA can show a remarkable histologic diversity, hence the term "pleomorphic." PAs are benign but can show local recurrences owing to their frequent pseudopodia-like growth. Long-standing PAs can undergo malignant change, i.e., carcinoma ex-pleomorphic adenoma. Carcinoma ex-PA is diagnosed when there is evidence of significant cytologic atypia, necrosis, and increased mitotic activity, with (invasive carcinoma ex-PA) or without (noninvasive carcinoma ex-PA) invasive growth. Salivary duct carcinoma is the most common histologic type of salivary gland carcinoma to arise ex-PA.

Approximately 50–70% of PAs show rearrangements of either the *PLAG1* (*pleomorphic adenoma gene 1* on 8q12) or less commonly *HMGA2* (*high mobility group A2* on 12q14-15) genes [2]. *PLAG1* appears to play a role in the pathogenesis of all pleomorphic adenomas, as the PLAG1 protein is consistently overexpressed regardless of the *PLAG1* or *HMGA2* rearrangement status [3]. Cryptic intrachromosomal rearrangements involving *PLAG1* or the adjacent *TCEA1* and *CHCHD7* genes may explain this PLAG1 overexpression in cases lacking *PLAG1* or *HMGA2* rearrangements by routine cytogenetic or FISH analysis. The *PLAG1* gene encodes a zinc-finger protein normally only expressed in fetal and embryonic tissue. PLAG1 protein has been shown to bind DNA, in particular the promoter of the *insulin-like growth factor II* gene, leading to its overexpression [4]. When rearranged, usually with the *CTNNB1* (*beta-catenin*), *LIFR* (*leukemia inhibitory factor receptor*), or *FGFR1* (*fibroblast*

T. M. Stevens (✉)
Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA
e-mail: tstevens@uabmc.edu

J. A. Bishop
Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_25

growth factor receptor I) genes as fusion partners, upregulation of *PLAG1* occurs leading to its overexpression [5].

PLAG1 rearrangements have also been identified in the cutaneous counterpart of PA, the benign mixed tumor [3].

Parenthetically, *PLAG1* rearrangements are also present in lipoblastoma, with either the *COL1A2* or *HAS2* genes as fusion partners [3, 4].

The *PLAG1* or *HMGA2* rearrangement persists in up to 86% of carcinomas ex-PA [6]. Furthermore, *PLAG1* and *HMGA2* rearrangements are not found in other de novo salivary gland carcinomas. Carcinomas ex-PA have also been shown to harbor amplification of the *MDM2* and *Her2* (*ERBB2*) genes and show deletions of 5q23.2-q31.2 and gains of *MYC*. *p53* mutations, however, may not be involved in malignant transformation of PA [3].

Because *PLAG1* and *HMGA2* rearrangements are encountered in both benign PAs and carcinomas ex-PA, the diagnostic utility of demonstrating these alterations is limited.

Basal Cell Adenoma and Adenocarcinoma

Basal cell adenoma (BCA) is a benign salivary gland neoplasm typically arising in the parotid gland that histologically is well circumscribed and grows in nests, tubular, and trabecular architectures. Neoplastic cells are bland with high nuclear to cytoplasmic ratios, giving them a basaloid appearance. Peripheral palisading and stellate reticulum-like areas are common, as is production of abundant basement membrane type material. Basal cell adenocarcinomas histologically appear very similar to BCA but show invasion. Basal cell adenocarcinoma is usually regarded as a low-grade malignancy with an excellent prognosis, carrying recurrence and metastatic rates of ~17% and 5%, respectively [5].

Forty-six percent of BCA (18/39) have been shown to carry *CTNNB1* (*beta-catenin*) mutations, with the p.I35T mutation being present in 17 of the 18 BCA cases with mutations. Membranous type basal cell adenomas are genetically related to cutaneous cylindromas, as they show alterations in the *CYLD* locus (16q12-13) implicated in the Brooke-Spiegler syndrome [5, 7]. One of four basal cell adenocarcinomas tested revealed a p.I35T *CTNNB1* mutation [7]. Activating mutations in *PIK3CA* and biallellic inactivation of *NFKB1A* have been identified in basal cell adenocarcinomas. The diagnostic role of molecular studies for diagnosing BCA or basal cell adenocarcinoma is currently very limited.

Epithelial-Myoepithelial Carcinoma

Epithelial-myoepithelial carcinoma (EMCA) is an adenocarcinoma showing two cells types: an outer myoepithelial cell layer, often with clear cytoplasm, and a tightly coupled inner luminal epithelial cell layer. They are usually of low cytologic grade, but intermediate- and high-grade examples exist. A recent study purported to show that 31 of 39 (80%) EMCAs studied may have arisen ex-pleomorphic adenoma, on the basis of the investigators identifying 12 of 39 EMCA showing histologic evidence of a preexisting PA but with intact *PLAG1* and *HMGA2* genes, 9/39 cases showing *PLAG1* rearrangements, and 10/39 cases showing *HMGA2* alterations. Eight of 39 EMCA in this study lacked histologic or molecular evidence of PA ("de novo" EMCA). This study awaits further confirmation. Up to 33% of EMCA harbor *HRAS* mutations in codon 61, although typically in cases with intact *PLAG1* and *HMGA2* [8]. Low-grade EMCA have been shown to lack *MYB* gene rearrangements [9]. Molecular diagnostics does not have a prominent role in diagnosing EMCA, as its genetic alterations overlap with other benign or malignant neoplasms.

Adenoid Cystic Carcinoma

Adenoid cystic carcinoma (ACC) is a malignant salivary gland neoplasm that arises in major and minor salivary glands as well as in the lung, skin, breast, and other sites. ACC displays a relentless clinical course and overall poor prognosis. Locally destructive growth is a common occurrence, and metastasis to the lung and other distant sites is not uncommonly seen. ACC grows in three main architectural patterns: tubular, cribriform, and solid, with cases showing >30% solid pattern being designated as high grade. Cytologically, the majority of nuclei are basophilic, angulated with inconspicuous nucleoli and typically scant cytoplasm, surrounded by a variably evident myoepithelial cell layer, making them biphasic tumors like PA and EMCA. The neoplasm produces a hyalinized and myxoid matrix and typically shows infiltrative borders with perineural invasion frequently identified [10]. At the current time, ACC are treated with surgery and radiation. ACC typically do not respond to chemotherapy [11].

Activation of the *MYB* oncogene, either by a t(6;9)(q22-23;p23-24) translocation resulting in *MYB-NFIB* gene fusion or by other mechanisms of *MYB* gene activation, is seen in about 80% of ACC. The t(6;9) fusion leads to loss of negative regulatory elements on the *MYB* gene, leading to its overexpression. *MYB* activation drives the expression of many downstream targets that are also activated by MYC signaling, leading to cellular proliferation and loss of differentiation. This ability of the *MYB-NFIB* fusion to activate downstream targets appears to occur through AKT-dependent IGFR1 signaling activation as inhibition of IGFR1 has been shown to reverse *MYB*-activated transcriptional program. While ACC has a low exonic somatic mutation rate, suggesting that *MYB* activation is the main oncogenic driver in ACC,

EGFR and MET signaling also promote growth of ACC cells, and mutations in genes involving the FGF-IGF-PI3K signaling pathway have also been identified. Hence, multiple complex signaling pathways appear to drive the growth of ACC, and therefore inhibition of multiple pathways, including, but not limited to, MYC, EGFR, MET, and IGFR1 signaling, could potentially prove effective in treating ACC [12]. In addition, NOTCH signaling pathways and genes controlling chromatin remodeling are also frequently altered in ACC [13, 14]. An *MYBL1-NFIB* t(8;9) fusion has been detected in some t(6;9)-negative ACC, and these appear to show a similar gene expression profile as *MYB*-rearranged ACC, suggesting that *MYBL1* fusion-positive ACC has similar oncogenic mechanisms to *MYB*-rearranged ACC [15, 16]. In cases of ACC showing solid patterns (so-called high-grade ACC), higher numbers of copy number alterations and losses of 1p and 6q have been identified [17].

In some cases, particularly small samples, differentiating ACC from pleomorphic adenoma, basal cell adenoma, basal cell adenocarcinoma, and other neoplasms can prove problematic. Identification of MYB protein expression by immunohistochemistry is supportive of the diagnosis of ACC, as it is seen in 82% of cases (Fig. 25.1). However, MYB labeling by immunohistochemistry is also seen in up to 14% of non-ACC salivary gland tumors [18]. Therefore, in problematic cases, identification of *MYB* gene rearrangements by FISH or RT-PCR is the gold standard for a definitive diagnosis of ACC and is a very useful tool for the surgical pathologist [16].

A recent analysis of classic, variant, and hybrid forms of epithelial-myoepithelial carcinomas revealed that 5 of 28 (18%) cases previously diagnosed as epithelial-myoepithelial carcinomas harbored *MYB* gene rearrangements. Three of these were cases containing a hybrid of typical adenoid cystic carcinoma and areas typical of epithelial-myoepithelial carcinoma (Fig. 25.2), and two resembled intermediate grade epithelial-myoepithelial carcinoma with extensive invasion. None of the 15 classic low-grade epithelial-myoepithelial carcinomas in this study contained *MYB* gene rearrangements [9]. These findings strongly suggest that adenoid cystic carcinoma may have a wider morphologic spectrum than previously appreciated, including tumors previously regarded as EMCA. This is an example of the usefulness of genomics in the classification of salivary gland tumors.

Myoepithelioma and Myoepithelial Carcinoma

Myoepithelial neoplasms, that is, tumors made up of cells with both epithelial and contractile features, arise in salivary gland, soft tissues, and rarely in bone. Cytologically, they

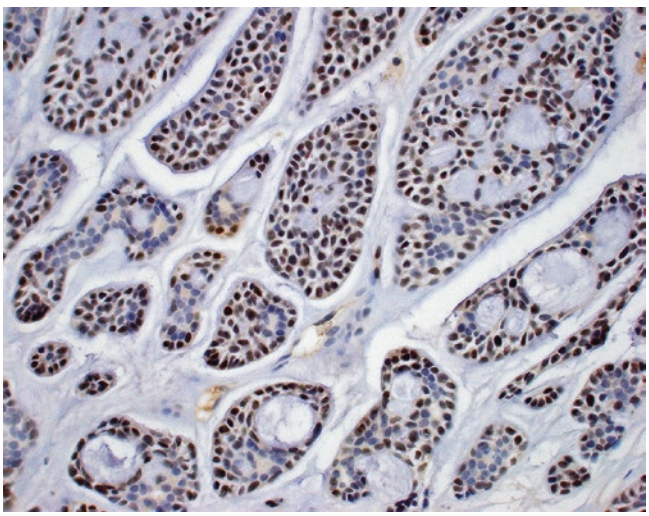


Fig. 25.1 Strong, diffuse nuclear expression of the MYB protein by immunohistochemistry is supportive of the diagnosis of adenoid cystic carcinoma. The cribriform architecture of this adenoid cystic carcinoma can be appreciated here

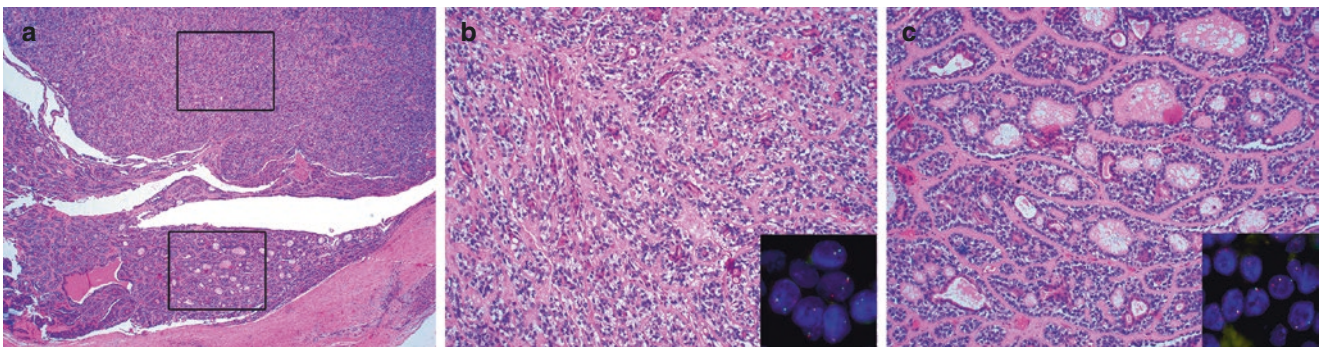


Fig. 25.2 "Hybrid" carcinoma, featuring (**a**) epithelial-myoepithelial carcinoma-like morphology (top box, **a**) and areas typical of adenoid cystic carcinoma (bottom box, **a**). Both the (**b**) epithelial-myoepithelial-like and (**c**) adenoid-cystic carcinoma components harbored *MYB* gene fusions, as assessed by *MYB* dual color break-apart FISH analysis (insets, **b** and **c**). Given such findings, some cases of apparent epithelial-myoepithelial carcinoma with infiltrative features and focal cribriforming may represent true adenoid cystic carcinoma. (From Bishop and Westra [9], Fig. 2, with permission of Wolters Kluwer Health, Inc.)

can be epithelioid, plasmacytoid, spindled, and may exhibit clear cytoplasm. Neoplastic cells often lie in a basement membrane-rich to focally myxoid matrix that they produce. Some expression of the myoepithelial markers S100 protein, calponin, smooth muscle actin, etc., in addition to keratins, p63, and EMA, is found. Diagnostic criteria making a myoepithelial neoplasm benign (myoepithelioma) or malignant (myoepithelial carcinoma) are not rigidly set, but the presence of invasion is the most important criterion for rendering a diagnosis of malignancy in this group of tumors. Other features such as cytologic atypia, necrosis, and increased mitotic activity are also useful. It should also be noted that some pleomorphic adenomas can show a preponderance of myoepithelial cells, and these cases are labeled "myoepithelial predominant pleomorphic adenomas." The presence of any component of pleomorphic adenoma would exclude a pure myoepithelial neoplasm.

While myoepithelial neoplasms of salivary glands and soft tissue share histologic features, increasing evidence suggests that they are genetically distinct. Namely, about 50% of deep-seated soft tissue myoepithelial tumors harbor rearrangement of the *EWSR1* gene, with several different partner genes identified. Deep-seated soft tissue examples can, in addition to the above-described histologic features, also show a small round blue cell phenotype and are often seen in children and young adults [19]. Soft tissue myoepithelial tumors with *EWRS1-POU5F1* fusions occur mostly in children and young adults and typically show a nested proliferation of epithelioid cells with clear cytoplasm and delicate fibrous septae. Similar *EWRS1-POU5F1* rearrangements have been found in some hidradenomas of the skin. POU5F1 expression appears limited to germ cells in the adult and is also present in human germ cell tumors. POU5F1 is involved in maintaining a stem cell state and when altered can induce divergent developmental programs, possibly explaining the epithelial and mesenchymal phenotypes seen in some myoepithelial tumors. Other partner genes in soft tissue myoepithelial tumors include *PBX1*, *PBX3*, *ZNF44*, and *KLF15* [20]. Myoepithelial neoplasms involving superficial soft tissue and skin and those that are benign are less likely to harbor *EWSR1* gene rearrangements [19].

Salivary gland myoepithelial tumors, on the other hand, mostly lack *EWRS1* rearrangements. Six myoepitheliomas in Shah et al. [21] and five myoepithelial carcinomas of salivary glands in Antonescu et al. had intact *EWSR1* genes [19]. This issue is not completely settled, however, as Skalova et al. identified *EWSR1* fusions (with an unidentified partner gene) in 39% of clear cell myoepithelial carcinomas of salivary gland [1]. Additional research is needed to determine if there is any link between deep soft tissue myoepithelial neoplasms and their salivary gland counterparts.

PLAG1 and *HMGA2* fusions, seen in pleomorphic adenomas, have not been identified in pure myoepithelial salivary gland tumors [19]. Similarly, *EWRS1* fusions have not been identified in pleomorphic adenomas. Interestingly, extraskeletal myxoid chondrosarcoma, a tumor with some histologic overlap with myoepithelial neoplasms, also shows *EWSR1* rearrangements, albeit with *NR4A3* [20].

The evolving story of myoepithelial neoplasms is therefore an excellent example showing the importance of modern molecular techniques in arriving at a correct diagnosis, as many morphologically similar entities (soft tissue myoepithelial tumors, myxoid chondrosarcoma, classic or adamantinoma-like Ewing sarcoma, etc.) can show *EWSR1* rearrangements. Therefore, accurate identification of the partner gene is critical in making a correct diagnosis.

Clear Cell Carcinoma

Clear cell carcinoma (CCC), also known as hyalinizing clear cell carcinoma, is a rare low-grade salivary gland carcinoma typically arising in intraoral sites. Histologically, there is a proliferation of cells with uniform, non-pleomorphic, low-grade nuclei with clear to focally eosinophilic cytoplasm. The cells are arranged in sheets and cords and are set in a hyalinized to desmoplastic stroma. Importantly and as an aid in the differential diagnosis with the histologically overlapping myoepithelial neoplasms, the cells of CCC do not express myoepithelial markers such as smooth muscle actin or S100 protein. Antonescu et al. identified a recurrent *EWSR1-ATF1* translocation, t(12;22)(q13;q12), in CCC, and this rearrangement has been confirmed to occur in about 80–90% of CCC, including clear cell odontogenic carcinoma, showing that these two neoplasms, if not representing the same entity, are at least closely related.

CCC is important to recognize from its mimics, especially squamous cell carcinoma, as the latter has significantly worse prognosis with different treatment. Indeed, CCC only rarely shows locoregional metastasis, and some cases may be treated with surgery alone. Mucoepidermoid carcinomas can also enter into the differential diagnosis with CCC, as CCC can show variable mucin production, but CCC lack *MAML2* rearrangements [21]. In problematic cases *EWSR1* and/or *MAML2* gene study will allow a correct diagnosis. Currently, while molecular testing serves a very useful diagnostic role in CCC, it is not theranostic.

It should be noted that approximately 45% of soft tissue myoepithelial tumors also show *EWSR1* rearrangements, albeit with different partner genes, none of which have as yet been identified in CCC, including, for example, *POU5F1*, *FUS*, *PBX1*, and *ZNF444* [1, 19, 22]. In addition, one study showed 39% of "pure" clear cell myoepithelial carcinomas, 24% of clear cell myoepithelial carcinoma ex-pleomorphic adenoma, and 1 of 11 (9%) epithelial-myoepithelial carcino-

mas to contain *EWSR1* fusions; this study did not, however, characterize the partner gene fused to *EWSR1* [1]. However, another study showed a lack of *ATF1* gene rearrangements in five epithelial-myoepithelial carcinomas and one clear cell myoepithelial carcinoma [22]. Interestingly, *EWSR1-ATF1* rearrangements can also be found in clear cell sarcoma ("melanoma of soft parts"), angiomatoid fibrous histiocytoma, and clear cell sarcoma-like tumor of the gastrointestinal tract [1, 23].

The genomics of myoepithelial carcinomas and clear cell carcinoma of salivary gland highlight the promiscuity of the *EWSR1* gene, which is also rearranged in many other tumor types, including, but not limited to, desmoplastic small round blue cell tumor, myxoid chondrosarcoma, classic and adamantinoma-like Ewing sarcoma, and some myxoid liposarcomas, albeit all with different fusion partners [20].

Mucoepidermoid Carcinoma

Mucoepidermoid carcinoma (MEC) is the most common salivary gland malignancy and involves both major and minor glands as well as the lung, intraosseous/intragnathic, thyroid, thymus, and other sites. It is characterized by a heterogeneous mixture of mucin-containing goblet cells, intermediate cells, and squamoid/epidermoid cells, with variable cystic, glandular, and solid architectures. Recurrent fusions between the *MAML2* gene (11q21) and either *CRTC1* (19p13) or *CRTC3* (15q26) genes, the former being the most common partner gene, have been described in MEC. These fusions result in the *MAML2* gene swapping out its Notch ligand-binding domain with the *CRTC1* promoter and *CREB*-binding domain, resulting in activation of the cAMP/CREB signaling pathways and disruption of NOTCH signaling. This fusion has also been shown to activate MYC transcription targets resulting in cellular proliferation and tumorigenesis and to be associated with expression of amphiregulin, a ligand for the EGFR receptor [24]. These fusions are found in most low-grade MEC and in up to 50% of high-grade MEC [25]. In a recent whole exome study of MEC, *p53* mutations were the most commonly mutated gene in MEC and were found only in intermediate- and high-grade MEC, and these cases also had a higher mutational burden than *p53* mutation-negative MEC. *POU6F2* was the second most commonly mutated gene in MEC, and this mutation was only identified in low-grade MEC [26]. *CDKN2A* and *PIK3CA* gene alterations have also been found in MEC, more commonly in high-grade MEC. *BAP1* and *BRCA1/2* alterations have been found in 20.8 and 10.5% of MEC, respectively, and *HER2* (*ERBB2*) amplification has been found in 8.3% of cases [27]. Low-grade MEC have also been shown to carry fewer copy number alterations than fusion-positive high-grade MEC [28].

Currently, the role of molecular testing in MEC is largely in its use as a diagnostic adjunct. For example, cases of MEC and metaplastic Warthin tumor can show significant overlap. Identification of a *MAML2* fusion allows for a diagnosis of mucoepidermoid carcinoma in this setting and excludes Warthin tumor and other entities that could enter into a differential diagnosis. Depending on a tumor's location, these include tumors such as glandular odontogenic cyst, secretory carcinoma, squamous cell carcinoma, salivary duct carcinoma, and adenosquamous carcinoma, all of which have been shown to lack *MAML2* gene rearrangements [29]. Indeed, *MAML2* gene testing has identified new variants of MEC known as the Warthin-like and ciliated variants which very closely mimic Warthin tumor and a benign developmental cyst, respectively [30]. It should be noted, however, that while *MAML2* rearrangements are specific for mucoepidermoid carcinoma in the salivary glands, clear cell hidradenomas of the skin have been shown to carry *CRTC1-MAML2* fusions [31].

It was initially suggested that MEC containing *MAML2* fusions were usually low grade and carried a better prognosis than fusion-negative MEC. However, fusion-positive MEC cases that are high grade resulting in death have been noted, and these cases are more likely to carry deletions in *CDKN2A* [25, 32, 33]. It has been suggested that cases previously called fusion-negative high-grade MEC may actually represent other as yet undefined higher-grade salivary gland carcinomas, thus leading to the initial suggestion that fusion-positive MEC were more indolent [25, 28, 34]. The fact remains that the utility of *MAML2* translocation status is currently more for its diagnostic value than its prognostic ability.

Secretory Carcinoma

Secretory carcinoma (SC) is a prototypical example of the discovery of an entity thanks to advanced molecular techniques. In 2010, Skalova et al. [35] described a salivary gland adenocarcinoma capable of growing as tubular, papillocystic, microcystic, cribriform, and solid architectures with low-grade nuclei with inconspicuous nucleoli and moderate to abundant eosinophilic, bubbly, and flocculent cytoplasm, often with colloid-like secretions, identical to secretory carcinoma of the breast (Fig. 25.3). Like secretory carcinoma of the breast, these cases displayed a t(12:15)(p13;q25) rearrangement resulting in the *ETV6-NTRK3* gene fusion. Parenthetically, this same fusion is identified in infantile fibrosarcoma, congenital mesoblastic nephroma, a subset of ALK-negative inflammatory myofibroblastic tumors, radiation-induced papillary thyroid carcinomas, and some hematologic malignancies [36, 37]. Rare cases of SC show *ETV6* gene rearrangements with a then-unknown fusion

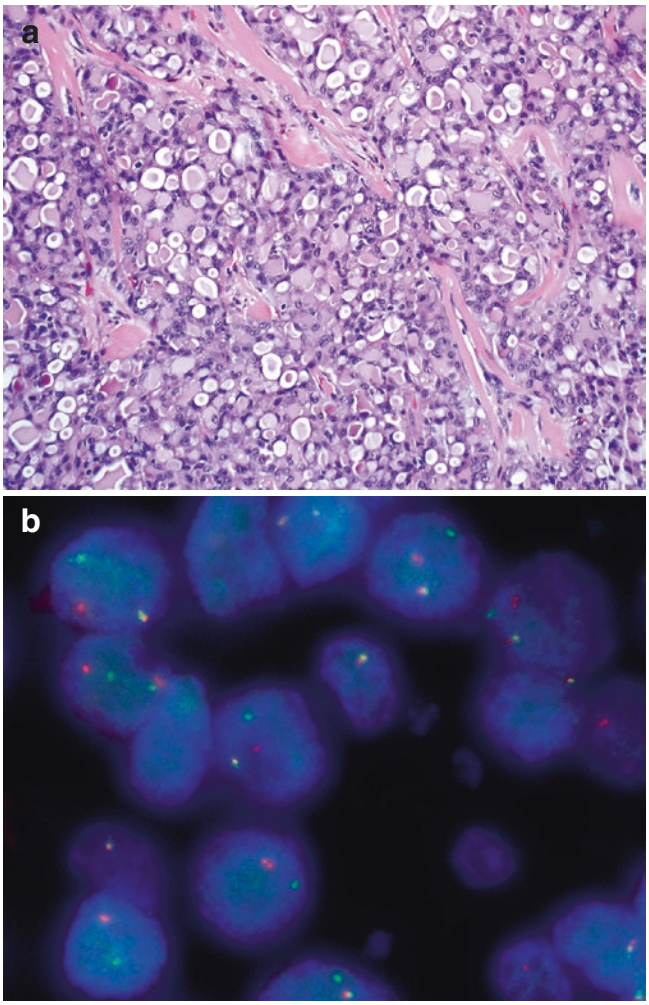


Fig. 25.3 Secretory carcinoma. (**a**) An example of secretory carcinoma showing abundant eosinophilic cytoplasm, colloid-like secretions, and low-grade nuclei with inconspicuous nucleoli. Secretory carcinomas harbor rearrangements in the *ETV6* gene, most commonly to the *NTRK* gene. (**b**) This example shows an *ETV6* fusion, as evidenced by the split signals on this *ETV6* break-apart FISH study

partner or have atypical *ETV6-NTRK3* exon fusion junctions, and these molecularly atypical SC cases often show an abundant fibrosclerotic stroma, infiltrative features, and more aggressive clinical course [38, 39]. Subsequently, a subset of secretory carcinoma cases were shown to harbor *ETV6-RET* gene fusions [40]. Prior to its recognition, SC were typically diagnosed as acinic cell carcinomas or adenocarcinoma, not otherwise specified [41].

Initially, SC was coined as "mammary analog secretory carcinoma." However, the World Health Organization (WHO) Classification of Head and Neck Tumors uses the term secretory carcinomas for these tumors, since they can occur at any site containing major and minor salivary glands, the skin, and even the thyroid [41–43].

Immunohistochemically, SC typically shows diffuse expression of S100 protein and mammaglobin and lacks significant p63 and DOG1 expression, as well as lacks *HER2* (*ERBB2*) amplifications [41]. This immunophenotype overlaps with some forms of intraductal carcinoma [41]. However, intraductal carcinomas lack *ETV6* gene rearrangements and show more prominent intraductal growth patterns, whereas if an intraductal component is present in SC, it is very focal [41, 44].

SC shows a relatively indolent clinical course with outcomes similar to acinic cell carcinoma [41, 45], with SC showing local recurrences, lymph node metastasis, and distant metastasis in approximately 15%, 20%, and 5% of cases, respectively [46]. SC are typically treated with surgery with adjuvant radiation given in some cases with aggressive features such as positive margins, high-grade features, or lymph node involvement. Occasional cases do metastasize, and systemic chemotherapy is occasionally used in these settings. It should be noted that typical SC is low-grade; however, cases exist showing high-grade transformation, typified by more pleomorphic nuclei, loss of typical architectural growth patterns, and increased mitosis and necrosis, and these cases have a high mortality rate [46].

Identification of SC and its *NTRK* fusion is important as anti-NTRK personalized therapies are a treatment option [47–49]. Anti-RET therapies are in development and may also be of use in cases of SC with *ETV6-RET* gene fusions. Therefore, not only is molecular diagnostics useful in separating SC from its histologic mimics acinic cell carcinoma, mucoepidermoid carcinoma, and polymorphous adenocarcinoma, it also may allow selection of more appropriate personalized therapy (e.g., anti-NTRK or anti-RET therapies).

Acinic Cell Carcinoma

Acinic cell carcinoma is a salivary gland carcinoma, more commonly identified in women, with architectural patterns that overlap with secretory carcinoma. That is they show variable solid, microcystic, tubular, and papillocystic growth patterns. These tumors typically have a blue low-power appearance owing to their basophilic cytoplasmic zymogen granules. Acinic cell carcinomas also show cytologic variability, as clear cells and zymogen-poor cells can be seen [41, 45].

Before its recognition as a unique, translocation defined entity, most cases of secretory carcinoma were previously diagnosed as "zymogen-poor" acinic cell carcinomas [35]. Secretory carcinomas can be separated from acinic cell carcinomas by their more uniform cellular composition, eosinophilic, bubbly cytoplasm, colloid-like secretions, and lack of zymogen granules. In addition, secretory carcinomas typically show diffuse, strong co-expression of mammaglobin and S100 protein, and DOG1 is usually negative, while

acinic cell carcinoma shows the reverse immunoprofile [41]. While there is little genomic data on acinic cell carcinomas, it is known that they lack *ETV6* and *MAML2* gene rearrangements [50]. Lastly, acinic cell carcinomas do not show *HER2* (*ERBB2*) amplifications [41].

While low-grade forms of acinic cell carcinoma and secretory carcinoma show overlapping rates of recurrence, regional metastasis, and overall prognosis, it is important to separate acinic cell carcinoma from secretory carcinoma, as the former would not respond to NTRK inhibitors [51].

Acinic cell carcinomas with high-grade transformation exist and show a more prominent nuclear pleomorphism, necrosis, and infiltration and carry a worse prognosis than their low-grade counterparts [45].

Salivary Duct Carcinoma

Salivary duct carcinomas (SDC) are highly malignant tumors characterized by an invasive adenocarcinoma growing as cribriform islands, large ductal structures with comedo necrosis, and solid sheets. SDC carry an approximately 20–40% 5-year survival rate [2, 52]. Cytologically, the neoplastic cells are high grade with prominent nucleoli and abundant eosinophilic cytoplasm imparting an apocrine appearance. Androgen receptor (AR) is almost always expressed and in some studies is required for the diagnosis. S100 protein is typically patchy to negative. Sarcomatoid, mucinous, papillary, micropapillary, and basaloid variants exist [52–54].

HER2 (*ERBB2*) gene amplification occurs in about 20–30% of cases [54–56], and it appears that the MAPK pathway is important in driving SDC, as mutations in the *BRAF*, *HRAS*, and *NF1* genes are commonly detected [57]. In addition, *PIK3CA* and *TP53* are frequently mutated, and *PTEN* is often lost, in SDCs [58, 59].

NCOA4-RET fusions have been described in two cases of salivary duct carcinoma [57], although it should be noted that these two tumors were not well described pathologically and one case showed just focal expression of AR, and therefore it is not entirely clear if these two reported cases carrying the *NCOA4-RET* fusion are true apocrine-type high-grade salivary duct carcinomas or instead represent intercalated-type intraductal carcinomas (see below). Single examples each of *ETV6-NTRK3*, *BCL6-TRADD*, and *ABL1-PPP2R2C* have been identified in tumors purported to be SDC [60]. SDC is the most common carcinoma to arise in pleomorphic adenomas, and most of these cases contain *PLAG1* and *HMGA2* gene rearrangements [60, 61].

Currently, the SDC are treated with surgery, radiation, and possible chemotherapy. The previously mentioned cases with *RET* rearrangements did show a response to anti-RET therapies [57]. Anti-HER2 therapies have shown some modest objective response in treating SDC, although mutations in *PTEN*, *TP53*, and *HRAS* may decrease the effectiveness of these therapies [58, 59, 62–65]. Lastly, androgen deprivation therapy, given frequent AR expression by these tumors, is attempted in some cases and has shown efficacy [66, 67].

Intraductal Carcinoma

Intraductal carcinoma is the new term endorsed by the WHO for lesions previously listed under various rubrics including low-grade salivary duct carcinoma, low-grade cribriform cystadenocarcinoma, and salivary duct carcinoma in situ. Intraductal carcinoma of salivary gland is now understood to occur in two forms, an intercalated duct type and an apocrine type [44]. The former resembles atypical ductal hyperplasia or low-grade ductal carcinoma in situ of the breast and as such is an intraductal proliferation of cells containing small, bland nuclei, inconspicuous nucleoli, and variable eosinophilic to amphophilic cytoplasm. The cells are arranged in micropapillary, papillary, fenestrated, and cribriform architectures and show S100 protein and mammaglobin expression, with an absence of AR expression and lack of *HER2* (*ERBB2*) gene amplification [41]. In rare cases, the cells of the intercalated type of intraductal carcinoma may breach the myoepithelial cells of the duct to show locally invasive growth. Even when showing areas of invasion, the intercalated type of intraductal carcinoma shows an excellent prognosis [41, 44].

The apocrine type of intraductal carcinoma is made up of cells with prominent central nucleoli and abundant eosinophilic cytoplasm with apocrine-type secretions and as such resembles conventional (high-grade) salivary duct carcinoma. Unlike the intercalated type, the apocrine type is positive for AR and typically negative for S100 protein. It is currently not clear what the genetic relationship is, if any, between the intercalated type and apocrine types of intraductal carcinomas or if either gives rise to conventional salivary duct carcinomas. That being said, cases of intraductal carcinoma with mixed intercalated- and apocrine-type features [44], as well as rare cases of intercalated-type intraductal carcinomas with areas of in situ and invasive disease showing higher-grade cytology, have been reported [68].

A recent study by Weinreb et al. [44] analyzed several pure intraductal carcinomas of intercalated type as well as cases of invasive apocrine-type salivary duct carcinomas containing a significant intraductal component with next-generation sequencing and FISH techniques. A novel *NCOA4-RET* fusion was identified in a case showing pure low-grade intercalated-type morphology both in the

intraductal and invasive components. Six other cases of pure low-grade intraductal, intercalated-type tumors showed *RET* gene rearrangements by FISH, for a total of 47% of low-grade intercalated-type intraductal carcinomas showing *RET* gene rearrangements in this study; further, none of these cases harbored *ETV6* gene rearrangement, eliminating the possibility that these cases represented secretory carcinomas.

Intraductal carcinomas have shown a lack of *ETV6*, *ROS-1*, *ALK-1*, *PLAG*, and *HMGA2* rearrangements [2, 41, 44]. In the Weinreb study, the apocrine-type intraductal carcinomas with low to intermediate nuclear grade with apocrine-type invasive areas showed *PIK3CA* and *HRAS* mutations, in the absence of *RET* rearrangements, with poor clinical outcomes, and thus may be closely related to conventional salivary duct carcinoma.

Future studies on the low-grade, intercalated type of intraductal carcinomas, with or without invasion, and those intraductal carcinomas with higher-grade apocrine features are needed to determine the relationship, if any, of both of these clinicopathologic spectrums of disease with conventional salivary duct carcinoma, and with each other, especially since a *NCOA4-RET* fusion was reportedly detected in a conventional salivary duct carcinoma [57]. Teasing apart the possible presence of *RET* gene rearrangements in both the low-grade intercalated type of intraductal carcinoma and conventional salivary duct carcinomas will be important as personalized therapies directed at the *RET* fusion would not be necessary in the former, a neoplasm with excellent prognosis, but may be useful in the latter, if they are in fact present [44].

Polymorphous Adenocarcinoma and Cribriform Adenocarcinoma of Minor Salivary Glands

Polymorphous adenocarcinoma (PLA) and cribriform adenocarcinoma of minor salivary glands (CASG) are two entities that share histologic overlap as well as *PRKD* gene alterations, albeit *PRKD* mutations in the former and rearrangements in the latter [69]. PLA typically involves the palate and histologically shows cytologically uniform cells arranged in many architectural patterns, most commonly streaming fascicles exhibiting prominent whorled, "targetoid" growth around nerves, but also solid, cribriform, tubular, and occasionally papillary patterns. PLA shows vesicular nuclei with scant cytoplasm and production of a myxohyaline matrix, overall giving PLA a light blue lower power hue. CASG involves minor salivary gland sites such as the base of the tongue, palate, lip, retromolar trigone, etc. Unlike PLA, however, CASG shows more prominent solid and cribriform architecture along with cleft-like spaces forming glomeruloid formations and lacks the fascicular, whorled growth typical of PLA to any significant degree. Furthermore, the nuclei of CASG show a striking similarity to papillary thyroid carcinoma, much more so than PLA. Ultrastructurally, CASG has a myosecretory phenotype, with both microvilli and microfilaments identified [70]. While PLA almost never shows regional lymph node metastasis, CASG frequently presents with regional lymph node metastasis. Despite this, CASG shares the same excellent outcome as PLA [71].

Recently, about 80% of cases morphologically classified as CASG harbored either rearrangements of *PRKD1*, *PRKD2*, or *PRKD3*. *ARID1A* and the *DDX3X* genes were identified as partners with *PRKD1* in two cases, respectively. CASG lacks mutations in *BRAF*, *KRAS*, *NRAS*, *HRAS*, *RET*, *C-KIT*, and *PDGFRa* genes [72, 73]. An E710D hotspot mutation in *PRKD1* has been identified in about 70% of PLA, but not in other salivary gland tumor types [74]. Mutations in *PRKD2* or *PRKD3* have not been identified in PLA [75]. The *PRKD1*, *PRKD2*, and *PRKD3* genes share similar serine-threonine kinase activity involved in the diacylglycerol and protein kinase C pathways, regulating cell adhesion, migration, and survival [69, 74].

As both neoplasms can co-express S100 protein and mammaglobin, they could be confused with secretory carcinoma. However, careful morphologic exam and, if needed, molecular studies for *ETV6* and/or the *PRKD* genes will allow a correct diagnosis, as PLA/CASG lack *ETV6* rearrangements [76]. That CASG lacks expression of TTF-1 and thyroglobulin and harbors *PRKD1-3* rearrangements [73] allows its diagnostic distinction from metastatic papillary thyroid carcinoma.

In the current WHO Classification of Head and Neck Tumors, CASG is listed as a subtype of PLA. However, it is currently up for debate if PLA and CASG represent distinct entities or exist on a spectrum [77]. Arguments that these entities exist on a spectrum would include that cases showing intermediate morphology are not uncommon and that about 45% of such intermediate cases harbor *PRKD* rearrangements, and a PLA with a *PRKD2* rearrangement has been identified [69].

Conclusions

This chapter has summarized the current state of the art of the application of modern molecular techniques as it relates to diagnosis, treatment, and taxonomic classification of salivary gland tumors. In addition, it has shown how the study of molecular alterations in salivary gland pathology has led to the development of new therapeutic protocols for treating salivary gland cancer. NTRK inhibitor therapies in secretory carcinomas are just one example of many. The story is only beginning, however, as molecular study of salivary gland tumors in the ensuing years will likely continue to uncover new entities and clarify relationships between known entities,

for example, the relationship, if any, between intraductal carcinoma and salivary duct carcinoma. Molecular diagnostics will also inevitably result in a gradual reduction in cases that would have previously been assigned to the wastebasket "adenocarcinoma, not otherwise specified" category. The major genetic alterations identified in salivary gland tumors to date are summarized in Table 25.1.

Table 25.1 Summary of major genetic alterations identified in salivary gland tumors to date

Salivary gland tumor	Major genetic alterations
Pleomorphic adenoma	*PLAG1* and *HMGA2* rearrangements
Carcinoma ex pleomorphic adenoma	Majority retain *PLAG1* or *HMGA2* rearrangements
Basal cell adenoma	*CTTNB1* (*beta-catenin*) mutations (p.I35T) in about 46% of cases *CYLD* locus (16q12-13) alterations in membranous examples
Basal cell adenocarcinoma	1 of 4 cases tested showed a *CTTNB1* (*beta-catenin*) mutations (p.I35T) Activating mutations in *PIK3CA* and biallelic inactivation of *NFKB1A*
Epithelial-myoepithelial carcinoma	*PLAG1* and *HMGA2* rearrangements in cases arising ex-pleomorphic adenoma *HRAS* mutations common in cases with intact *PLAG1* and *HMGA2*
Adenoid cystic carcinoma	t(6;9)(q22-23;p23-24) resulting in an *MYB-NFIB* gene fusion and other mechanisms of *MYB* oncogene in ~80% of cases Less commonly a t(8;9) resulting in *MYBL1-NFIB* gene fusion
Myoepithelial neoplasms	*Salivary gland origin:* Largely unknown, but *EWSR1* fusions can be seen in clear cell myoepithelial carcinomas *Soft tissue origin: EWSR1* rearrangement in about 50%, with either *POU5F1*, *ZNF44*, *PBX1, PBX3*, *KLF15*
Clear cell carcinoma	t(12;22)(q13;q12) *EWSR1-ATF1* in 80–90%
Mucoepidermoid carcinoma	t(11;19)(q21;p13) *CRTC1-MAML2* t(11;15)(q21;q26) *CRTC3-MAML2* *p53* is most commonly mutated gene
Secretory carcinoma	*ETV6-NTRK* (~95–98%) *ETV6-RET* (~2–5%)
Acinic cell carcinoma	No specific abnormality
Salivary duct carcinoma	*HER2* gene amplifications and mutations in *TP53*, *PIK3CA,* and *HRAS.* Loss of or mutations in *PTEN* *PLAG1* and *HMGA2* rearrangements in cases that are ex-pleomorphic adenoma
Intraductal carcinoma	*RET* rearrangements in the intercalated type *PIK3CA* and *HRAS* mutations in the apocrine type
Polymorphous adenocarcinoma	Somatic hotspot activating point mutation (E710D) in *PRKD1* (14q12)
Cribriform adenocarcinoma of salivary glands	*ARID1A-PRKD1* t(1;14)(p36.11;q12) *DDX3X-PRKD1* t(X;14)(p11.4;q12) *PRKD2* and *PRKD3* rearrangements

References

1. Skalova A, Weinreb I, Hyrcza M, Simpson RH, Laco J, Agaimy A, et al. Clear cell myoepithelial carcinoma of salivary glands showing EWSR1 rearrangement: molecular analysis of 94 salivary gland carcinomas with prominent clear cell component. Am J Surg Pathol. 2015;39(3):338–48.
2. Bahrami A, Perez-Ordonez B, Dalton JD, Weinreb I. An analysis of PLAG1 and HMGA2 rearrangements in salivary duct carcinoma and examination of the role of precursor lesions. Histopathology. 2013;63(2):250–62.
3. Bahrami A, Dalton JD, Shivakumar B, Krane JF. PLAG1 alteration in carcinoma ex pleomorphic adenoma: immunohistochemical and fluorescence in situ hybridization studies of 22 cases. Head Neck Pathol. 2012;6(3):328–35.
4. Voz ML, Agten NS, Van de Ven WJ, Kas K. PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. Cancer Res. 2000;60(1):106–13.
5. Jo VY, Sholl LM, Krane JF. Distinctive patterns of CTNNB1 (beta-catenin) alterations in salivary gland basal cell adenoma and basal cell adenocarcinoma. Am J Surg Pathol. 2016;40(8):1143–50.
6. Katabi N, Ghossein R, Ho A, Dogan S, Zhang L, Sung YS, et al. Consistent PLAG1 and HMGA2 abnormalities distinguish carcinoma ex-pleomorphic adenoma from its de novo counterparts. Hum Pathol. 2015;46(1):26–33.
7. Sato M, Yamamoto H, Hatanaka Y, Nishijima T, Jiromaru R, Yasumatsu R, et al. Wnt/beta-catenin signal alteration and its diagnostic utility in basal cell adenoma and histologically similar tumors of the salivary gland. Pathol Res Pract. 2018;214(4):586–92.
8. El Hallani S, Udager AM, Bell D, Fonseca I, Thompson LDR, Assaad A, et al. Epithelial-myoepithelial carcinoma: frequent morphologic and molecular evidence of preexisting pleomorphic adenoma, common HRAS mutations in PLAG1-intact and HMGA2-intact cases, and occasional TP53, FBXW7, and SMARCB1 alterations in high-grade cases. Am J Surg Pathol. 2018;42(1):18–27.
9. Bishop AJ, Westra HW. MYB translocation status in salivary gland epithelial-myoepithelial carcinoma: evaluation of classic, variant, and hybrid forms. Am J Surg Pathol. 2018;42(3):319–25.
10. Coca-Pelaz A, Rodrigo JP, Bradley PJ, Vander Poorten V, Triantafyllou A, Hunt JL, et al. Adenoid cystic carcinoma of the head and neck– an update. Oral Oncol. 2015;51(7):652–61.
11. Carlson J, Licitra L, Locati L, Raben D, Persson F, Stenman G. Salivary gland cancer: an update on present and emerging therapies. Am Soc Clin Oncol Educ Book. 2013;33:257–63.
12. Andersson MK, Afshari MK, Andren Y, Wick MJ, Stenman G. Targeting the oncogenic transcriptional regulator MYB in adenoid cystic carcinoma by inhibition of IGF1R/AKT signaling. J Natl Cancer Inst. 2017;109(9) https://doi.org/10.1093/jnci/djx017.
13. Rettig EM, Talbot CC Jr, Sausen M, Jones S, Bishop JA, Wood LD, et al. Whole-genome sequencing of salivary gland adenoid cystic carcinoma. Cancer Prev Res. 2016;9(4):265–74.
14. Stephens PJ, Davies HR, Mitani Y, Van Loo P, Shlien A, Tarpey PS, et al. Whole exome sequencing of adenoid cystic carcinoma. J Clin Invest. 2013;123(7):2965–8.
15. Brayer KJ, Frerich CA, Kang H, Ness SA. Recurrent fusions in MYB and MYBL1 define a common, transcription factor-driven oncogenic pathway in salivary gland adenoid cystic carcinoma. Cancer Discov. 2016;6(2):176–87.
16. Mitani Y, Liu B, Rao PH, Borra VJ, Zafereo M, Weber RS, et al. Novel MYBL1 gene rearrangements with recurrent MYBL1-NFIB fusions in salivary adenoid cystic carcinomas lacking t(6;9) translocations. Clin Cancer Res. 2016;22(3):725–33.
17. Persson M, Andren Y, Moskaluk CA, Frierson HF Jr, Cooke SL, Futreal PA, et al. Clinically significant copy number alterations and complex rearrangements of MYB and NFIB in head and

neck adenoid cystic carcinoma. Genes Chromosomes Cancer. 2012;51(8):805–17.
18. Brill LB 2nd, Kanner WA, Fehr A, Andren Y, Moskaluk CA, Loning T, et al. Analysis of MYB expression and MYB-NFIB gene fusions in adenoid cystic carcinoma and other salivary neoplasms. Mod Pathol. 2011;24(9):1169–76.
19. Antonescu CR, Zhang L, Chang NE, Pawel BR, Travis W, Katabi N, et al. EWSR1-POU5F1 fusion in soft tissue myoepithelial tumors. A molecular analysis of sixty-six cases, including soft tissue, bone, and visceral lesions, showing common involvement of the EWSR1 gene. Genes Chromosomes Cancer. 2010;49(12):1114–24.
20. Stevens TM, Qarmali M, Morlote D, Mikhail FM, Swensen J, Gatalica Z, et al. Malignant Ewing-like neoplasm with an EWSR1-KLF15 fusion: at the crossroads of a myoepithelial carcinoma and a Ewing-like sarcoma. A case report with treatment options. Int J Surg Pathol. 2018;26(5):440–7. https://doi.org/10.1177/1066896918755009.
21. Shah AA, LeGallo RD, van Zante A, Frierson HF Jr, Mills SE, Berean KW, et al. EWSR1 genetic rearrangements in salivary gland tumors: a specific and very common feature of hyalinizing clear cell carcinoma. Am J Surg Pathol. 2013;37(4):571–8.
22. Antonescu CR, Katabi N, Zhang L, Sung YS, Seethala RR, Jordan RC, et al. EWSR1-ATF1 fusion is a novel and consistent finding in hyalinizing clear-cell carcinoma of salivary gland. Genes Chromosomes Cancer. 2011;50(7):559–70.
23. Zucman J, Delattre O, Desmaze C, Epstein AL, Stenman G, Speleman F, et al. EWS and ATF-1 gene fusion induced by t(12;22) translocation in malignant melanoma of soft parts. Nat Genet. 1993;4(4):341–5.
24. Shinomiya H, Ito Y, Kubo M, Yonezawa K, Otsuki N, Iwae S, et al. Expression of amphiregulin in mucoepidermoid carcinoma of the major salivary glands: a molecular and clinicopathological study. Hum Pathol. 2016;57:37–44.
25. Seethala RR, Dacic S, Cieply K, Kelly LM, Nikiforova MN. A reappraisal of the MECT1/MAML2 translocation in salivary mucoepidermoid carcinomas. Am J Surg Pathol. 2010;34(8):1106–21.
26. Kang H, Tan M, Bishop JA, Jones S, Sausen M, Ha PK, et al. Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. Clin Cancer Res. 2017;23(1):283–8.
27. Wang K, McDermott JD, Schrock AB, Elvin JA, Gay L, Karam SD, et al. Comprehensive genomic profiling of salivary mucoepidermoid carcinomas reveals frequent BAP1, PIK3CA, and other actionable genomic alterations. Ann Oncol. 2017;28(4):748–53.
28. Jee KJ, Persson M, Heikinheimo K, Passador-Santos F, Aro K, Knuutila S, et al. Genomic profiles and CRTC1-MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. Mod Pathol. 2013;26(2):213–22.
29. Bishop JA, Yonescu R, Batista D, Warnock GR, Westra WH. Glandular odontogenic cysts (GOCs) lack MAML2 rearrangements: a finding to discredit the putative nature of GOC as a precursor to central mucoepidermoid carcinoma. Head Neck Pathol. 2014;8(3):287–90.
30. Bishop JA, Cowan ML, Shum CH, Westra WH. MAML2 rearrangements in variant forms of mucoepidermoid carcinoma: ancillary diagnostic testing for the ciliated and Warthin-like variants. Am J Surg Pathol. 2018;42(1):130–6.
31. Behboudi A, Winnes M, Gorunova L, van den Oord JJ, Mertens F, Enlund F, et al. Clear cell hidradenoma of the skin-a third tumor type with a t(11;19)--associated TORC1-MAML2 gene fusion. Genes Chromosomes Cancer. 2005;43(2):202–5.
32. Behboudi A, Enlund F, Winnes M, Andren Y, Nordkvist A, Leivo I, et al. Molecular classification of mucoepidermoid carcinomas-prognostic significance of the MECT1-MAML2 fusion oncogene. Genes Chromosomes Cancer. 2006;45(5):470–81.
33. Anzick SL, Chen WD, Park Y, Meltzer P, Bell D, El-Naggar AK, et al. Unfavorable prognosis of CRTC1-MAML2 positive mucoepidermoid tumors with CDKN2A deletions. Genes Chromosomes Cancer. 2010;49(1):59–69.
34. Chiosea SI, Dacic S, Nikiforova MN, Seethala RR. Prospective testing of mucoepidermoid carcinoma for the MAML2 translocation: clinical implications. Laryngoscope. 2012;122(8):1690–4.
35. Skalova A, Vanecek T, Sima R, Laco J, Weinreb I, Perez-Ordonez B, et al. Mammary analogue secretory carcinoma of salivary glands, containing the ETV6-NTRK3 fusion gene: a hitherto undescribed salivary gland tumor entity. Am J Surg Pathol. 2010;34(5):599–608.
36. Alassiri AH, Ali RH, Shen Y, Lum A, Strahlendorf C, Deyell R, et al. ETV6-NTRK3 is expressed in a subset of ALK-negative inflammatory myofibroblastic tumors. Am J Surg Pathol. 2016;40(8):1051–61.
37. Leeman-Neill RJ, Kelly LM, Liu P, Brenner AV, Little MP, Bogdanova TI, et al. ETV6-NTRK3 is a common chromosomal rearrangement in radiation-associated thyroid cancer. Cancer. 2014;120(6):799–807.
38. Skalova A, Vanecek T, Simpson RH, Laco J, Majewska H, Baneckova M, et al. Mammary analogue secretory carcinoma of salivary glands: molecular analysis of 25 ETV6 gene rearranged tumors with lack of detection of classical ETV6-NTRK3 fusion transcript by standard RT-PCR: report of 4 cases harboring ETV6-X gene fusion. Am J Surg Pathol. 2016;40(1):3–13.
39. Ito Y, Ishibashi K, Masaki A, Fujii K, Fujiyoshi Y, Hattori H, et al. Mammary analogue secretory carcinoma of salivary glands: a clinicopathologic and molecular study including 2 cases harboring ETV6-X fusion. Am J Surg Pathol. 2015;39(5):602–10.
40. Skalova MA, Vanecek HWT, Martinek JP, Weinreb JI, Stevens JT, Simpson JR, et al. Molecular profiling of mammary analog secretory carcinoma revealed a subset of tumors harboring a novel ETV6-RET translocation: report of 10 cases. Am J Surg Pathol. 2018;42(2):234–46.
41. Stevens TM, Kovalovsky AO, Velosa C, Shi Q, Dai Q, Owen RP, et al. Mammary analog secretory carcinoma, low-grade salivary duct carcinoma, and mimickers: a comparative study. Mod Pathol. 2015;28(8):1084–100.
42. Dettloff J, Seethala RR, Stevens TM, Brandwein-Gensler M, Centeno BA, Otto K, et al. Mammary analog secretory carcinoma (MASC) involving the thyroid gland: a report of the first 3 cases. Head Neck Pathol. 2017;11(2):124–30.
43. Bishop JA, Taube JM, Su A, Binder SW, Kazakov DV, Michal M, et al. Secretory carcinoma of the skin harboring ETV6 gene fusions: a cutaneous analogue to secretory carcinomas of the breast and salivary glands. Am J Surg Pathol. 2017;41(1):62–6.
44. Weinreb I, Bishop JA, Chiosea SI, Seethala RR, Perez-Ordonez B, Zhang L, et al. Recurrent RET gene rearrangements in intraductal carcinomas of salivary gland. Am J Surg Pathol. 2018;42(4):442–52.
45. Chiosea SI, Griffith C, Assaad A, Seethala RR. The profile of acinic cell carcinoma after recognition of mammary analog secretory carcinoma. Am J Surg Pathol. 2012;36(3):343–50.
46. Skalova A, Vanecek T, Majewska H, Laco J, Grossmann P, Simpson RH, et al. Mammary analogue secretory carcinoma of salivary glands with high-grade transformation: report of 3 cases with the ETV6-NTRK3 gene fusion and analysis of TP53, beta-catenin, EGFR, and CCND1 genes. Am J Surg Pathol. 2014;38(1):23–33.
47. Chi HT, Ly BT, Kano Y, Tojo A, Watanabe T, Sato Y. ETV6-NTRK3 as a therapeutic target of small molecule inhibitor PKC412. Biochem Biophys Res Commun. 2012;429(1–2):87–92.
48. Tognon CE, Somasiri AM, Evdokimova VE, Trigo G, Uy EE, Melnyk N, et al. ETV6-NTRK3-mediated breast epithelial cell transformation is blocked by targeting the IGF1R signaling pathway. Cancer Res. 2011;71(3):1060–70.
49. Drilon A, Li G, Dogan S, Gounder M, Shen R, Arcila M, et al. What hides behind the MASC: clinical response and acquired

resistance to entrectinib after ETV6-NTRK3 identification in a mammary analogue secretory carcinoma (MASC). Ann Oncol. 2016;27(5):920–6.
50. Lei Y, Chiosea SI. Re-evaluating historic cohort of salivary acinic cell carcinoma with new diagnostic tools. Head Neck Pathol. 2012;6(2):166–70.
51. Skalova A, Michal M, Simpson RH. Newly described salivary gland tumors. Mod Pathol. 2017;30(s1):S27–s43.
52. Simpson RH. Salivary duct carcinoma: new developments--morphological variants including pure in situ high grade lesions; proposed molecular classification. Head Neck Pathol. 2013;7(Suppl 1):S48–58.
53. Simpson RH, Skalova A, Di Palma S, Leivo I. Recent advances in the diagnostic pathology of salivary carcinomas. Virchows Arch. 2014;465(4):371–84.
54. Masubuchi T, Tada Y, Maruya S, Osamura Y, Kamata SE, Miura K, et al. Clinicopathological significance of androgen receptor, HER2, Ki-67 and EGFR expressions in salivary duct carcinoma. Int J Clin Oncol. 2015;20(1):35–44.
55. Di Palma S, Simpson RH, Marchio C, Skalova A, Ungari M, Sandison A, et al. Salivary duct carcinomas can be classified into luminal androgen receptor-positive, HER2 and basal-like phenotypes. Histopathology. 2012;61(4):629–43.
56. Skalova A, Starek I, Vanecek T, Kucerova V, Plank L, Szepe P, et al. Expression of HER-2/neu gene and protein in salivary duct carcinomas of parotid gland as revealed by fluorescence in-situ hybridization and immunohistochemistry. Histopathology. 2003;42(4):348–56.
57. Wang K, Russell JS, McDermott JD, Elvin JA, Khaira D, Johnson A, et al. Profiling of 149 salivary duct carcinomas, carcinoma ex pleomorphic adenomas, and adenocarcinomas, not otherwise specified reveals actionable genomic alterations. Clin Cancer Res. 2016;22(24):6061–8.
58. Seethala RR, Griffith CC. Molecular pathology: predictive, prognostic, and diagnostic markers in salivary gland tumors. Surg Pathol Clin. 2016;9(3):339–52.
59. Griffith CC, Seethala RR, Luvison A, Miller M, Chiosea SI. PIK3CA mutations and PTEN loss in salivary duct carcinomas. Am J Surg Pathol. 2013;37(8):1201–7.
60. Dalin MG, Desrichard A, Katabi N, Makarov V, Walsh LA, Lee KW, et al. Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast Cancer. Clin Cancer Res. 2016;22(18):4623–33.
61. Chiosea SI, Thompson LD, Weinreb I, Bauman JE, Mahaffey AM, Miller C, et al. Subsets of salivary duct carcinoma defined by morphologic evidence of pleomorphic adenoma, PLAG1 or HMGA2 rearrangements, and common genetic alterations. Cancer. 2016;122(20):3136–44.
62. Falchook GS, Lippman SM, Bastida CC, Kurzrock R. Human epidermal receptor 2-amplified salivary duct carcinoma: regression with dual human epidermal receptor 2 inhibition and anti-vascular endothelial growth factor combination treatment. Head Neck. 2014;36(3):E25–7.
63. Limaye SA, Posner MR, Krane JF, Fonfria M, Lorch JH, Dillon DA, et al. Trastuzumab for the treatment of salivary duct carcinoma. Oncologist. 2013;18(3):294–300.
64. Nardi V, Sadow PM, Juric D, Zhao D, Cosper AK, Bergethon K, et al. Detection of novel actionable genetic changes in salivary duct carcinoma helps direct patient treatment. Clin Cancer Res. 2013;19(2):480–90.
65. Qiu W, Tong GX, Turk AT, Close LG, Caruana SM, Su GH. Oncogenic PIK3CA mutation and dysregulation in human salivary duct carcinoma. Biomed Res Int. 2014;2014:810487.
66. Soper MS, Iganej S, Thompson LD. Definitive treatment of androgen receptor-positive salivary duct carcinoma with androgen deprivation therapy and external beam radiotherapy. Head Neck. 2014;36(1):E4–7.
67. Delgado R, Klimstra D, Albores-Saavedra J. Low grade salivary duct carcinoma. A distinctive variant with a low grade histology and a predominant intraductal growth pattern. Cancer. 1996;78(5):958–67.
68. Brandwein-Gensler M, Hille J, Wang BY, Urken M, Gordon R, Wang LJ, et al. Low-grade salivary duct carcinoma: description of 16 cases. Am J Surg Pathol. 2004;28(8):1040–4.
69. Weinreb I, Zhang L, Tirunagari LM, Sung YS, Chen CL, Perez-Ordonez B, et al. Novel PRKD gene rearrangements and variant fusions in cribriform adenocarcinoma of salivary gland origin. Genes Chromosomes Cancer. 2014;53(10):845–56.
70. Michal M, Skalova A, Simpson RH, Raslan WF, Curik R, Leivo I, et al. Cribriform adenocarcinoma of the tongue: a hitherto unrecognized type of adenocarcinoma characteristically occurring in the tongue. Histopathology. 1999;35(6):495–501.
71. Michal M, Kacerovska D, Kazakov DV. Cribriform adenocarcinoma of the tongue and minor salivary glands: a review. Head Neck Pathol. 2013;7(Suppl 1):S3–11.
72. Skalova A, Sima R, Kaspirkova-Nemcova J, Simpson RH, Elmberger G, Leivo I, et al. Cribriform adenocarcinoma of minor salivary gland origin principally affecting the tongue: characterization of new entity. Am J Surg Pathol. 2011;35(8):1168–76.
73. Laco J, Kamaradova K, Vitkova P, Sehnalkova E, Dvorakova S, Vaclavikova E, et al. Cribriform adenocarcinoma of minor salivary glands may express galectin-3, cytokeratin 19, and HBME-1 and contains polymorphisms of RET and H-RAS proto-oncogenes. Virchows Arch. 2012;461(5):531–40.
74. Weinreb I, Piscuoglio S, Martelotto LG, Waggott D, Ng CK, Perez-Ordonez B, et al. Hotspot activating PRKD1 somatic mutations in polymorphous low-grade adenocarcinomas of the salivary glands. Nat Genet. 2014;46(11):1166–9.
75. Piscuoglio S, Fusco N, Ng CK, Martelotto LG, da Cruz PA, Katabi N, et al. Lack of PRKD2 and PRKD3 kinase domain somatic mutations in PRKD1 wild-type classic polymorphous low-grade adenocarcinomas of the salivary gland. Histopathology. 2016;68(7):1055–62.
76. Patel KR, Solomon IH, El-Mofty SK, Lewis JS Jr, Chernock RD. Mammaglobin and S-100 immunoreactivity in salivary gland carcinomas other than mammary analogue secretory carcinoma. Hum Pathol. 2013;44(11):2501–8.
77. Xu B, Aneja A, Ghossein R, Katabi N. Predictors of outcome in the phenotypic Spectrum of polymorphous low-grade adenocarcinoma (PLGA) and cribriform adenocarcinoma of salivary gland (CASG): a retrospective study of 69 patients. Am J Surg Pathol. 2016;40(11):1526–37.

Genomic Applications in Breast Carcinoma

26

Fresia Pareja, Leticia De Mattos-Arruda, Britta Weigelt, and Jorge S. Reis-Filho

Introduction

Advances in molecular technologies have led to a paradigm shift in the way we define breast cancer, resulting in the transition from purely morphological classification systems to combined histologic and molecular taxonomies. The advent of massively parallel sequencing has solidified the notion that breast cancer comprises multiple diseases with different biology, clinico-pathological features, natural history, and response to therapy [1–5]. Moreover, microarray-based gene expression profiling has led to the implementation of a molecular classification of breast cancer [1] and to the development of prognostic gene signatures, some of which have now been incorporated into clinical practice [6].

Molecular studies have shed light into the vast tumor heterogeneity of breast cancer, illustrated by the dissimilar genetic makeup of primary tumors and metastatic foci [7]. While the complexity and heterogeneity of breast cancer poses significant diagnostic and therapeutic challenges, it also provides opportunities for the realization of the potentials of precision medicine [8]. Novel strategies, such as the implementation of liquid biopsies, are being developed to overcome these diagnostic and therapeutic hurdles.

In this chapter, we will discuss the key contributions of molecular pathology in the dissection of the biology of breast cancer, focusing on the role of gene expression profiling and massively parallel sequencing in the classification and prognostication of the disease. We will contextualize the diagnostic and therapeutic challenges posed by breast cancer heterogeneity, as well as strategies to overcome them.

F. Pareja · B. Weigelt · J. S. Reis-Filho (✉)
Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
e-mail: reisfilj@mskcc.org

L. De Mattos-Arruda
Department of Medical Oncology, Vall d'Hebron Institute of Oncology, Vall d'Hebron University Hospital, Barcelona, Spain

Universitat Autònoma de Barcelona, Barcelona, Spain

Molecular Classification of Breast Cancer

Gene expression studies have solidified the notion that breast cancer should not be regarded as a single disease but rather as a group of entities with different molecular landscapes and clinical outcomes. Pioneering microarray-based gene expression profiling has led to the development of a breast cancer classification comprising five "intrinsic" molecular subtypes, namely, luminal A, luminal B, HER2 (also known as HER2-enriched), basal-like, and normal-like [1]. The "intrinsic" molecular classification has made it evident that ER-positive and ER-negative breast cancer are essentially different diseases at the transcriptomic level [1–3, 5, 9]. Furthermore, studies investigating the clinico-pathological features of these cancers revealed that if luminal A and basal-like breast cancers are compared, they differ in terms of risk factors, clinico-pathological presentation, histopathological features, response to therapy, and outcomes [5].

In-depth analyses of the transcriptomic profiles of luminal A, luminal B, HER2-enriched, basal-like, and normal breast-like revealed important characteristics of these molecular subtypes. Luminal tumors are characterized by the expression of the ER gene (*ESR1*) and ER-related genes. There is marked intrinsic heterogeneity within the luminal subgroup. Luminal tumors are subclassified into luminal A and luminal B subtypes based on the level of expression of proliferation-related genes, whereby luminal A tumors display low levels of expression of proliferation-related genes, whereas luminal B cancers display higher levels [10–12]. Luminal A tumors may be further subclassified into four groups, which differ in terms of their somatic mutation profiles, copy number alterations, and clinical behavior [13]. Among these subgroups, a copy number high (CNH) luminal subgroup was recognized, characterized by high genomic instability, *TP53* mutations, increased Aurora kinase signaling, and poor clinical outcome [13]. HER2-enriched cancers are characterized by expression of the *HER2* gene (*ERBB2*) and of genes found in the *HER2* amplicon. It should be noted, however, that not all HER2-enriched breast cancers

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_26

display *HER2* gene amplification and not all cases diagnosed as HER2-positive according to the ASCO/CAP guidelines are classified as HER2-enriched by microarray analysis [14]. In fact, all intrinsic breast cancer subtypes may be recognized among clinically defined HER2-positive breast cancers [15]. In light of the not uncommonly observed primary and secondary resistance to HER2 blockade, the identification of biomarkers predictive of response is of paramount importance. Along these lines, the determination of the molecular intrinsic subtype in the realm of HER2-positive disease is paving the road for the development of a therapeutically sound molecular stratification of HER2-positive breast cancer. In fact, the analysis of HER2-positive breast cancers from the NCCTG (Alliance) N9831 trial, using the Prosigna algorithm, showed that HER2-enriched and luminal tumors benefited the most from the addition of trastuzumab to chemotherapy, whilst basal-like tumors did not show a significant benefit [16].

Similarly, it was recently shown that in patients with clinically defined HER2-positive breast cancer from the PAMELA trial who were managed with dual HER2 blockade with trastuzumab and lapatinib, the pathologic complete response varied according to the intrinsic molecular subtype [17]. Indeed, HER2-positive breast tumors of the HER2-enriched subtype showed a significantly higher rate of pathologic response compared to patients from non HER2-enriched subtypes, further suggesting that the intrinsic subtype might greatly aid in the discrimination of patients who will benefit from HER2 blockage, in whom chemotherapy might potentially be spared [17].

The basal-like subtype was so named because the transcriptomic profiles of these cancers comprise genes that are usually expressed by normal breast epithelial/ basal cells. Normal-like breast cancers, on the other hand, have proven to be more controversial. There are several lines of evidence to suggest that this subtype is a mere artifact of gene expression profiling, being the result of "intrinsic" subtyping of samples with a disproportionately high content of normal breast epithelial cells and/or stromal cells [5, 10, 18, 19].

Due to limitations of hierarchical clustering analysis for the classification of single breast cancer samples in a prospective manner [20], single sample predictors have been developed [3]. They allow for gene expression-based subtyping of individual tumors based on microarray gene expression profiling. Microarray-based single sample predictors, however, seem to have limited reproducibility and to require extensive and rather complex processing of the microarray data to be applied for the classification of individual samples [11, 21]. To overcome these limitations and to allow for the use of archival material, the PAM50 assay has been developed. This is an nCounter-based assay based on the expression of 50 genes and classifies breast cancers into the four major intrinsic subtypes (i.e., luminal A, luminal B, HER2-enriched, and basal-like; the normal-like subtype was removed as it is currently perceived as a likely artifact of having a high percentage of normal cell contamination) [18]. Importantly, immunohistochemical surrogate definitions have gained widespread use in the last few years due to their similarities with breast cancer molecular subtypes as defined by gene expression profiling. Indeed, based on the recognition of "intrinsic" breast cancer subtypes, this immunohistochemical surrogate classification was accepted by the 12th St. Gallen International Breast Cancer Conference Expert Panel as a new approach for therapeutic purposes [22]. Nevertheless, it has been recognized that disagreement between the PAM50 assay and immunohistochemistry may lead to different treatment decisions [23].

In addition to the "intrinsic" subtypes, microarray-based class discovery studies have resulted in the identification of additional molecular subtypes, which are predominantly of ER-negative phenotype. The molecular apocrine subtype of breast cancer has been identified by independent investigators [24–26] and is characterized by low or no expression of ER and expression of androgen receptor (AR) and AR-related genes [24–26]. These tumors have been shown to have an aggressive clinical outcome [26] and to display some molecular and histopathological features consistent with apocrine differentiation. Through an analysis of conditional mouse models, breast cancer cell lines, and primary breast cancers, the claudin-low subtype has been identified [19, 27]. These tumors are characterized by low levels of expression of the tight junction proteins claudins 3, 4, and 7 and other adhesion molecules, including E-cadherin, and display transcriptomic features similar to those of breast cancer-initiating cells and epithelial-to-mesenchymal transition. In comparison with other intrinsic subtypes, claudin-low tumors display low levels of expression of ER and ER-related genes and intermediate levels of expression of proliferation-related genes. Although initially perceived as a variant of triple-negative breast cancers (TNBCs), up to 33% and 22% of claudin-low cancers may be ER and HER2 positive by immunohistochemical analysis [19]. From an immunohistochemical standpoint, it should be emphasized that up to 41% and 55% of tumors classified as claudin-low by gene expression profiling express claudin 3 and E-cadherin, respectively [19].

TNBC, defined by the lack of expression of ER, progesterone receptor, and HER2, is vastly heterogeneous at the molecular level, and despite the large overlap between TNBC and the basal-like intrinsic subgroup of breast cancer, it is nowadays recognized that these definitions are not synonymous. Indeed, seminal studies by Lehmann et al. [28] revealed the existence of six molecular TNBC subtypes, namely, basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M), mesenchymal stem-like (MSL),

immunomodulatory (IM), and luminal androgen receptor (LAR). Underscoring the therapeutic relevance of this molecular TNBC taxonomy, murine xenografts of breast cancer cell lines representative of the different TNBC subtypes were found to display differential sensitivity to therapeutic agents [28]. While basal-like cell lines displayed sensitivity to cisplatin, mesenchymal stem-like and LAR cell lines were shown to be sensitive to a dual PI3K and mTOR inhibitor (BEZ235) and an antiandrogen (bicalutamide), respectively [28]. The clinical implications of this taxonomy were further supported by the different responsiveness of the various TNBC molecular subtypes to neoadjuvant chemotherapy [29] and by their different survival outcomes [28]. Follow-up studies conducted by the same group revealed, nonetheless, that the transcriptional profiles of IM and MSL tumors derive from tumor-infiltrating lymphocytes and stromal cells, respectively, rather than from tumor cells, and this classification was therefore refined to include only the four remaining molecular TNBC subgroups [30]. Subsequent independent studies by Burstein et al. [31] proposed the existence of four transcriptomic TNBC subgroups. The TNBC molecular subtypes proposed by Burstein et al. [31], i.e., luminal/androgen receptor, mesenchymal, basal-like/immune-suppressed, and basal-like/immune-activated, also differed in terms of their clinical outcomes and were analogous to the ones put forward by Lehmann et al. [30], indicating that the most parsimonious number of molecular TNBC subtypes is likely four.

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) implemented an alternative molecular breast cancer taxonomy, based on the integration of genome-wide copy number alterations and transcriptomic profiles [32]. In their pioneering study, Curtis et al. [32] analyzed approximately 2,000 breast cancers and, using this integrative approach, classified them into 10 integrative clusters (IntClust 1–10). The molecular subtypes identified by this strategy had a limited correlation with the "intrinsic" subtypes and have been shown to be associated with different outcomes [32]. These investigators later devised a simplified gene expression-based methodology to subtype breast cancer into the ten IntClusts [33], which could facilitate the application of this taxonomy. A validation study in 7,544 breast cancers using this classifier confirmed the reproducibility of the IntClust molecular classification, as well as its association with survival outcome and response to neoadjuvant chemotherapy [33].

Gene Expression Prognostic Signatures

Gene expression studies have solidified the notion that breast cancer is markedly heterogeneous and that ER-positive and ER-negative breast cancer are different diseases. The identification of breast cancer patients who may benefit from adjuvant chemotherapy, and of those in whom chemotherapy could be spared, remains challenging. Nonetheless, it is nowadays recognized that the assessment of panels of genes, namely, "first-generation" signatures, could aid in the prognostication of breast cancer. It should be noted however that "first-generation" signatures, which identify the patient population having poor prognosis [34, 35], have been shown to be useful only for ER-positive breast cancer patients, as they have negligible discriminatory power in ER-negative disease, because the levels of expression of proliferation-related genes are uniformly high in these tumors (Fig. 26.1). In fact, several meta-analyses [10, 34, 36] have demonstrated that "first generation" signatures identify as poor prognosis those patients whose tumors have high levels of expression of proliferation-related genes, which have been shown to constitute one of the strongest prognostic factors in ER-positive disease [36, 37]. Microarray-based technologies allowed the initial implementation of various multigene assays, which were further developed and are nowadays commercially available [38]. Several multigene assays have been implemented in clinical practice in the context of ER-positive disease, including MammaPrint®, Breast Cancer Index, Oncotype DX®, Prosigna, and EndoPredict (Table 26.1) [5, 6, 39–43].

Even though these assays provide similar information at the population level, the pairwise concordance between different assays for individual patients is only moderate [44]. Indeed, the comparison of the EndoPredict score and Oncotype DX® RS in the same cancer samples revealed major discrepancies in 18% of cases [44]. A third of cases classified by MammaPrint® as high risk were classified as low risk by Oncotype DX® [45]. The OPTIMA prelim study [46] compared risk stratification by different multigene assays and revealed that while they provided equivalent risk information at the population level in patients with ER-positive breast cancer, they assigned individual patients to different subtypes and risk strata [46]. Indeed, there was a disagreement in risk categorization in 61% of tumors [46]. These discrepancies might be, at least in part, due to the differences in the weight of proliferation-related genes and ER signaling-related genes in the different assays, which are more relevant in early and late recurrences, respectively. Although all of these multigene assays have the power to predict early recurrences (within 5 years of diagnosis), they differ in their ability to predict late recurrences (beyond 5 years of diagnosis). Indeed, Prosigna ROR, EPclin (EndoPredict), and BCI appear to have the best predictive power for late recurrences [47].

Despite these limitations, prognostic signatures are changing clinical practice and play an important role in the management of ER-positive disease. Indeed, their incorporation in

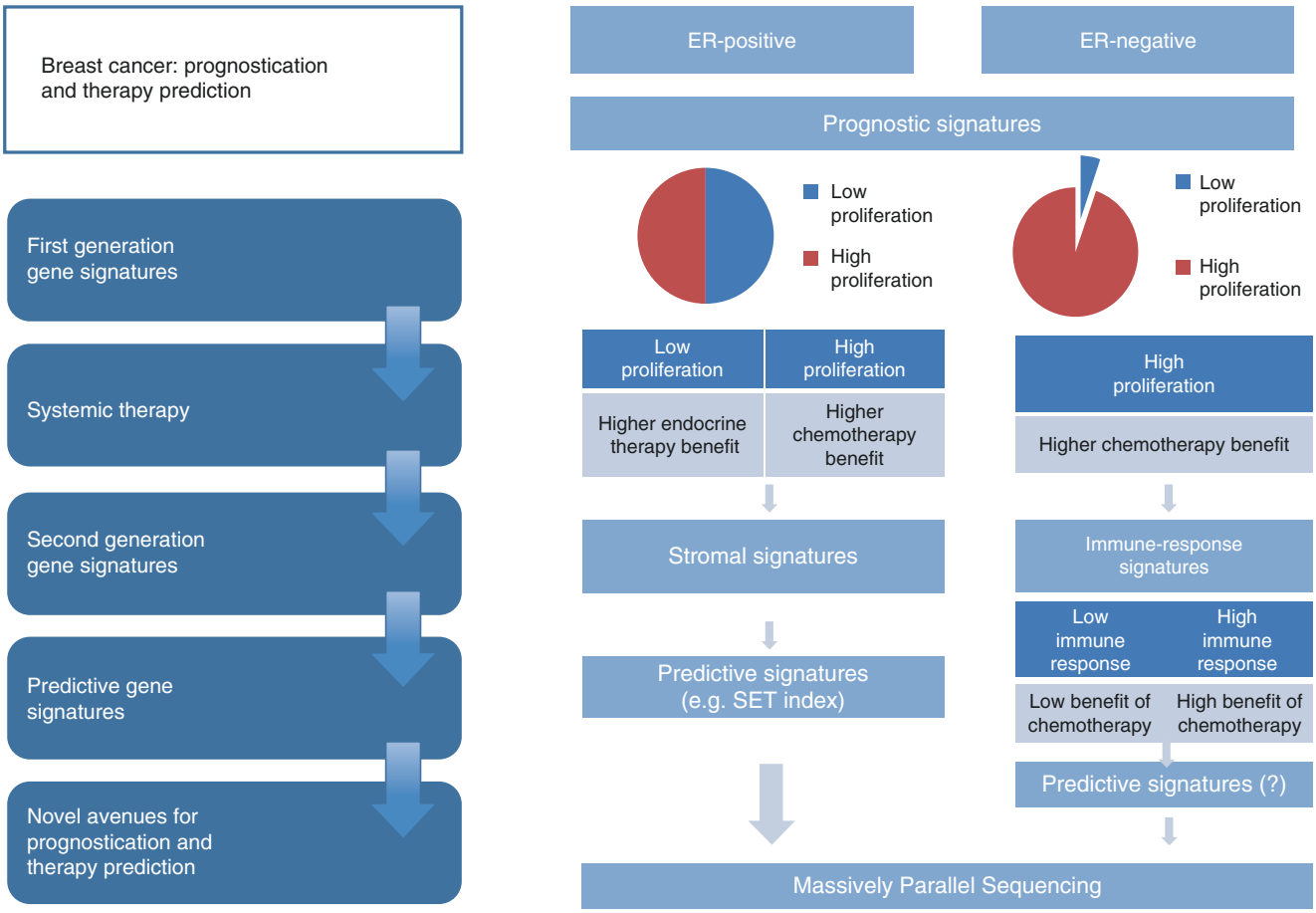


Fig. 26.1 Schematic representation of gene expression signatures and their prognostic and predictive value for estrogen receptor (ER)-positive and ER-negative breast cancer. First-generation prognostic gene expression signatures are clinically useful for ER-positive disease and classify patients into good or poor prognosis. Second-generation signatures, which are underpinned by the prognostic value conferred by the expression of immune response-related genes, may play a role in the prognostication of patients with ER-negative breast cancer. The stromal gene signatures and endocrine predictive signatures (such as the SET index) also have the potential to help personalize the therapy for patients with ER-positive disease. New genomic platforms for discovering and validating prognostic and predictive biomarkers (e.g., massively parallel sequencing) are expected to have a dramatic impact on systemic therapy decision-making for patients with breast cancer

Table 26.1 Main characteristics of commercially available gene expression signatures in breast cancer

	Gene expression signature				
Characteristic	MammaPrint®	Breast cancer index	Oncotype DX®	Prosigna ROR	EndoPredict
Material	Frozen FFPE	FFPE	FFPE	FFPE	FFPE
Central vs local lab	Central	Local	Central	Local	Local
Platform	Microarray qRT-PCR	qRT-PCR	qRT-PCR	Nanostring	qRT-PCR
Gene signature	70 genes	2 gene-ratio HOXB13:IL17R/ 5-gene molecular grade index	21 genes	55 genes (5 genes used for normalization)	8 cancer-related genes and 3 reference genes
Breast cancer population	pT1-2, N0, <61 years	ER+	ER+ and N0 disease treated with tamoxifen	All	ER+/HER2-
Level of evidence	IA	IA	IB	IB	IB
Early recurrence	Yes	Yes	Yes	Yes	Yes
Late recurrence		Good		Good	Good

Abbreviations: *ER* estrogen receptor, *FFPE* formalin-fixed paraffin-embedded, *ROR* risk of relapse, *qRT-PCR* quantitative reverse transcriptase polymerase chain reaction

the 8th edition of the American Joint Committee on Cancer (AJCC) staging system in the subset of ER-positive HER2-negative breast cancers has been recommended by a multidisciplinary team of breast cancer experts [48].

Owing to the fact that the prognostic power of these first-generation signatures largely stems from the information provided by proliferation-related genes, the classification of breast cancers according to these signatures correlates with response to conventional chemotherapy agents [49–51]. This is not surprising, given that chemotherapy preferentially targets cells that are cycling/ proliferating. An important observation, however, is that most of the low-risk/good prognosis groups identified by first-generation prognostic signatures may potentially benefit from specific chemotherapy agents (e.g., taxanes) [52, 53].

MammaPrint®

The 70-gene assay (MammaPrint®, Agendia, Netherlands) is a widely used breast cancer multigene classifier assay and the first US Food and Drug Administration (FDA)-cleared breast cancer recurrence assay. MammaPrint® is a microarray-based gene expression profiling assay that uses DNA microarray technology to predict risk of developing distant metastasis. The application of this assay is intended for patients with ER-positive node-negative, stage I-II invasive breast cancer. Although it originally required RNA extracted from fresh-frozen tumor specimens, technology improvements have eliminated the need of frozen tissue, and this assay is now available for formalin-fixed, paraffin-embedded (FFPE) tissue. Of note, the analysis of FFPE samples has been shown to be comparable to that of frozen material [54, 55].

This gene signature was originally developed by the supervised expression analysis of 25,000 genes from 78 patients with node-negative stage I-II breast cancer who did not receive adjuvant systemic therapy, which resulted in a list of 70 genes [56]. A prognostic score that categorizes patients into "good" (i.e., no distant metastasis within 5 years of follow-up) and "poor" (i.e., distant metastasis within 5 years of follow-up) outcome groups was developed. Although this prognostic signature consists of genes that are to some extent associated with proliferation, invasion, metastasis, and angiogenesis, its prognostic power seems to stem from the expression levels of proliferation-related genes alone [34].

This signature was further validated in various cohorts of breast cancer patients (e.g., node-negative, node-positive, HER2-positive) and was shown to provide prognostic information in addition to that provided by standard clinicopathological variables [56–62]. Furthermore, the prognostic groups identified by MammaPrint® seem to correlate with response to chemotherapy; MammaPrint®-defined good prognosis tumors have been reported to derive minimal benefit from chemotherapy, whereas a subset of tumors classified as of poor prognosis have higher rates of chemotherapy response [60].

The first prospective validation of the MammaPrint® assay was provided by the RASTER study [63], which included 427 node-negative breast cancer patients and showed that patients with a low-risk signature had a 5-year relapse-free survival rate of 97%, compared to 91.7% among patients with a high-risk signature. Later on, the clinical utility of the MammaPrint® assay was validated by the MINDACT randomized phase III trial [64], which included 6,693 patients with negative or 1–3 positive nodes. The results of this trial showed that patients with clinically high-risk disease based on clinico-pathological parameters (Adjuvant! Online) and a MammaPrint®-defined low genomic risk who did not receive chemotherapy had a 5-year distant metastasis-free survival of 94.7%, supporting the utility of the MammaPrint® assay in the selection of patients in whom chemotherapy could be spared [64].

Oncotype Dx®

The 21-gene assay (Oncotype DX®, Genomic Health, Redwood City, CA, USA) is one of the most widely used multigene classifier assays. It consists of a qRT-PCR-based signature in which RNA is extracted from FFPE tissue samples [65, 66]. The signature measures the expression of 21 genes, of which 16 are cancer-related genes and 5 are reference genes. An algorithm is used to calculate a "recurrence score" (RS) based on the 21-gene list ranging from 0 to 100 and classifies patients into three risk groups: low risk (RS <18), intermediate risk (RS from 18 to <31), and high risk (RS ≥31). The RS has been shown to predict the 10-year risk of distant relapse for ER-positive node-negative breast cancer patients, based on the analyses of samples from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-20 clinical trial [67]. The RS was validated in a large cohort of ER-positive, node-negative tamoxifen-treated patients from the NSABP B-14 trial which results in level I evidence to support its prognostic value [68]. In addition, RS has also been shown to be associated with benefit from chemotherapy in patients with ER-positive disease. Chemotherapy benefit is observed in patients whose tumors have a high-RS, whereas the benefit from chemotherapy is negligible in patients with low-RS cancers [69]. The first prospective study to validate the clinical utility of Oncotype DX® was the TAILORx trial [70]. To minimize undertreatment, the Oncotype DX® RS ranges were modified in this trial, with an RS of 11–25 defining the intermediate-risk group. The initial results of TAILORx showed that the risk of recurrence in patients with hormone receptor-positive,

HER2-negative, node-negative breast cancer with an RS < 10, receiving endocrine therapy alone, is very low [70], indicating that this population can safely forgo chemotherapy.

Multiple studies have evaluated the clinical utility of Oncotype DX® to determine whether patients with an intermediate RS may benefit from the addition of adjuvant chemotherapy. A recent prospective-retrospective study showed that patients with an intermediate RS (11–25) had very low 5-year distant recurrence rates, suggesting that chemotherapy did not confer clinical benefit in this group [71]. The results of the TAILORx trial in patients with an intermediate RS will be presented soon and are eagerly awaited.

Based on these studies, Oncotype Dx® has been incorporated in clinical guidelines and it use is recommended by expert panels; furthermore, it has received support from the American Society of Clinical Oncology for its use in early ER-positive node-negative breast cancer [22, 72, 73].

Oncotype DX® has been shown to provide prognostic information above and beyond that of histologic grade and tumor size [74–76]. The applications of Oncotype Dx® have been expanded, as this assay has also been revealed to be a useful prognostic test in other scenarios such as (i) ER-positive node-positive patients treated with tamoxifen, (ii) ER-positive patients treated with aromatase inhibitors, (iii) ER-positive, node-negative patients receiving no adjuvant therapy, and (iv) node-positive patients treated with doxorubicin-containing chemotherapy [75, 77, 78].

Prosigna®

The prediction analysis of microarrays 50 (PAM50) assay was originally intended as a means to identify breast cancer "intrinsic" gene subtypes with high prognostic validity [18]. Prosigna®, a commercially available assay using NanoString technology in RNA extracted from FFPE samples, was later developed, and its use in postmenopausal women with hormone receptor-positive tumors with or without node involvement was approved by the FDA [79]. Assessment of the expression of 50 classifier genes and 5 control genes can be used to classify breast tumors in the intrinsic subtypes. In addition, this assay provides a risk of recurrence score (ROR), which ranges between 0 and 100, defining low-, intermediate-, and high-risk categories. The ROR score in the training dataset predicted the probability of cancer recurrence over 10 years for patients with node-negative tumors who did not receive adjuvant systemic therapy [18]. The prognostic value of ROR score has been further validated for 786 patients with ER-positive breast cancer treated with tamoxifen, showing that PAM50 and tumor size might give more prognostic information than other clinico-pathological variables [80]. Notably, an 11-gene proliferation signature, which is related to cell cycle function, was derived from the 50 genes of the PAM50 assay. The 11-gene signature was found to improve the original model as it was found to have more prognostic value than expression of Ki67 [80]. A study comparing the prognostic information provided by Oncotype Dx® and PAM50 using over 1,000 samples from the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial revealed that the PAM50 ROR score yielded significantly more prognostic information than the Oncotype Dx® RS, and that the PAM50 ROR provides independent prognostic information above and beyond that offered by nodal status, tumor size, histopathologic grade, age, and type of endocrine treatment [81]. Another validation study included 1,478 postmenopausal patients with hormone receptor-positive, HER2-negative breast cancer receiving adjuvant endocrine therapy, and showed that the ROR score was able to predict relapse-free survival [82]. Similarly, a recent comprehensive study conducted in a nationwide Danish cohort, including postmenopausal women with hormone receptor-positive, HER2-negative breast cancer, solidified the notion that Prosigna ROR may identify patients with negative- or one to three positive-nodes in whom adjuvant chemotherapy could be spared [83].

Breast Cancer Index^{SM}(BCI)

The Breast Cancer Index (BCI) molecular assay (BioTheranostics, San Diego, CA) was developed to assess the risk of distant recurrence in ER-positive, node-negative breast cancer patients [74, 84, 85]. It is a prognostic assay which combines two gene expression signatures: the HOXB13:IL17BR (H:I) two-gene ratio, which predicts distant recurrence in patients with ER-positive breast cancer treated with tamoxifen [84], and a proliferation-related five-gene molecular grade index (MGI) [74] that distinguishes grade 1 from grade 3 cancers. This dichotomous index (MGI together with *HOXB13:IL17BR*) is based on quantitative RT-PCR (qRT-PCR) using RNA from FFPE tissues, and provides more accurate prognosis than either biomarker alone. Furthermore, the BCI, is a continuous risk model that enables prediction of distant recurrence risk, and is significantly associated with distant recurrence and breast cancer death [85].

The BCI assay, 21-gene recurrence score, and an immunohistochemical prognostic model (IHC4) were prospectively compared for both early (0–5 years) and late (5–10 years) recurrence in ER-positive, node-negative patients in the TransATAC study (i.e., patients enrolled in the Arimidex, Tamoxifen, Alone or in Combination (ATAC) clinical trial) [86]. The BCI has been shown to be a signifi-

cant prognostic test for risk of both early and late distant recurrence and could assist in the identification of high-risk patients who would derive benefit from extended endocrine therapy or additional therapy.

A recent retrospective analysis comparing the prognostic accuracy of BCI and Oncotype DX® RS showed that BCI possessed a higher prognostic accuracy than the RS [87]. Notably, the BCI was able to identify subsets of patients with low- and intermediate-RS tumors with significant rates of distant recurrence [87], indicating that BCI may aid in the selection of patients with hormone receptor-positive and node-negative breast cancer who could benefit from adjuvant chemotherapy or extended endocrine therapy. A novel Breast Cancer Index model (BCIN+) was later developed for the assessment of the risk for distant recurrence in patients with hormone receptor-positive breast cancer and one to three positive lymph nodes [88]. BCIN+ integrates BCI gene expression and tumor size and grade and could identify a patient population with limited risk of recurrence over 15 years, who could safely forgo extended endocrine therapy [88].

EndoPredict Test

EndoPredict is an RNA-based multigene assay that interrogates proliferation and ER signaling-related genes for the assessment of the probability of distant recurrence in patients with ER-positive, HER2-negative breast cancer treated with adjuvant endocrine therapy [39, 41–43]. The EndoPredict test is based on the quantification of mRNA levels of eight cancer genes plus three reference genes in FFPE specimens by qRT-PCR and was shown to provide additional prognostic information, which is independent from clinico-pathological parameters (i.e., Adjuvant!Online and Ki67 labeling index) [40]. In two validation cohorts, the EndoPredict test was combined with clinical risk factors (i.e., nodal status and tumor size) into a comprehensive risk score called EPclin, which has been shown to identify a subgroup of "very-low"-risk patients who may be satisfactorily treated with adjuvant endocrine therapy only [39]. The clinical utility of EndoPredict was also validated in the patients with ER-positive, HER2-negative node-positive breast cancer from the GEICAM 9906 trial treated with adjuvant chemotherapy and endocrine therapy [89]. The EndoPredict and EPclin scores showed independent prognostic power for the prediction of metastasis-free survival and low-risk and high-risk patients [89]. A recent study comparing the performance of EPclin and Oncotype Dx® RS for the prediction of 10-year distant recurrence showed that EPclin provided more prognostic information then Oncotype Dx® RS [90].

Gene Expression Predictive Signatures

Predictive gene signatures aim to define the therapeutic response to chemotherapy, endocrine therapy, or other target agents [5, 6, 91–95]. Akin to the prognostic gene expression signatures, ER status and proliferative index have been shown to be major determinants of response to combinatorial chemotherapy. Thus far, the clinical value of gene expression signatures predictive of response to single chemotherapy agents remains controversial for breast cancer. In fact, there is no robust available gene signature capable of predicting responses to specific therapeutic agents. Several hypotheses have been advanced to explain the limited success in developing and validating predictive signatures. First, resistance to chemotherapy can be caused by functional alterations in few or single genes, and it is plausible that microarray-based gene expression profiling would not be sufficiently sensitive to identify such genes [91]. Second, intra-tumor genetic heterogeneity plays an important role in determining the emergence of drug resistance. Breast tumors often comprise heterogeneous collections of cancer cells that encompass rare clonal subpopulations, which have different genetic and epigenetic aberrations [96, 97]. Some genetic aberrations, which may be found in single clones of tumors, may drive therapeutic resistance [98]. In fact, as microarrays give an average of the expression profile of the tumor, this technique would not be reliable to identify those rare resistant clones. Finally, multiple genetic and epigenetic factors and also drug-resistance mechanisms not related to the tumor itself (e.g., tissue microenvironment, patient metabolism) may determine resistance to therapy [6]. Although some predictive gene expression signatures appear to have predictive value in validation studies (e.g., SET index) [99], their accuracy to determine the response of individual patients may be limited [6].

Massively Parallel Sequencing and the Impact in Intra-tumor Genetic Heterogeneity

The advent of massively parallel sequencing has enabled the analysis of the entire constellation of genetic alterations in cancers to be defined in a matter of days at reasonable costs. Several large-scale massively parallel sequencing-based studies of breast cancer have now been completed and demonstrated that (i) the collection of genetic aberrations found in breast cancers is complex with a limited number of highly recurrently mutated genes in a substantial proportion of unselected cases [32, 96, 100, 101], (ii) the number of genes mutated in small minorities of breast cancers is vast, (iii) the repertoire of mutations in luminal and basal-like

breast cancers is vastly different, and (iv) despite these differences, there is no gene or mutation that defines a subtype of breast cancer [100–103].

Genomic analyses of human cancers have provided direct evidence of spatial [104–106] and temporal [104, 107, 108] intra-tumor genetic heterogeneity and have shown that a substantial proportion of cancers at diagnosis are composed of mosaics of tumor cells [96, 106], where subclones of cells harbor private mutations in addition to the founder genetic events. Although intra-tumor genetic heterogeneity is recognized for many years [109], it has been explored in primary breast cancers using massively parallel sequencing approaches in a limited number of studies (Fig. 26.2) [96, 97, 110]. The impact of intra-tumor genetic heterogeneity on the biology and, consequently, on treatment design of breast cancer remains to be fully understood. Genomic analysis of two pairs of matched primary tumors and distant metastatic relapses after adjuvant treatment revealed differences in their mutational makeup [107, 108], and suggested that clonal selection during the metastatic process is likely to occur. Along these lines, the integrative whole exome sequencing and gene expression analysis of a cohort of 500 metastatic solid tumors, which was enriched for breast cancer patients, identified *TP53*, *CDKN2A*, *PTEN*, *PIK3CA*, and *RB1* as the most frequently somatically mutated genes in metastatic cancer [7]. A recent study portrayed the mutational landscape of 216 metastatic breast cancers and compared it to the one of 772 primary breast cancers from the TCGA [111]. This study identified *ESR1* and *RB1* as driver genes enriched in breast cancer metastases, with *ESR1* being the most frequently metastasis-specific mutated gene [111]. Among other frequently mutated actionable genes identified in ER-positive HER2-negative metastatic breast cancer were *TSC1* and *TSC2*, *ERBB4*, *NOTCH3*, and *ALK* [111].

Mutations targeting *ESR1* are also among the actionable targets that differ between primary and metastatic breast cancers. While *ESR1* mutations are found in <1% of primary breast cancers, they may be identified in up to 54% of

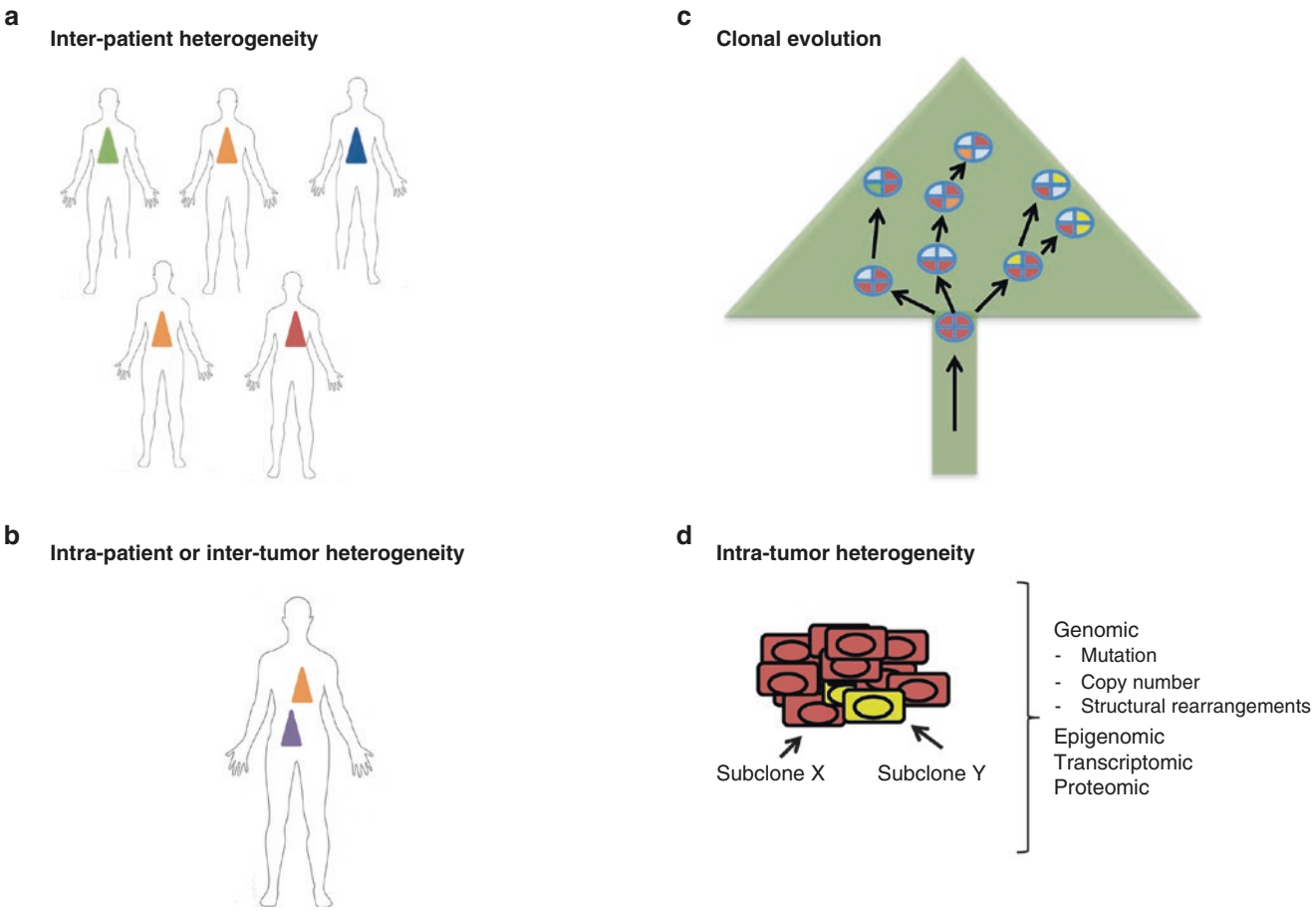


Fig. 26.2 Tumor heterogeneity. (**a**) Inter-patient heterogeneity. (**b**) Inter-patient heterogeneity. (**c**) Clonal evolution and the tree model: mutations shared by all tumor cells proceed from the founder clone which is depicted as the trunk of the tree. The branches are composed by tumor cells that acquire mutations present only in a subset of the tumor cells. (**d**) Intra-tumor genetic heterogeneity and the approaches for the characterization of the molecular aberrations in breast cancers

relapses following endocrine therapy [112]. *ESR1* mutations affect the DNA binding domain, and some of these mutations have been shown to result in the activation of ER-dependent genes even in the absence of E2 and to require higher doses of tamoxifen and fulvestrant for the inhibition of ER activity [113–115]. Along these lines, *ESR1* mutations may be identified in the cfDNA of patients with metastatic breast cancer who progress despite endocrine therapies [116]. Moreover, the detection of Y537S and D538G *ESR1* mutations in cfDNA of patients with ER-positive metastatic breast cancer from the BOLERO-2 trial receiving aromatase inhibitors was associated with a shorter survival [117].

HER2 mutations are also enriched in metastatic breast cancer [118]. Of note, not all *HER2* mutations result in activation of downstream pathways [119]. Indeed, *in vitro* and *in vivo* assays revealed that only a subset of *HER2* mutations are *bona fide* activating mutations [119]. Importantly, upon therapeutic pressure, passenger *HER2* mutations may become drivers. Along these lines, massively parallel sequencing of lapatinib-resistant cell models showed that acquisition of the *HER2* L755S mutation may result HER2 reactivation, representing a mechanism of resistance to lapatinib, which may be overcome by irreversible HER2 inhibition [120]. A recent "basket" trial across 21 cancer types using the pan-HER kinase inhibitor neratinib showed that its efficacy in *HER2* mutant cases varied according to tumor type and individual mutant variant [121]. Breast tumors and missense mutations targeting the kinase domain of *HER2* were found to be associated with the greatest sensitivity to neratinib [121].

The spatial and temporal intra-tumor genetic heterogeneity observed in solid cancers constitutes a challenge for the realization of the potentials of precision medicine, given that the results of genetic biomarker analyses performed in single biopsies for treatment decision-making may differ according to the area of the tumor sampled [104], between the primary tumor and its distant metastases, or even between different metastatic sites [104, 122]. This multiregional separation of molecular aberrations can lead to sampling bias, potentially impairing the interpretation of genomics results derived from individual biopsies. Therefore, approaches to provide a global assessment of the repertoire of somatic genetic aberrations in a tumor are important for the accurate selection of targeted therapies for individual patients.

Deciphering intra-tumor heterogeneity using massively parallel sequencing approaches has important implications that may refine our understanding of breast cancer biology, its genetic diversity and the mechanisms that lead to therapeutic resistance [103, 122–125]. Much effort has been put in this direction, including massively parallel sequencing of single cells [106] and circulating biomarkers [126–129].

Liquid Biopsies in Breast Cancer

Tumors are composed of multiple subclones with different genetic alterations, and minor subpopulations of the primary tumor may be the ones that develop into metastasis [108]. Despite their many advantages, traditional DNA sequencing approaches, in which the bulk of the tumor is analyzed, lack the power to detect minor tumor subclones [130] which may be the source of disease progression and resistance to therapy [123]. Moreover, occasionally, the anatomic inaccessibility of metastatic outgrowths precludes their sampling [131]. Liquid biopsies, which encompass the study of circulating cell-free tumor DNA (cfDNA) and circulating tumor cells (CTCs), have the potential to circumvent the limitations inherent to tissue-based DNA sequencing and to monitor dynamic changes in tumor genomes, in a noninvasive manner [129, 132].

Multiple lines of evidence indicate that the study of liquid biopsies has a potential use in tailoring therapy in early and metastatic breast cancer [133]. In the context of early disease, mutation tracking in ctDNA in plasma in early breast cancer patients receiving neoadjuvant chemotherapy could predict metastatic relapse in a shorter median lead time than the methods currently used [134]. Moreover, it allowed for the identification of the genetic events in minimal residual disease that could in turn predict the genetic alterations in subsequent metastasis with more accuracy than sequencing of the primary tumor [134].

Liquid biopsies may also play a role in the detection of genetic alterations that drive therapeutic resistance in metastatic breast cancer, such as *ESR1* mutations [135]. The detection of *ESR1* mutations in liquid biopsies might aid in the triage of patients with metastatic hormone receptor-positive breast cancer for further endocrine therapies, as illustrated in the study of archived baseline plasma of patients of the SoFEA trial [136]. In this study, patients with *ESR1* mutations detected in plasma treated with fulvestrant had a better progression-free survival than those treated with exemestane, whereas no difference was observed in patients with wild-type *ESR1*.

Other potential uses of liquid biopsies in tailoring the management of breast cancer patients are currently being explored [133]. *BRCA1/2* reversion mutations in *BRCA1/2* germline mutation carriers may functionally restore *BRCA1* and *BRCA2* and mediate resistance to platinum salts or PARP inhibition [137]. MPS analysis of cfDNA detected *BRCA1/2* reversion mutations in *BRCA1/2* germline mutation carriers with metastatic breast cancer pretreated with platinum and/or PARP inhibitors, underscoring the potential of liquid biopsies to aid in the selection of patients amenable to PARP inhibition [138].

Taken together, a burgeoning body of evidence indicates that analysis of liquid biopsies represents a robust approach to tackle breast intratumor heterogeneity and to guide the management of breast cancer patients, both in early and in advanced stages.

Molecular Advances in Histologic Subtyping of Breast Tumors

Comprehensive genomic portrayals of breast cancer have analyzed cohorts of unselected breast cancers, where invasive ductal carcinomas of no special type (IDC-NST) were overrepresented [102]. Special types of breast cancer, which collectively account for up to 20% of all invasive breast cancers, were largely not investigated in those studies. In fact, the second breast TCGA study, which focused on lobular breast cancer [139], and independent investigators of invasive lobular carcinomas (ILC) [140] confirmed that inactivating *CDH1* mutations, the hallmark of lobular carcinomas, are not present in IDC-NSTs. Furthermore, the genetic alterations activating the estrogen pathway differ according to tumor histology, with *FOXA1* and *GATA3* mutations being more frequent in ILCs and IDC–NSTs, respectively [139, 140].

The analysis of special types of breast cancer, however, has provided important insights in regard to the taxonomy of breast cancer. Studies focusing on the genomic characterization of rare breast cancer types have demonstrated that the vast histologic heterogeneity of breast cancer is paralleled by marked heterogeneity at the molecular level, which is more overt in the realm of TNBC [141, 142]. Indeed, studies conducted by our group and others have shown that contrary to the common perception of TNBC as a group of tumors with uniformly aggressive biology and poor prognosis, low-grade variants of triple negative disease exist [141, 142]. Among these entities, the "low-grade triple-negative breast neoplasia" family, which includes microglandular adenosis (MGA), atypical MGA, and acinic cell carcinoma (ACC), can be recognized. Notwithstanding their low-grade morphology, MGA and ACC display complex genomic profiles and frequent *TP53* mutations, similar to conventional high-grade TNBCs [143].

Salivary gland-like tumors of the breast are also low-grade TNBC variants and encompass tumors that despite being more frequent in the salivary glands arise also in the breast and are underpinned by pathognomonic genetic alterations, such as secretory carcinomas and adenoid cystic carcinomas [144, 145]. Secretory carcinomas are characterized by the t(12;15)(p13;q25) translocation that results in the *ETV6–NTRK3* fusion gene [144]. The hallmark genetic alteration of adenoid cystic carcinomas is the t(6;9)(q22-23;p23-24) translocation which results in the *MYB-NFIB* fusion gene [145]. Interestingly, our study of *MYB-NFIB*-negative adenoid cystic carcinomas revealed that these tumors harbor *MYBL1* rearrangements (*MYBL1-ACTN1* and *MYBL1-NFIB*) or *MYB* amplification, showing that this entity is driven by MYB or MYBL1 activation achieved by different mechanisms, and constitutes an example of convergent phenotype [146].

Adenomyoepitheliomas (AMEs) and solid papillary carcinomas with reverse polarity (SPCRPs) constitute additional examples of genotypic-phenotypic correlations in the breast. Our recent analysis of breast AMEs revealed that their genetic makeup varies according to their ER status [147]. ER-positive AMEs display frequent *PIK3CA* or *AKT1* activating mutations, whereas ER-negative AMEs are characterized by *HRAS* Q61 hotspot mutations co-occurring with *PIK3CA* or *PIK3R1* mutations [147]. Notably, epithelial-myoepithelial carcinomas of the salivary glands harbor frequent *HRAS* Q61 hotspot mutations which co-occur with *PIKC3A* mutations in almost half of cases [148], showing that the aforementioned mutational co-occurrence results in epithelial-myoepithelial differentiation regardless of anatomic location. Importantly, this study [147] qualifies *HRAS* mutations as pathognomonic for AMEs in a breast-specific context.

"SPCRPs are extremely rare breast tumors, which morphologically resemble the tall cell variant of papillary thyroid carcinoma and constitute an additional example of genotypic-phenotypic association in the breast. Our analysis of two independent cohorts of SPCRPs revealed that these tumors are underpinnned by *IDH2* R172 hotspot or *TET2* mutations, concurrent with *PIK3CA* or *PIK3R1* mutations [149, 150]. Simultaneous *IDH2* and *PIK3CA* mutations in breast cell lines resulted in the recapitulation of the characteristic morphology of SPCRPs [149] illustrating how the integration of molecular studies and classic pathology resulted in the definition of a discrete breast cancer subtype with a distinctive morphology and molecular underpinning.

Conclusions

Molecular diagnostics play a key role in the management of breast cancer patients, and molecular assays are being increasingly incorporated in routine clinical practice. Gene expression profiling has provided significant advances in the molecular classification and prognostication of breast cancer and has given new insights regarding therapeutic prediction. Microarray-based gene expression studies have changed the way breast cancer is perceived and have highlighted the fact that breast cancer comprises a heterogeneous collection of diseases with distinct molecular characteristics and outcomes. Along these lines, the development of multigene signatures has allowed the identification of patients with ER-positive, HER2-negative breast cancer in whom chemotherapy could be spared.

The identification of actionable targets by massively parallel sequencing approaches is becoming a cornerstone for the realization of the potentials of precision medicine. Indeed, the implementation of liquid biopsies in the monitoring of early and advanced breast cancer, in the near future, as means to overcome the challenges posed by intra-tumor heterogeneity is not hard to envision.

The integration of molecular studies and classic pathology in the recent years has facilitated the dissection of the morphologic and genetic heterogeneity of breast cancer. Thus, the taxonomy of the breast is becoming increasingly more reliant on the genetic makeup of tumors rather than solely on classical histomorphological parameters. Molecular techniques are developing at an unprecedented pace. Nevertheless, to achieve the goals of individualized therapy, molecular methods must be incorporated into clinical practice after undergoing the same level of scrutiny that current diagnostic techniques have been subjected to.

References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–52.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869–74.
3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100(14):8418–23.
4. Jeffrey SS, Lonning PE, Hillner BE. Genomics-based prognosis and therapeutic prediction in breast cancer. J Natl Compr Cancer Netw. 2005;3(3):291–300.
5. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet. 2011;378(9805):1812–23.
6. Weigelt B, Pusztai L, Ashworth A, Reis-Filho JS. Challenges translating breast cancer gene signatures into the clinic. Nat Rev Clin Oncol. 2012;9(1):58–64.
7. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548(7667):297–303.
8. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013;153(1):17–37.
9. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. 2001;61(16):5979–84.
10. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7:96.
11. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J Pathol. 2010;220(2):263–80.
12. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009;360(8):790–800.
13. Ciriello G, Sinha R, Hoadley KA, Jacobsen AS, Reva B, Perou CM, et al. The molecular diversity of Luminal A breast tumors. Breast Cancer Res Treat. 2013;141(3):409–20.
14. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: american society of clinical oncology/college of american pathologists clinical practice guideline update. J Clin Oncol. 2013;31(31):3997–4013.
15. Kim J, Pareja F, Weigelt B, Reis-Filho JS. Prediction of trastuzumab benefit in HER2-positive breast cancers: is it in the intrinsic subtype? J Natl Cancer Inst. 2017;109(2):djw218.
16. Perez EA, Ballman KV, Mashadi-Hossein A, Tenner KS, Kachergus JM, Norton N, et al. Intrinsic subtype and therapeutic response among HER2-positive breaty st tumors from the NCCTG (alliance) N9831 trial. J Natl Cancer Inst. 2017;109(2):djw207.
17. Llombart-Cussac A, Cortes J, Pare L, Galvan P, Bermejo B, Martinez N, et al. HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. Lancet Oncol. 2017;18(4):545–54.
18. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160–7.
19. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.
20. Mackay A, Weigelt B, Grigoriadis A, Kreike B, Natrajan R, A'Hern R, et al. Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. J Natl Cancer Inst. 2011;103(8):662–73.
21. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, Quackenbush J, et al. A three-gene model to robustly identify breast cancer molecular subtypes. J Natl Cancer Inst. 2012;104(4):311–25.
22. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes – dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 2011;22(8):1736–47.
23. Bastien RR, Rodriguez-Lescure A, Ebbert MT, Prat A, Munarriz B, Rowe L, et al. PAM50 breast Cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med Genet. 2012;5:44.
24. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene. 2005;24(29):4660–71.
25. Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene. 2006;25(28):3994–4008.
26. Guedj M, Marisa L, de Reynies A, Orsetti B, Schiappa R, Bibeau F, et al. A refined molecular taxonomy of breast cancer. Oncogene. 2012;31(9):1196–206.
27. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.
28. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750–67.
29. Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, et al. Differential response to neoadju-

vant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin Cancer Res. 2013;19(19):5533–40.
30. Lehmann BD, Jovanovic B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of triple-negative breast Cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. PLoS One. 2016;11(6):e0157368.
31. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res. 2015;21(7):1688–98.
32. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346–52.
33. Ali HR, Rueda OM, Chin SF, Curtis C, Dunning MJ, Aparicio SA, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biol. 2014;15(8):431.
34. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.
35. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med. 2006;355(6):560–9.
36. Reyal F, van Vliet MH, Armstrong NJ, Horlings HM, de Visser KE, Kok M, et al. A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. Breast Cancer Res. 2008;10(6):R93.
37. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res. 2008;14(16):5158–65.
38. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. Nat Rev Clin Oncol. 2017;14(10):595–610.
39. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17(18):6012–20.
40. Dubsky P, Filipits M, Jakesz R, Rudas M, Singer CF, Greil R, et al. EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. Ann Oncol. 2013;24(3):640–7.
41. Kronenwett R, Bohmann K, Prinzler J, Sinn BV, Haufe F, Roth C, et al. Decentral gene expression analysis: analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test. BMC Cancer. 2012;12:456.
42. Azim HA Jr, Michiels S, Zagouri F, Delaloge S, Filipits M, Namer M, et al. Utility of prognostic genomic tests in breast cancer practice: the IMPAKT 2012 Working Group Consensus Statement. Ann Oncol. 2013;24(3):647–54.
43. Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. Br J Cancer. 2013;109:2959.
44. Varga Z, Sinn P, Fritzsche F, von Hochstetter A, Noske A, Schraml P, et al. Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. PLoS One. 2013;8(3):e58483.
45. Iwamoto T, Lee JS, Bianchini G, Hubbard RE, Young E, Matsuoka J, et al. First generation prognostic gene signatures for breast cancer predict both survival and chemotherapy sensitivity and identify overlapping patient populations. Breast Cancer Res Treat. 2011;130(1):155–64.
46. Bartlett JM, Bayani J, Marshall A, Dunn JA, Campbell A, Cunningham C, et al. Comparing breast Cancer multiparameter tests in the OPTIMA prelim trial: no test is more equal than the others. J Natl Cancer Inst. 2016;108(9):djw050.
47. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Denkert C, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor-positive breast cancer: a secondary analysis of a randomized clinical trial. JAMA Oncol. 2018;4(4):545–53.
48. Amin MB, et al. AJCC cancer staging manual. 8th ed. New York: Springer; 2016.
49. Iwamoto T, Bianchini G, Booser D, Qi Y, Coutant C, Ya-Hui Shiang C, et al. Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J Natl Cancer Inst. 2010;
50. Iwamoto T, Pusztai L. Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? Genome Med. 2010;2(11):81.
51. Reis-Filho JS, Weigelt B, Fumagalli D, Sotiriou C. Molecular profiling: moving away from tumor philately. Sci Transl Med. 2010;2(47):47ps3.
52. A'Hern RP, Jamal-Hanjani M, Szasz AM, Johnston SR, Reis-Filho JS, Roylance R, et al. Taxane benefit in breast cancer – a role for grade and chromosomal stability. Nat Rev Clin Oncol. 2013;10(6):357–64.
53. Martin M, Prat A, Rodriguez-Lescure A, Caballero R, Ebbert MT, Munarriz B, et al. PAM50 proliferation score as a predictor of weekly paclitaxel benefit in breast cancer. Breast Cancer Res Treat. 2013;138(2):457–66.
54. Sapino A, Roepman P, Linn SC, Snel MH, Delahaye LJ, van den Akker J, et al. MammaPrint molecular diagnostics on formalin-fixed, paraffin-embedded tissue. J Mol Diagn. 2014;16(2):190–7.
55. Mittempergher L, de Ronde JJ, Nieuwland M, Kerkhoven RM, Simon I, Rutgers EJ, et al. Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue. PLoS One. 2011;6(2):e17163.
56. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530–6.
57. Mook S, Schmidt MK, Viale G, Pruneri G, Eekhout I, Floore A, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. Breast Cancer Res Treat. 2009;116(2):295–302.
58. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst. 2006;98(17):1183–92.
59. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol. 2008;26(5):729–35.
60. Knauer M, Cardoso F, Wesseling J, Bedard PL, Linn SC, Rutgers EJ, et al. Identification of a low-risk subgroup of HER-2-positive breast cancer by the 70-gene prognosis signature. Br J Cancer. 2010;103(12):1788–93.
61. Mook S, Knauer M, Bueno-de-Mesquita JM, Retel VP, Wesseling J, Linn SC, et al. Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. Ann Surg Oncol. 2010;17(5):1406–13.
62. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ, et al. Long-term impact of the 70-gene signature on breast cancer outcome. Breast Cancer Res Treat. 2014;143(3):587–92.
63. Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J, et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int J Cancer. 2013;133(4):929–36.

64. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med. 2016;375(8):717–29.
65. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol. 2008;26(5):721–8.
66. Kim C, Paik S. Gene-expression-based prognostic assays for breast cancer. Nat Rev Clin Oncol. 2010;7(6):340–7.
67. Fisher B, Dignam J, Wolmark N, DeCillis A, Emir B, Wickerham DL, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. J Natl Cancer Inst. 1997;89(22):1673–82.
68. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817–26.
69. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol. 2006;24(23):3726–34.
70. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med. 2015;373(21):2005–14.
71. Stemmer SM, Steiner M, Rizel S, Soussan-Gutman L, Ben-Baruch N, Bareket-Samish A, et al. Clinical outcomes in patients with node-negative breast cancer treated based on the recurrence score results: evidence from a large prospectively designed registry. NPJ Breast Cancer. 2017;3:33.
72. NCCN Clinical Practicce Guidelines in Oncology™. Breast Cancer [online]. http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. 2012.
73. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007;25(33):5287–312.
74. Goldstein LJ, Gray R, Badve S, Childs BH, Yoshizawa C, Rowley S, et al. Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. J Clin Oncol. 2008;26(25):4063–71.
75. Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res. 2006;8(3):R25.
76. Tang CG, Lin AY. Molecular prediction of recurrence of breast cancer. N Engl J Med. 2005;352(15):1605–7. author reply −7.
77. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol. 2010;11(1):55–65.
78. Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. J Clin Oncol. 2010;28(11):1829–34.
79. Brumbaugh CD, Kim HJ, Giovacchini M, Pourmand N. NanoStriDE: normalization and differential expression analysis of NanoString nCounter data. BMC Bioinformatics. 2011;12:479.
80. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res. 2010;16(21):5222–32.
81. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol. 2013;31(22):2783–90.
82. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. Ann Oncol. 2014;25(2):339–45.
83. Laenkholm AV, Jensen MB, Eriksen JO, Rasmussen BB, Knoop AS, Buckingham W, et al. PAM50 risk of recurrence score predicts 10-year distant recurrence in a comprehensive Danish cohort of postmenopausal women allocated to 5 years of endocrine therapy for hormone receptor-positive early breast cancer. J Clin Oncol. 2018;36(8):735–40.
84. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell. 2004;5(6):607–16.
85. Jerevall PL, Ma XJ, Li H, Salunga R, Kesty NC, Erlander MG, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. Br J Cancer. 2011;104(11):1762–9.
86. Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. Lancet Oncol. 2013;14(11):1067–76.
87. Sestak I, Zhang Y, Schroeder BE, Schnabel CA, Dowsett M, Cuzick J, et al. Cross-stratification and differential risk by breast cancer index and recurrence score in women with hormone receptor-positive lymph node-negative early-stage breast cancer. Clin Cancer Res. 2016;22(20):5043–8.
88. Zhang Y, Schroeder BE, Jerevall PL, Ly A, Nolan H, Schnabel CA, et al. A novel breast cancer index for prediction of distant recurrence in HR(+) early-stage breast cancer with one to three positive nodes. Clin Cancer Res. 2017;23(23):7217–24.
89. Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K, et al. Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. Breast Cancer Res. 2014;16(2):R38.
90. Buus R, Sestak I, Kronenwett R, Denkert C, Dubsky P, Krappmann K, et al. Comparison of EndoPredict and EPclin with Oncotype DX Recurrence Score for prediction of risk of distant recurrence after endocrine therapy. J Natl Cancer Inst. 2016;108(11):djw149.
91. Borst P, Wessels L. Do predictive signatures really predict response to cancer chemotherapy? Cell Cycle. 2010;9(24):4836–40.
92. Jansen MP, Foekens JA, van Staveren IL, Dirkzwager-Kiel MM, Ritstier K, Look MP, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. J Clin Oncol. 2005;23(4):732–40.
93. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol. 2004;22(12):2284–93.
94. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet. 2003;362(9381):362–9.
95. Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J, et al. Multifactorial approach to predicting resistance to anthracyclines. J Clin Oncol. 2011;29(12):1578–86.

96. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395–9.
97. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994–1007.
98. Collisson EA, Cho RJ, Gray JW. What are we learning from the cancer genome? Nat Rev Clin Oncol. 2012;9(11):621–30.
99. Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. J Clin Oncol. 2010;28(27):4111–9.
100. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012;486(7403):405–9.
101. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 2012;486(7403):353–60.
102. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70.
103. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486(7403):400–4.
104. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
105. Geyer FC, Weigelt B, Natrajan R, Lambros MB, de Biase D, Vatcheva R, et al. Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. J Pathol. 2010;220(5):562–73.
106. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90–4.
107. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009;461(7265):809–13.
108. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010;464(7291):999–1005.
109. Fidler IJ. Tumor heterogeneity and the biology of cancer invasion and metastasis. Cancer Res. 1978;38(9):2651–60.
110. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149(5):979–93.
111. Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, et al. Mutational profile of metastatic breast cancers: a retrospective analysis. PLoS Med. 2016;13(12):e1002201.
112. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, et al. Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. Clin Cancer Res. 2014;20(7):1757–67.
113. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. Cell Rep. 2013;4(6):1116–30.
114. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet 2013;45(12):1446-1451.
115. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat Genet. 2013;45:1439.
116. Chu D, Paoletti C, Gersch C, VanDenBerg D, Zabransky D, Cochran R, et al. ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients. Clin Cancer Res. 2015;22(4):993–9.
117. Chandarlapaty S, Chen D, He W, Sung P, Samoila A, You D, et al. Prevalence of ESR1 mutations in cell-free DNA and outcomes in metastatic breast cancer: a secondary analysis of the BOLERO-2 clinical trial. JAMA Oncol. 2016;2(10):1310–5.
118. Weigelt B, Reis-Filho JS. Activating mutations in HER2: neu opportunities and neu challenges. Cancer Discov. 2013;3(2):145–7.
119. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, et al. Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer Discov. 2013;3(2):224–37.
120. Xu X, De Angelis C, Burke KA, Nardone A, Hu H, Qin L, et al. HER2 reactivation through acquisition of the HER2 L755S mutation as a mechanism of acquired resistance to HER2-targeted therapy in HER2(+) breast cancer. Clin Cancer Res. 2017;23(17):5123–34.
121. Hyman DM, Piha-Paul SA, Won H, Rodon J, Saura C, Shapiro GI, et al. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. Nature. 2018;554(7691):189–94.
122. Swanton C. Intratumor heterogeneity: evolution through space and time. Cancer Res. 2012;72(19):4875–82.
123. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323–34.
124. Turner NC, Reis-Filho JS. Genetic heterogeneity and cancer drug resistance. Lancet Oncol. 2012;13(4):e178–85.
125. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. Sci Transl Med. 2012;4(127):127ps10.
126. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One. 2012;7(5):e33788.
127. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
128. De Mattos-Arruda L, Cortes J, Santarpia L, Vivancos A, Tabernero J, Reis-Filho JS, et al. Circulating tumour cells and cell-free DNA as tools for managing breast cancer. Nat Rev Clin Oncol. 2013;10(7):377–89.
129. Dawson SJ, Rosenfeld N, Caldas C. Circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013;369(1):93–4.
130. Martelotto LG, Ng CK, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast cancer intra-tumor heterogeneity. Breast Cancer Res. 2014;16(3):210.
131. Stewart CM, Kothari PD, Mouliere F, Mair R, Somnay S, Benayed R, et al. The value of cell-free DNA for molecular pathology. J Pathol. 2018;244(5):616–27.
132. De Mattos-Arruda L, Mayor R, Ng CK, Weigelt B, Martinez-Ricarte F, Torrejon D, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun. 2015;6:8839.
133. Beddowes E, Sammut SJ, Gao M, Caldas C. Predicting treatment resistance and relapse through circulating DNA. Breast. 2017;34(Suppl 1):S31–S5.
134. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015;7(302):302ra133.
135. Chu D, Paoletti C, Gersch C, VanDenBerg DA, Zabransky DJ, Cochran RL, et al. ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients. Clin Cancer Res. 2016;22(4):993–9.
136. Fribbens C, O'Leary B, Kilburn L, Hrebien S, Garcia-Murillas I, Beaney M, et al. Plasma ESR1 mutations and the treatment of

estrogen receptor-positive advanced breast cancer. J Clin Oncol. 2016;34(25):2961–8.
137. Ashworth A. Drug resistance caused by reversion mutation. Cancer Res. 2008;68(24):10021–3.
138. Weigelt B, Comino-Mendez I, de Bruijn I, Tian L, Meisel JL, Garcia-Murillas I, et al. Diverse BRCA1 and BRCA2 reversion mutations in circulating cell-free DNA of therapy-resistant breast or ovarian cancer. Clin Cancer Res. 2017;23(21):6708–20.
139. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive molecular portraits of invasive lobular breast cancer. Cell. 2015;163(2):506–19.
140. Desmedt C, Zoppoli G, Gundem G, Pruneri G, Larsimont D, Fornili M, et al. Genomic characterization of primary invasive lobular breast cancer. J Clin Oncol. 2016;34(16):1872–81.
141. Pareja F, Geyer FC, Marchio C, Burke KA, Weigelt B, Reis-Filho JS. Triple-negative breast cancer: the importance of molecular and histologic subtyping, and recognition of low-grade variants. NPJ Breast Cancer. 2016;2:16036.
142. Geyer FC, Pareja F, Weigelt B, Rakha E, Ellis IO, Schnitt SJ, et al. The Spectrum of triple-negative breast disease: high- and low-grade lesions. Am J Pathol. 2017;187(10):2139–51.
143. Geyer FC, Berman SH, Marchio C, Burke KA, Guerini-Rocco E, Piscuoglio S, et al. Genetic analysis of microglandular adenosis and acinic cell carcinomas of the breast provides evidence for the existence of a low-grade triple-negative breast neoplasia family. Mod Pathol. 2017;30(1):69–84.
144. Tognon C, Knezevich SR, Huntsman D, Roskelley CD, Melnyk N, Mathers JA, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. Cancer Cell. 2002;2(5):367–76.
145. Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci U S A. 2009;106(44):18740–4.
146. Kim J, Geyer FC, Martelotto LG, Ng CK, Lim RS, Selenica P, et al. MYBL1 rearrangements and MYB amplification in breast adenoid cystic carcinomas lacking the MYB-NFIB fusion gene. J Pathol. 2018;244(2):143–50.
147. Geyer FC, Li A, Papanastasiou AD, Smith A, Selenica P, Burke KA, et al. Recurrent hotspot mutations in HRAS Q61 and PI3K-AKT pathway genes as drivers of breast adenomyoepitheliomas. Nat Commun. 2018;9(1):1816.
148. Grunewald I, Vollbrecht C, Meinrath J, Meyer MF, Heukamp LC, Drebber U, et al. Targeted next generation sequencing of parotid gland cancer uncovers genetic heterogeneity. Oncotarget. 2015;6(20):18224–37.
149. Chiang S, Weigelt B, Wen HC, Pareja F, Raghavendra A, Martelotto LG, et al. IDH2 mutations define a unique subtype of breast Cancer with altered nuclear polarity. Cancer Res. 2016;76(24):7118–29.
150. Lozada JR, Basili T, Pareja F, Alemar B, Paula AC, Gularte-Merida R, et al. Solid papillary breast carcinomas resembling the tall cell variant of papillary thyroid neoplasms (solid papillary carcinomas with reverse polarity) harbor recurrent mutations affecting IDH2 and PIK3CA: a validation cohort. Histopathology. 2018;73:339.

Genomic Applications in Pulmonary Malignancies

27

Reinhard Büttner, Carina Heydt, and Sabine Merkelbach-Bruse

Introduction

As a result of comprehensive insights into cancer genomes, it has been appreciated that human malignancies arise from a limited set of somatic genetic aberrations driving oncogenic signaling networks. Hence, pathologically altered genes causally driving and maintaining the cancer phenotype are being referred to as "oncogenic drivers." Lung cancer in particular may serve as a paradigm for personalized cancer medicine (PCM) because diagnosing and interfering with the individual set of pathogenic driver mutations result in highly effective, personalized, and frequently less toxic treatment regimens in a significant portion of tumors. This has led to the approval of selective small molecules targeting pathologically activated receptors and signaling molecules (EGFR, ALK, ROS1, BRAF) and a larger pipeline of druggable genetic alterations, which are under current clinical evaluation [1]. More recently immunotherapies targeting signaling pathways driving immune-escape of lung cancer cells have entered clinical practice. Hence, quantitative expression of immunomodulatory cell surface molecules and potential other markers including mutational tumor load emerged as clinically relevant biomarkers for selecting therapies. However, there are strict requirements for discovery and preclinical validation of oncogenic targets as well as for understanding and responding to mechanisms of resistance [2]. Biomarkers sensu stricto define companion diagnostics rather than classical, correlative biomarkers. Lung cancer may also serve as a paradigm for a comprehensive reclassification of tumor entities combining both morphologic and genomic data.

R. Büttner · C. Heydt · S. Merkelbach-Bruse (✉)
Center for Integrated Oncology, Institute of Pathology, University Hospital Cologne, Cologne, Germany
e-mail: sabine.merkelbach-bruse@uk-koeln.de

Driver Genetic Alterations in Lung Cancer: Targets of Approved Therapeutics

EGFR: Response and Resistance to Targeted Therapy

EGFR (HER1) is the prototypic member of the ERBB family of transmembrane tyrosine kinases also comprising HER2, HER3, and HER4 [3]. It is composed of an extracellular growth factor-binding domain, a transmembrane segment, and an intracellular protein-tyrosine kinase catalytic domain. As a result of ligand binding, the inactive receptor monomers undergo conformational change and receptor dimerization. Receptor activation leads to autophosphorylation as well as phosphorylation of tyrosine residues of adaptor or signaling molecules. Thereby, two key oncogenic pathways, the RAS/RAF/MAPK pathway and the PI3K/AKT pathway, are activated promoting cellular proliferation and survival [4]. Due to its function of providing oncogene dependency in NSCLC [5], EGFR has become a therapeutic target using tyrosine kinase inhibitors (TKIs) and/or inhibitory monoclonal antibodies [6].

Since 2004, several investigators have suggested the *EGFR* mutation status but not the *EGFR* gene copy number or expression as the optimal predictor of clinical benefit from EGFR inhibitors in NSCLC [7, 8]. Activating mutations usually occur in the region that encodes the intracellular tyrosine kinase domain and abolish autoinhibition, which keeps the wild-type receptor silent in the absence of ligand [9]. The growth of *EGFR*-mutant NSCLC cells is dependent on aberrant kinase activation. Additionally, the mutant receptor has a higher affinity for the competitive tyrosine kinase inhibitors than for ATP [10].

Several clinical trials have proven the superiority of EGFR tyrosine kinase inhibitors (TKI; i.e., gefitinib, erlotinib, and afatinib) as first-line treatment [11, 12]. Therefore, analysis of *EGFR* activation mutations and subsequent use of appropriate TKIs is now recommended in all guidelines with the exception of patients with definitive diagnosis of

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_27

Table 27.1 Summary of the most common *EGFR* mutations in NSCLC [15–18]

Exon 18	Exon 19	Exon 20	Exon 21
Glu709Ala	Glu746_Ala750del	Val769_Asp770insAlaSerVal	Leu861Gln
Glu709Gly	Glu746_Ser752delinsVal	Asp770_Asn771insSerValAsp	Leu858Arg
Glu709Lys	Leu747_Glu749del	Pro772_His773insProArg	
Gly719Ala	Leu747_Ala750delinsPro	His773_Val774insAsnProHis	
Gly719Ser	Leu747_Thr751del	Thr790Met	
Gly719Cys	Leu747_Ser752del	Ser768Ile	
	Leu747_Pro753delinsSer	Cys797Ser	

squamous cell carcinoma provided they are not "never or former light" smokers [13, 14].

Activating mutations conferring sensitivity to EGFR-TKI can be found in exons 18–21 of the *EGFR* gene (Table 27.1). In-frame deletions of exon 19 which account for about 49% of mutated cases comprise almost always the amino acid residues p.Leu747 to p.Ala750. A rare mutation type in exon 19 is an in-frame insertion of six amino acids accounting for approximately 1% of all *EGFR*-mutant NSCLC [19]. This insertion has been reported to be sensitive to EGFR-TKI with a slightly lower response rate than "classic" EGFR mutations [20].

The second most prevalent mutations are substitutions in exon 21 which account for approximately 42%. Nearly 75% of these substitutions are exchanges from leucine to arginine at codon 858 (p.Leu858Arg) and approximately 10% from leucine to glutamine at codon 861 (p.Leu861Gln).

Mutations in exon 18 account for 9% of all *EGFR*-mutated cases and are predominantly point mutations in codon 719 and less frequently in codon 709 (experiential data from genotyping >2000 cases). Codon 709 of exon 18 is also involved in a rare deletion occurring in exon 18 (p.Glu709_Thr710delinsAsp) [21]. Patients harboring these short deletions as well as point mutations in p.Glu709 or p.Gly719 seem to response to treatment with afatinib [22]. The activity of afatinib in tumors with mutations in p.Gly719 was further supported by the post hoc analysis of three LUX-lung trials which also found clinical benefit for the rare *EGFR* exon 21 mutation p.Leu861Gln [23].

Rare activating aberrations were described in the entire region of the kinase domain in the past few years. The EGFR-kinase domain duplication (EGFR-KDD) results from a tandem duplication of the exons 18–25 encoding the EGFR tyrosine kinase domain. Two case reports suggest the EGFR-KDD confers increased sensitivity to different EGFR-TKI [24, 25]. Recently, the presence of oncogenic *EGFR* fusions joining the EGFR tyrosine kinase domain to various partners, mainly *RAD51*, was reported. The fusion proteins were proven to be oncogenic in vitro, and patients harboring these fusions showed clinical benefit when treated with EGFR-TKI [26].

EGFR exon 20 insertions comprise approximately 4% of all mutations in NSCLC and most frequently occur between amino acid residues 767 and 774 (Table 27.1). They are associated with lower sensitivity to inhibitor-based therapy or even primary resistance [27]. Unlike the *EGFR* mutations described above, exon 20 insertion mutations do not lead to an altered ATP binding pocket, but instead the inserted amino acids form a wedge at the end of the C-helix which fosters the active kinase conformation [28]. In the above cited study from Yang et al. [23], a cohort of patients with insertion mutations in exon 20 was also analyzed. Combining data from three trials, they found an objective response of less than 10% and a median progression-free survival of only 2.7 month on treatment with afatinib. In preclinical studies, activity of the third-generation TKI EGFR816 on three different exon 20 insertion mutations was proven [29].

Another rare primary mutation in exon 20 is the point mutation p.Ser768Ile that occurs with a frequency below 1% and often shows co-occurrence with mutations in exon 18 and exon 21. Whereas the isolated mutation seems to be not sensitive to treatment with first-generation TKI, it does not limit the efficiency of these compounds in combination with sensitizing mutations [30]. In contrast, afatinib was active in tumors harboring p.Ser768Ile [23].

A wide spectrum of techniques has been employed for the detection of *EGFR* mutations. After preparing DNA lysates from microdissected tissue samples, the relevant parts of the *EGFR* gene are amplified by PCR followed by different techniques including Sanger sequencing, pyrosequencing, allele-specific PCR, fragment length analyses, or next-generation massively parallel sequencing techniques.

Most patients who initially respond to tyrosine kinase inhibitor therapies will develop secondary resistance. In 50% of cases, acquired resistance is due to the occurrence of p.Thr790Met mutations (Table 27.1) [31, 32]. This mutation increases the affinity of the binding pocket for ATP, thus interfering with drug binding [33]. Additionally the substitution by the bulky methionine side chain results in steric hindrance of first-generation TKI binding [31]. Osimertinib, an irreversible covalently bound third-generation EGFR-TKI, shows potent and highly specific activity against p.Thr790Met-mediated TKI resistance and has been approved

by the US Food and Drug Administration (FDA) and its European counterpart (EMA) since 2016 [34, 35]. Other resistance mechanisms include *MET* and *ERBB2* amplification, which lead to the activation of parallel signaling pathways [36, 37]. *MET* amplification drives HER3-dependent activation of PI3K [38]. It can either co-occur with the Thr790Met resistance mutation or represents an independent mechanism [39, 40].

Amplification of *ERBB2* has been detected with a frequency of 12–13% in patients with progressive disease following first-line EGFR-TKI treatment [37]. Several studies with ERBB2-directed antibodies and TKIs did not reveal clinical benefit of targeted treatment in NSCLC [41, 42].

Rarely, secondary mutations in downstream effectors such as *BRAF* and *PIK3CA* were identified [36, 43]. Loss of PTEN, which controls the PI3K/AKT signaling pathway, also contributes to resistance to EGFR-TKI [44]. Two pathway-independent mechanisms, namely, transformation to small-cell lung cancer and epithelial-to-mesenchymal transition, have also been described [36, 45].

Furthermore, *MET* and *ERBB2* amplification are mechanisms of resistance to third-generation EGFR-TKIs (osimertinib as well as rociletinib) [46–48]. The mutation p.Cys797Ser in *EGFR*, *MAPK1* amplification, downregulation of negative regulators of ERK, and *NRAS* and *KRAS* mutation or amplification have also been revealed as mechanisms of resistance to third-generation TKI. Hence, current strategies intend to deliver combinatorial therapies in first line (i.e., combination of EGFR and MEK inhibitors and EGFR and MET inhibitors) to avoid or prolonged emergence of resistance [49].

Molecular analysis of the resistance mechanism in formalin-fixed paraffin-embedded (FFPE) tissue can be complemented by analysis of ctDNA from plasma ("liquid biopsy"), especially in cases where the tissue availability is limited due to tumor location or risk to the patient. Additionally, ctDNA analysis overcomes the problem of false negativity due to tumor heterogeneity and allows for repeated sampling while monitoring treatment response.

Assays used to detect the resistance mutation in plasma vary widely in their sensitivity and specificity compared to tissue analysis, and there is a need for standardization of preanalytical plasma processing, ctDNA extraction, and *EGFR* mutation detection methods [50]. Using technologies like BEAMing, sensitivities up to 70% have been reached, and patients with p.Thr790Met positivity detected by plasma testing had similar outcomes with osimertinib compared to patients with p.Thr790Met positivity detected by FFPE tissue testing [51].

One major disadvantage of plasma testing is the inability to detect morphological changes within the tumor (e.g., transformation to small-cell lung cancer). Moreover, in real-world practice, the test sensitivity is also affected by the different stages of disease influencing the release of ctDNA into blood [52]. In summary, given the 30% false-negative rate of plasma testing, p.Thr790Met negative plasma results always should be complemented by tissue testing.

BRAF: Activating and Inactivating Mutations

BRAF is a serine-threonine kinase that mediates the RAS family members' activation of downstream proteins in the MAPK pathway [53]. *BRAF* mutations are found in approximately 3% of Caucasian NSCLC patients. In contrast to melanomas, NSCLC often harbor mutations outside codon p.Val600 [54]. Besides the kinase domain (exon 15), the G loop of the activation domain encoded by exon 11 may be mutated [55, 56]. Mutations outside codon p.Val600 can be either kinase activating (e.g., p.Lys601Glu, p.Leu597Gln, p.Gly464Val, or p.Gly466Ala) or inactivating (e.g., p.Asp594Gly, p.Gly466Val) [57]. In contrast to mutations in codon p.Val600, non-p.Val600 mutations are found to be associated with activating *KRAS* mutations [57].

BRAF fusions with different fusion partners are described in less than 1% of NSCLC, which were all of adenocarcinoma subtype [58]. Another study found the *SND1-BRAF* fusion in about 3% of lung adenocarcinoma from never smokers. In vitro experiments showed an increase in phosphorylation of the MAPK pathway as well as cell proliferation and spheroid formation [59].

Case reports showed promising responses to vemurafenib [60, 61], and in a vemurafenib basket trial, the overall response rate (ORR) in 20 NSCLC patients carrying a p.Val600Glu mutation was 42% [62]. With dabrafenib monotherapy, the ORR was 33% [63]. Due to the compensatory increase in KRAS signaling triggered by the single-agent BRAF inhibition, the responses typically last up to only 5 months [64]. Therefore, the combined inhibition of BRAF and MEK using dabrafenib and trametinib has also been explored [65]. The high activity of this combination (ORR 63%) reached in NSCLC with p.Val600Glu mutations led to the approval by the EMA in April 2017.

Up to now, there are no clinical data for treatment response in NSCLC patients with non-p.Val600 mutations. p.Val600-mutated *BRAF* acts as activated monomer, whereas activating mutations outside codon 600 lead to BRAF dimers which render the tumors insensitive to BRAF TKIs [66]. In preclinical studies, a combination of vemurafenib and trametinib showed higher efficacy compared to the single-agent approach [67].

The common point mutations occurring in limited regions of the *BRAF* or *KRAS* genes can be detected with single gene assays like allele-specific PCR, high-resolution melting analysis, or primer extension assays. For the p.Val600Glu mutation in exon 15 of *BRAF*, these assays have been in rou-

tine use for several years for the detection of targeted therapy options in melanoma. Special attention has to be paid to the fact that half of the *BRAF* mutations occur in codons outside p.Val600, which may become important for the selection of therapy options in the future.

ALK and ROS1: Chromosomal Inversions and Translocations and Therapy Resistance

The *ALK* gene encodes a receptor tyrosine kinase, referred to as anaplastic lymphoma kinase. This designation derives from anaplastic large-cell lymphomas, in which a gene fusion between *ALK* and *NPM* (*nucleophosmin*) was first detected [68]. In 2007, fusion of *ALK* with the upstream partner *EML4* was found in NSCLC [69]. The fusion is the result of inversion in chromosome 2.

Approximately 3–5% of lung adenocarcinomas harbor *ALK* rearrangements [69–71]. Different fusion variants have been reported, all comprising the entire tyrosine kinase domain of *ALK* and varying portions of the *EML4* gene [72]. The fusion results in protein dimerization and therefore constitutive activation of the kinase function [69, 70]. Most of the reported variants start with exon 20 of *ALK* as the first exon of the 3′ part, but, rarely, variants starting with exon 19 are described [73, 74]. In the past few years, at least 27 variants of ALK fusion were described [75]. *EML4-ALK* fusions are usually found in tumors with *EGFR* and *KRAS* wild-type sequences, and positivity was found to be associated with resistance to EGFR-targeted inhibitors [76, 77]. In a phase I clinical trial, 57% of patients with an *EML4-ALK* fusion showed an overall response to the dual ALK/MET inhibitor crizotinib [78]. Based on these results, the compound received regulatory approval for clinical use in the USA in 2011. Two trials comparing crizotinib versus chemotherapy in first- or second-line setting found significantly longer progression-free survival in the crizotinib group and thereby were the basis for approval in the EU 1 year later [79, 80].

In addition to the fusion variants describe above, fusion partners other than *EML4* have been identified, including *KIF5B*, which is a microtubule-based motor protein involved in organelle transport. The translocation t(2;10)(p23;p11) results in the fusion of the first domains of *KIF5B* including the motor domain and the coiled-coil domain with the tyrosine kinase domain of *ALK* [81]. Another rarely occurring fusion partner is *TFG* (TRK-fused gene), and, recently, *HIP1-ALK* was described as fusion variant responding to crizotinib [82, 83]. Due to small patient numbers, it remains not clear whether the specific fusion type influences the sensitivity to ALK inhibition. A retrospective analysis of 55 patients with *EML4-ALK* translocation revealed a higher disease control rate and longer median progression-free survival under crizotinib treatment in patients with *EML4-ALK* variant 1 compared to other variants [84].

The presence of several fusion variants described above makes it difficult to detect the fusion by RT-PCR. Therefore, the core approach for testing chromosomal rearrangements in FFPE tissue is still fluorescence in situ hybridization (FISH). For *ALK* testing, numerous studies have shown excellent concordance between immunohistochemistry (IHC) and FISH [85, 86], and thereby ALK-IHC is now widely used for patient screening. Depending on the antibody used, *ALK* FISH is recommended as confirmatory test in IHC 1+ or 2+ positive or equivocal cases [87], but also methods based on extracted nucleic acids such as amplicon-based parallel sequencing or hybridization-based NanoString technology can be used for orthogonal confirmation (Fig. 27.1) [88–91].

The success of crizotinib therapy is limited by the development of acquired drug resistance [92–94]. Approximately 20% of patients show secondary mutations in the tyrosine kinase domain of *ALK* leading to resistance, but also *ALK* copy-number gain; emergence of other oncogenic driver mutations, e.g., in *EGFR* or *KRAS*; or amplification of *KIT* has been detected [93–95]. Approximately one-third of resistant patients develop brain metastases. This might be due to the poor ability of crizotinib to cross the blood-brain barrier which is the consequence of drug efflux mediated by the ABCB1 pump [96]. Next-generation ALK inhibitors such as ceritinib and alectinib, overcoming different resistance mutations and showing improved CNS penetrance, were developed and tested in clinical phase I/II studies in crizotinib-pretreated as well as crizotinib-naïve patients with response rates around 50–60% [97, 98]. Both have been approved by the FDA and the EMA for use after progression on crizotinib. However, despite the higher inhibitory potency, patients also relapse upon treatment with second-generation TKI. More than half of the patients show resistance mutations in the fused *ALK* gene, and also the spectrum of mutations differed from that detected upon resistance to first-generation TKI [99].

Brigatinib is another second-generation ALK inhibitor which shows activity against several *ALK* resistance mutations and has recently received therapy breakthrough approval status from the FDA [99]. Finally, lorlatinib, a third-generation inhibitor, is effective against all known resistance mutations [99]. Both compounds are currently tested in preclinical and clinical studies. In the near future, continuous assessment of different resistance mechanisms in repeated biopsies following disease progression will be of importance in choosing the right therapy options.

The *ROS1* gene is located on chromosome 6 and encodes a receptor tyrosine kinase of the insulin receptor family. The chromosomal rearrangement of *ROS1* as driver mutation in lung cancer was first described by Rikova [83].

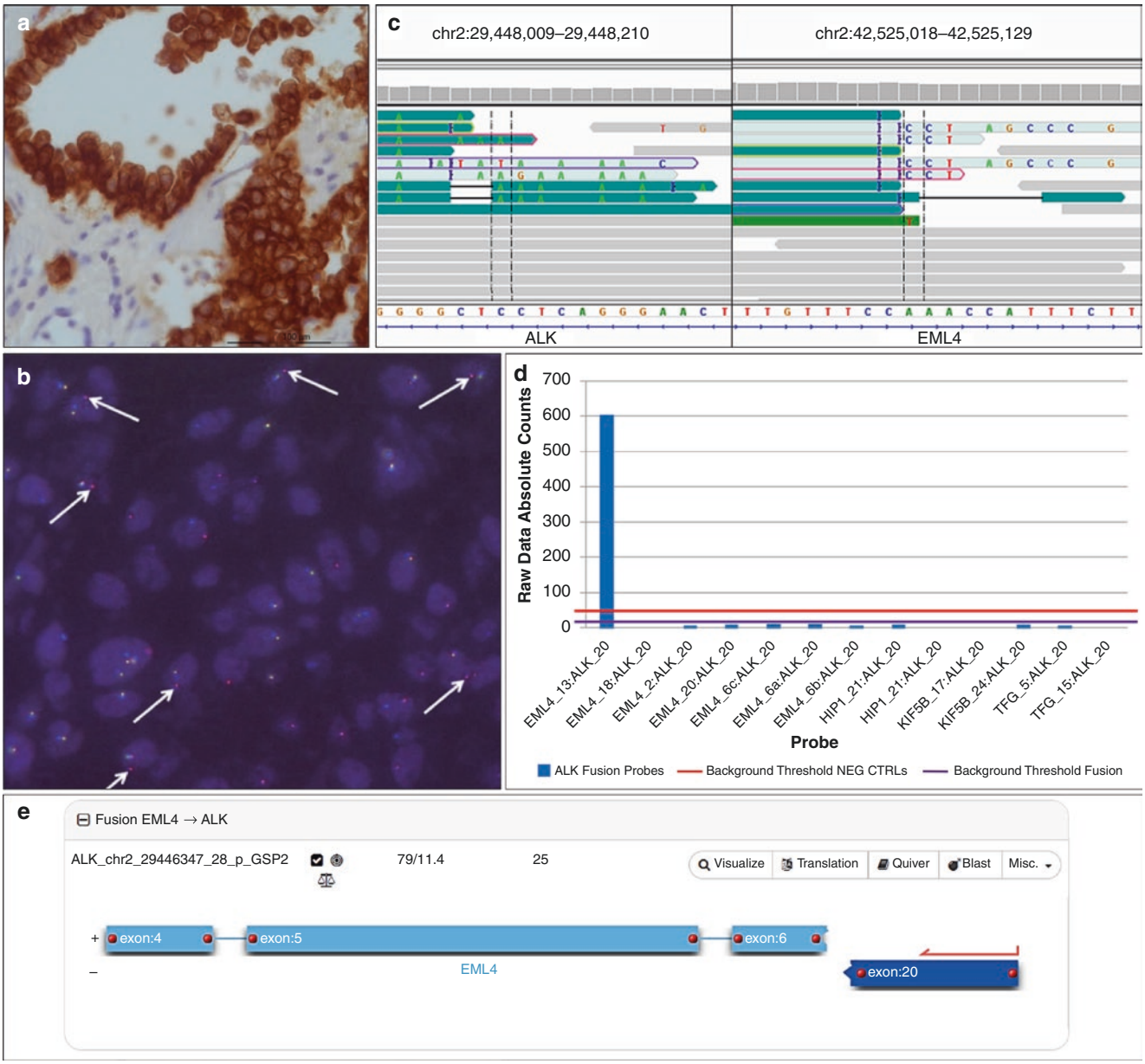


Fig. 27.1 Detection of *ALK* rearrangement with different methods. (**a**) ALK-IHC analysis was carried out with the mouse monoclonal antibody clone 1A4. ALK-positive tumor cells demonstrate strong ALK-IHC staining at a magnification of 400× (**c**) (Zytomed, Berlin, Germany). (**b**) FISH analysis was carried out with the ZytoLight ® SPEC ALK/EML4 TriCheck™ Probe (ZytoVision, Bremerhaven, Germany): Orange and green fluorescent signals for the *ALK* break-apart probes and blue fluorescent signals for the *EML4* probe. The arrows point to *EML4-ALK* rearranged nuclei. (**c**) Hybrid capture-based sequencing data visualized on chromosome 2 by the IGV (Integrative Genomics Viewer). Highlighted are normal *ALK* and *EML4* reads (gray), as well as *EML4-ALK* fusion reads identified by discordant and split reads (colored reads). *EML4-ALK* paired fusion reads are highlighted with the same colored border. (**d**) Detection of *EML4-ALK* rearrangement with the NanoString nCounter technology. The *EML4-ALK* fusion was detected with a specific *ALK* gene fusion probe. (**e**) RNA-based parallel sequencing assay results visualized with the ArcherDX analysis software (Boulder, CO, USA). (Illustration of the EML4-ALK fusion. Shown is the exact fusion point of the *EML4* and the *ALK* gene)

Rearrangements of *ROS1* by translocation or interstitial deletion leading to a fusion with several different partners have been detected in 0.8–1.7% of NSCLC [100–102]. The fusion products contain the intact tyrosine kinase domain of *ROS1* and a truncated fraction of the fusion partner, e.g., *TPM3*, *SDC4*, *SLC4A2*, *CD74*, *EZR*, *FIG*, or *LRIG3* [100, 103, 104]. *ROS1* rearrangements were found to be one of the best prognostic factors in NSCLC and rarely co-occur with other genetic drivers [105, 106].

Preclinical data suggest that rearranged *ROS1* can be targeted by ALK inhibitors due to highly similar tyrosine kinase domains [107]. Clinical studies confirmed this observation

and showed that rearranged *ROS1* can be efficiently inhibited by the tyrosine kinase inhibitor crizotinib [108]. Consequently, in 2016, crizotinib received regulatory approval for clinical use in the USA and the EU.

As for ALK testing, IHC is an effective screening tool to detect *ROS1*-positive samples. ROS1 protein expression may be observed in normal cells, and evaluation has to be done in comparison with an external positive control [109, 110]. Further confirmation by *ROS1* FISH is recommended in IHC-positive or doubtful cases. Alternatively, the same technologies using extracted nucleic acids as described for *ALK* testing can be used.

Like its *ALK* counterpart, *ROS1*-rearranged lung adenocarcinoma develops acquired drug resistance upon crizotinib therapy [111–113]. Most patients show secondary mutations in the tyrosine kinase domain of *ROS1* leading to resistance. Activating *KIT* mutations as well as alternative pathway activation have also been detected as a resistance mechanism [94, 114–118]. Several other ROS1 inhibitors like cabozantinib, lorlatinib, and entrectinib are currently under clinical investigation and showed efficacy in cases with acquired resistance [119–122].

Driver Genetic Alterations in Lung Cancer: Targets in Clinical Studies

KRAS: Response to TKI Treatment and Therapeutic Options

The *KRAS* gene encodes a member of the family of membrane-bound GTP-binding proteins that regulate proliferation, differentiation, and apoptosis through the MAPK, STAT, and PI3K signaling pathways [123]. Activating point mutations occurring in the GTPase domain of *KRAS* are found in 30% of NSCLC [2]. More than 95% of mutations are localized in codons 12 and 13 (exon 2), the predominant mutation being p.Gly12Cys (42.3% of all *KRAS* mutations). These mutations are mutually exclusive with those in *EGFR*. *KRAS* mutations were proposed to be a negative prognostic factor in NSCLC [124, 125]. Additionally they are associated with resistance to EGFR inhibitor-based therapy as confirmed by two meta-analyses that found *KRAS* mutations to be a negative predictor of response to single-agent EGFR-TKIs in advanced NSCLC [126, 127].

Inhibition of KRAS is difficult as the mutated RAS protein harbors reduced GTPase activity. For the p.Gly12Cys mutation, specific inhibitors have shown preclinical activity and are under clinical evaluation [128]. A more promising approach than targeting the mutant *KRAS* itself seemed to be the inhibition of signaling pathways downstream of *KRAS*, but targeting MEK alone showed only modest efficiency with response rates of 11% for selumetinib and 12% for trametinib [129, 130]. Preclinical studies revealed the efficacy of combined inhibition of the PI3K and MAPK pathways [131, 132]. Clinical studies combining MAPK inhibitors either with a CHK1 inhibitor (ralimetinib and prexasertib, NCT02860780 [133, 134]) or a pan-RAF inhibitor (LTT462 and LXH254, NCT02974725 [134, 135]) are under way. There is growing evidence for genetic heterogeneity of NSCLC bearing *KRAS* mutations as they may co-occur with different additional genetic changes such as *TP53*, *STK11*, or *CDKN2A/B* mutations and show different gene expression profiles depending on the co-mutation [136]. As for *BRAF*, the point mutations in *KRAS* can be easily analyzed with sensitive single gene assays like allele-specific PCR or primer extension assays. Like other genes (e.g., *ERBB2* or *MET*), *KRAS* status can be assessed as part of larger testing panels and can be useful to "rule out" other less common driver alterations in NSCLC that occur in mutually exclusive fashion.

ERBB22: Amplification and Mutation

ERBB2 is the sole member of the ERBB receptor family without an identified ligand, and hence, it is activated through dimerization with other members of the ERBB family. Activation of the ERBB2 tyrosine kinase domain turns on signaling via the MAPK and the PI3K pathways.

ERBB2 gene amplification assessed by fluorescence in situ hybridization (FISH) is demonstrable in only 2% of NSCLC, whereas 20% of cases test positive by immunohistochemistry. This is thought to be due to polysomy of chromosome 17 which occurs in 81% of NSCLC [137]. Possible benefit from therapy with trastuzumab, an antibody to ERBB2, was reported for a group of patients with strong ERBB2 overexpression and FISH positivity, but not for patients solely selected by immunohistochemistry [41, 138]. So far, clinical data do not support routine clinical use of ERBB2-directed therapies in patients with *ERBB2* amplification or overexpression [139], but recent in vitro studies have proven activity of afatinib in NSCLC cell lines as well as in xenograft mouse models [140].

ERBB2 mutations are present in 2–4% of NSCLC and occur mostly in exon 20 [141]. In the majority of cases, mutations are in-frame insertions or duplications of amino acids around codon 775 in exon 20 which are mutually exclusive with mutations in *KRAS*, *EGFR*, *BRAF*, and *PIK3CA*, as well as with *ALK* rearrangements [142, 143]. These *ERBB2* aberrations lead to constitutive receptor activation [144] and occur in a structurally analogous position as exon 20 insertion/duplication mutations in *EGFR*. Two recently identified mutations in the transmembrane domain of ERBB2 encoded by exon 17, p.Val659Glu and p.Gly660Asp, also seem to be onco-

genic in lung adenocarcinoma [140]. Mutations in exon 19, e.g., p.Leu755Ser, which are described in breast and gastrointestinal tumors, occur only rarely in NSCLC [142].

Over the past years, several tyrosine kinase inhibitors targeting simultaneously EGFR and ERBB2 were investigated in preclinical as well as clinical studies for *ERBB2*-mutated NSCLC. In a phase II trial, dacomitinib, an irreversible pan-EGFR inhibitor, showed durable partial responses in 12% of NSCLC cases with *ERBB2* mutations [145]. In this study, durable responses were only shown in patients with a three-amino-acid duplication (p.Gly778_Pro780dup), but none of the 13 patients with the most common duplication p.Tyr772_Ala775dup responded. In an in vitro study, three additional rarer mutations in exon 20 of *ERBB2* were identified as being sensitive to dacomitinib [146].

A partial response to the irreversible tyrosine kinase inhibitor afatinib was described in a small study for three out of three patients harboring *ERBB2* exon 20 mutations [147]. One of the patients in this study was reported with the above-described common duplication. The finding that afatinib might be effective in patients carrying this specific mutation was further confirmed in a case report of a patient with p.Tyr772_Ala775dup showing durable response to afatinib [148]. Additionally, one patient carrying this mutation was reported to respond to trastuzumab emtansine (T-DM1) [149].

Kris et al. summarized their own data and results from other preclinical and clinical studies [145] and concluded that the benefits of an individual agent may be confined to a specific aberration. The differences between the effects of various TKIs on various *ERBB2* mutations suggest that not all *ERBB2* mutations are alike although the reason for this variation remains unknown. Specific alterations have to be proven by comprehensive tumor sequencing, and *ERBB2* might be part of larger testing panels for parallel sequencing.

MET: Activation in Untreated NSCLC

The receptor tyrosine kinase hepatocyte growth factor receptor (HGFR) is encoded by the *MET* gene located on chromosome 7 [150]. *MET* point mutations in the semaphorin and juxtamembrane domains occur with a low frequency (1–3%) in NSCLC [151–156]. *MET* mutations are mutually exclusive with mutations in *EGFR*, *KRAS*, and *ERBB2* and fusions in *ALK*, *RET*, *ROS*, and *ERBB2* [151, 153, 156]. Some of the previously reported *MET* mutations seem to represent SNPs; thus, their clinical importance is highly questionable [157]. Mutations affecting the splice site between exons 13 and 14 in the juxtamembrane of *MET*, which leads to a deletion of exon 14 (exon 14 skipping), are associated with enhanced ligand-mediated proliferation and tumor growth [154, 156]. More than 100 mutations resulting in *MET* exon 14 skipping have been reported thus far. Case reports as well as preclinical studies have shown that *MET* exon 14 skipping mutations have increased sensitivity to MET inhibitors [151–153, 158–161]. Over the last years, different case reports have evaluated the two tyrosine kinase inhibitors crizotinib (ALK/MET inhibitor) and capmatinib (MET inhibitor). Paik et al. showed four cases with partial response to crizotinib and stable disease under cabozantinib therapy [162]. Several studies showed partial or complete response to crizotinib (nine cases) and partial response to capmatinib (two cases by Frampton et al.) in the presence of a *MET* exon 14 splicing event [151, 153, 159, 161]. Clinical studies investigating the role of MET inhibitors in *MET* exon 14-mutated tumors are under way [134]. *MET* exon 14 skipping mutations can be found in the presence or absence of MET amplifications [151, 153, 156, 162].

In NSCLC not previously treated with EGFR-specific tyrosine kinase inhibitors, high-level *MET* amplification is detected in approximately 2–3% and is associated with poor prognosis [155, 156, 163, 164]. Only limited data about the treatment of *MET*-amplified EGFR treatment-naïve NSCLC patients are available. Whether the so-called low-level *MET* amplification resulting in part from polysomy confers oncogene dependency remains to be shown. The efficiency of MET TKIs and monoclonal antibodies is currently under investigation. Response to crizotinib has been reported in lung cancer cells with *MET* amplification [165, 166]. Capmatinib alone, or in combination with the immune checkpoint inhibitor nivolumab, is currently evaluated in patients with wild-type *EGFR* status and *MET* amplification [134, 167]. It has also been shown that a *MET* amplification is not a mutually exclusive genetic event. Low-, medium-, and high-level amplifications can co-occur with *EGFR* and *KRAS* mutations in a therapy-naïve tumor; therefore, it might be necessary to administer an EGFR-TKI and MET inhibitor at the same time in *EGFR*-mutated and *MET*-amplified patients [153, 164].

MET amplification is more common in previously treated NSCLC patients and has shown to be a mechanism of resistance to third-generation EGFR-TKIs (osimertinib as well as rociletinib) as described above.

RET and NTRK1: Chromosomal Inversions and Translocations

The *RET* gene (rearranged during transfection) located on chromosome 10 encodes a receptor tyrosine kinase that belongs to the RET family. Activation of RET results in increased MAPK and PI3K pathway signaling [168]. The inversion inv.(10)(p11.22q11.2) was detected in adenocarci-

noma of the lung and results in the fusion of *KIF5B* with *RET*, which encodes a receptor tyrosine kinase. The fusion transcript comprises the coiled-coil domain of *KIF5B* and the tyrosine kinase domain of *RET*. It is detected in 0.7–2% of lung adenocarcinomas [169–171], and its occurrence is mutually exclusive with other driver mutations in *EGFR*, *KRAS*, *BRAF*, and *HER2* [172]. It has been shown to co-occur with missense and nonsense mutations in *TP53* [171]. Several other fusion partners like *CCDC6*, *NCOA4*, and *TRIM33* were identified [173].

Cells expressing the *KIF5B-RET* fusion transcript are sensitive to multikinase inhibitors [174]. The *RET* fusion protein might provide a new therapeutic target as shown in several phase II clinical studies with different tyrosine kinase inhibitors. Only moderate activity was detected for vandetanib in a phase II clinical trial enrolling 18 patients tested by FISH [175]. In another phase II study, stratifying patients according to their specific fusion partner revealed a much higher sensitivity toward vandetanib for the *CCDC6-RET* subtype (ORR, 83% vs. 20%) [176]. Alectinib, an ALK inhibitor, has shown activity against *RET*-translocated tumor cells in vitro [177] and preliminary antitumor activity in advanced RET-translocated NSCLC [178]. Two out of three *RET*-positive patients presented with partial responses in a phase II trial with cabozantinib [179], but a larger study showed only moderate activity for this inhibitor [180]. To summarize, the available tyrosine kinase inhibitors have limited activity in *RET*-translocated NSCLC, and further investigation of the tumor biology is needed.

Unlike for *ALK* and *ROS1* testing, no validated RET antibody for IHC use is available. Keeping in mind that the fusion partner seems to be important for the therapy decision, nucleic acid-based methods like massively parallel sequencing or NanoString technology might be the detection method of choice.

NTRK1 (neurotrophic tyrosine kinase receptor, type 1) belongs to the tropomyosin-related kinases superfamily. The *NTRK* gene is located on chromosome 1q21–22 and encodes the high-affinity nerve growth factor receptor (TRKA). When activated by the ligand NRG, TRKA controls the MAPK, PI3K, and PLC-γ pathways [181, 182].

Rearrangements of *NTRK1* are identified in 0.1% of all NSCLC patients [183]. In patients without other known driver mutations, the frequency increases to 3%. The entire kinase domain of TRKA is involved in the fusion events, and fusion partners include *MPRIP* and *CD74* [184]. Treatment with entrectinib, a pan-TRK inhibitor, which also shows activity against ALK and ROS1 fusion proteins, showed promising results in preclinical models in patients with *NTRK* translocation [122, 185, 186]. In other solid tumor entities, response to the tropomyosin-related kinase inhibitor LOXO-101 has been reported [187].

Molecular Alterations in Squamous Cell Carcinoma

Although recurrent alterations including mutations in *TP53*, *CDKN2A*, *PIK3CA*, *DDR2*, *NFE2L2*, *KEAP1*, *FGFR2*, *FGFR3*, and *MLL2* and amplifications of *FGFR1* have been identified in squamous cell carcinoma, only limited personalized therapy strategies have been developed [188]. Testing for *EGFR* mutations and *ALK* and *ROS1* rearrangements is only recommended for young patients and never or former light smokers [13, 189].

The FGFR (fibroblast growth factor receptor) family comprises four members which are activated by 18 ligands (fibroblast growth factors, FGFs) and thereby stimulate multiple pathways including MAPK and the PI3K signaling cascades [190]. A large screen for *FGFR* aberrations in solid tumors found 9% incidence of *FGFR1* amplification, 3% of each *FGFR2* and *FGFR3* mutations and 1% of *FGFR4* mutations in a cohort of 93 squamous cell carcinoma of the lung. This is in contrast to only 4% of lung adenocarcinoma harboring any *FGFR* aberrations [191]. In preclinical studies it was already proven that some of these mutations in *FGFR2* and *FGFR3* confer sensitivity toward FGFR kinase inhibitors [192]. Similar rates of *FGFR* aberrations was reported in a smaller study analyzing 75 squamous cell carcinoma that additionally found *FGFR1* mutations in 2.7% of cases and one case with *FGFR3-TACC3* fusion [193]. Copy-number changes and mutations were mutually exclusive, and the presence of FGFR aberrations in general was an indicator for significantly worse prognosis among patients with disease recurrence after surgery.

Several in vitro studies described inhibitor sensitivity of the *FGFR1* amplification in squamous cell carcinoma [194, 195]. The region on chromosome 8 spanning the *FGFR1* locus is amplified in about 9–20% of lung cancer patients and is associated with smoking. The standard method for the detection of gene copy-number changes in FFPE tissue is FISH. The patterns of amplification are diverse, mainly due to the incidence of different degrees of polysomy in the tumor cells [196]. In a phase I study with the FGFR tyrosine kinase inhibitor BGJ398, the overall response rate of 11% in patients with squamous cell carcinoma of the lung was lower than expected from clinical data suggesting that the *FGFR1* amplification alone is not the sole driver in these tumors [197].

Discoidin domain receptors (DDR) 1 and 2 were identified as targets of imatinib, nilotinib, and dasatinib [198]. Gain-of-function mutations in *DDR2* have been described in about 4% of squamous cell carcinoma of the lung with no hotspots in their distribution [199, 200]. A squamous cell lung cancer patient with the point mutation p.Ser768Arg responded to treatment with dasatinib and erlotinib in an

early-phase clinical trial [200]. One of the challenges in advancing *DDR2* mutations as therapeutic targets is the lack of recurrent point mutations as well as the limited number of these mutations [201]. As the TCGA datasets show overlap, for example, with *KRAS* mutation, it seems that additional drivers are relevant in *DDR2*-mutated lung cancer, and up to now clinical studies could not confirm the preclinical results.

Phosphatidyl 3-kinases (PI3K) are heterodimeric lipid kinases, composed of a catalytic and a regulatory subunit, and involved in a wide range of vital cellular processes including proliferation and differentiation. The *PIK3CA* gene encoding the catalytic subunit is frequently mutated in human cancers [202]. Somatic mutations are found in 1–3% of NSCLC [203, 204] and occur within two hot spot regions, the helical binding domain encoded by exon 9 (p.Glu542Lys and p.Glu545Lys) and the catalytic subunit encoded by exon 20 (p.His1047Arg or Leu). Mutations seem to be more common in tumors with squamous cell histology than in adenocarcinoma (6.5% versus 1.5%) and are not mutually exclusive with *EGFR*, *KRAS*, or *BRAF* mutations [205]. In up to 70% of cases, coexisting mutations were found, leading to doubt regarding the role of *PIK3CA* lesions as driver mutations conferring oncogene dependency [203, 204]. Analyzing a cohort of 308 early-stage squamous cell carcinoma, Mc Gowan et al. found that patients with *PIK3CA* mutations had a significantly longer overall survival and time to relapse than non-mutated cases [206]. However, for stage IV patients, association of *PIK3CA* aberrations with worse survival, higher metastatic burden, and greater incidence of brain metastases was described [160]. As PD-L1 expression is less frequent in *PIK3CA* aberrant cases, it was also speculated that *PIK3CA* mutations might serve as biomarker for the response to immune checkpoint targeted therapies [206, 207].

The incidence of *PIK3CA* amplification in NSCLC is 31% with the overwhelming majority (93.3%) occurring in squamous cell carcinoma and the remainder in adenocarcinoma [208].

Multiple drugs targeting the PI3K pathway are currently under development. These include pan and selective inhibitors of PI3K, AKT inhibitors, rapamycin and rapalogs for mTOR inhibition, dual mTORC1-mTORC2 inhibitors, and dual PI3K-mTOR inhibitors. Several preclinical and clinical trials are ongoing [209].

A comprehensive description of standards for the molecular analysis of lung cancers can be found in the "Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for treatment with Targeted Tyrosine Kinase Inhibitors" published by the College of American Pathologists (CAP), the International Association for the Study of Lung Cancer (IASLC), and the Association for Molecular Pathology (AMP) [210].

Molecular Diagnostics Methods

Characteristics of FFPE Tissue

Formalin-fixed and paraffin-embedded tissue (FFPE) remaining after conventional histological and immunohistochemical stainings is widely used for molecular analysis because fresh frozen tumor tissue is only rarely available. The integrity and stability of DNA in FFPE is a limiting factor for the reliability of mutation testing [211]. During tissue fixation, DNA quality is affected mainly by degradation of target DNA due to the reaction of the phosphodiester backbone with formalin. The main factors affecting the degree of DNA degradation in FFPE are the duration of specimen archiving, the type of fixative used, and the duration of fixation prior to paraffin embedding [212, 213]. The small size of the biopsy presents another challenge to molecular analysis, especially in lung cancer as fine needle aspiration biopsy is a commonly used procedure. In practice, however, these parameters are highly variable, particularly in a reference lab setting where the material is received from different diagnostic centers.

During the fixation process, formalin causes deamination of cytosine and adenine, resulting in uracil or hypoxanthine residues in the template DNA [214]. During subsequent PCR, these changes result in C → T/G → A or A → G/T → C transitions. Given that PCR starts from few templates, especially in biopsy samples with small amounts of extracted genomic DNA and thereby low copy numbers of the desired fragment, these artifacts may be amplified and detected as false mutations. C → T/G → A transitions can be prevented by using uracil-*N*-glycosylase prior to PCR. Thereby, uracil is removed from the DNA strands and a strand break is created [215]. The occurrence of artifacts can be circumvented by using higher amounts of template DNA wherever available. If that is not possible, we strongly recommend performing multiple independent PCR amplifications from the same sample.

Our group and others have suggested to use 4% buffered formalin for biopsy specimens for at least 6 h and overnight for larger specimens for optimal fixation [216, 217]. An important aim of the extraction protocol is to preserve the quality of the DNA as well as possible by applying gentle extraction procedures [218]. For subsequent analyses, it is required to purify raw DNA extracts from potential PCR inhibitors such as hemoglobin or melanin [219]. In our experience, automated systems are favored over manual systems due to better standardization of handling steps and the use of magnetic beads for DNA binding instead of spin columns.

During all steps of the analysis process, the prevention of carry-over contaminations is an important issue. Precautionary measures are described below together with the individual analysis steps and the corresponding notes.

A comprehensive description of the whole workflow can be found in the "Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors" published by the Association for Molecular Pathology [220] and available online (http://www.amp.org/documents/LungBiomarker-AMP-2013-proof.pdf).

Isolation of Nucleic Acids

Macrodissection

Prior to DNA and RNA extraction, an experienced pathologist has to evaluate a hematoxylin and eosin (H&E)-stained tissue section to define the most appropriate tissue area. Furthermore, the pathologist ideally would mark on the H&E-stained section the tumor area for appropriate manual microdissection in order to reduce the amount of non-tumorous tissue. Laser-capture microdissection is also possible in principle; however, it is labor-intensive and difficult to apply in a high-throughput setting.

For manual microdissection, two different methods can be used: First, the marking of the H&E-stained slide is transferred to the tissue block before cutting one to six 10-µm-thick section rolls in a reaction tube. This method impairs the quality of tissue blocks for further investigations. The alternative protocol is to prepare tissue sections first and to perform the manual microdissection on glass slides after deparaffinization.

Avoiding Contamination During Nucleic Acid Isolation and Amplification

In general, because PCR amplification is a highly sensitive method and susceptible to carry-over contamination, the following fundamental rules should be the gold standard in every diagnostic molecular pathology laboratory. The working area should be divided in pre- and post-PCR sections, optimally three independent rooms for extracting DNA, preparing the PCR, and performing the post-PCR analysis. Each section should have specifically assigned equipment and reagents. Plugged pipette tips have to be used throughout and gloves to be changed between the working sections. Also, during cutting the paraffin blocks by microtome, some rules have to be observed to avoid contamination between blocks:

1. Blocks that are proposed for the same analysis should not be cut consecutively.
2. The microtome and the working area should be cleaned from paraffin material after each individual block.
3. Depending on the type of microtome, the blades should be removed or relocated after each block.
4. If the sections are mounted on slides for manual microdissection, the water bath should be cleaned regularly with filter paper, and the water should be changed as often as possible.

Purification of Nucleic Acids

The need for efficient nucleic acid extraction from FFPE material has increased over the years as new molecular diagnostic methods require high quality and quantity of nucleic acids. Over the last years, different in-house as well as well as commercially available kits have been developed and evaluated for DNA, RNA, and/or total nucleic acid extraction [221–225]. The majority of commercially available assays are based on the purification by silica. Silica has the ability to bind nucleic acids in the presence of salts [226–228]. There are two ways in which silica is used in kits, either as silica coated-membranes in spin columns or as silica-coated paramagnetic beads. As initial step, the paraffin has to be removed from the tissue, and the tissue has to be lysed with proteinase K to release the nucleic acids from the tissue. There are different ways to remove the paraffin from the tissue, e.g., with a combination of xylene and ethanol or with mineral oil. There are also some kits that do not include a paraffin removal step. After lysation of the tissue, the lysate can be purified. When using spin columns, the lysate is added to a spin column, followed by a centrifugation step, where the lysate is driven through the spin column and the nucleic acids bind to the silica-coated membrane. This step is followed by ethanol washes of the membrane-bound nucleic acids to remove contaminants like proteins. Then the DNA or RNA is released under low-salt conditions with nuclease-free water, TE buffer, or TRIS-HCl.

In paramagnetic bead-based nucleic acid purification kit, the lysate is added to the magnetic beads, and the silica on the beads binds the nucleic acids. The bound nucleic acids can then be separated from the solution by magnetic force. The bound nucleic acids are then washed with ethanol and released with nuclease-free water, TE buffer, or TRIS-HCl.

For nucleic acid purification from FFPE material, commercially available assays are specifically optimized for FFPE RNA or FFPE DNA extraction. In the RNA extraction protocols, most often a DNase digestion step is included to remove the DNA. To extract total nucleic acids, the RNA extraction kits can be used without the DNase digestion step. There are also kits available which combine the simultaneous extraction of RNA and DNA from one sample.

Additionally, the commercially available kits are either designed for manual extraction or for manual and automated extraction or for automated extraction only. Especially to those laboratories with a high sample throughput, automated extraction systems are essential. For automated nucleic acid extraction, different automated extraction machines are available from different companies [221, 224].

Selected Protocols

(a) Manual extraction FFPE DNA
- ReliaPrep FFPE gDNA Miniprep System (Promega): https://www.promega.de/products/dna-purification-quantitation/genomic-dna-purification/reliaprep-ffpe-gdna-miniprep-system/
- QIAamp DNA FFPE Tissue Kit (Qiagen): https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/qiaamp-dna-ffpe-tissue-kit/#orderinginformation
- Ion AmpliSeq Direct FFPE DNA Kit (Thermo Fisher Scientific): http://www.thermofisher.com/de/de/home/life-science/dna-rna-purification-analysis/dna-extraction/genomic-dna-extraction/dna-extractions-working-with-ffpe-samples.html
- High Pure FFPET DNA Isolation Kit (Roche): https://shop.roche.com/shop/products/high-pure-ffpet-dna-isolation-kit

(b) Manual extraction FFPE RNA
- ReliaPrep™ FFPE Total RNA Miniprep System (Promega): https://www.promega.de/products/rna-purification-and-analysis/rna-analysis-workflow/rna-purification/reliaprep-ffpe-total-rna-miniprep-system/
- RNeasy FFPE Kit (Qiagen): https://www.qiagen.com/us/shop/sample-technologies/rna/total-rna/rneasy-ffpe-kit/#orderinginformation
- RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific): https://www.thermofisher.com/de/de/home/life-science/dna-rna-purification-analysis/rna-extraction/rna-sample-extraction/working-with-ffpe-samples.html
- High Pure FFPE RNA Micro Kit (Roche): https://shop.roche.com/shop/en/de/products/high-pure-ffpe-rna-micro-kit

(c) Automated extraction:
- Maxwell RSC (Promega): https://www.promega.de/products/instruments/maxwell-instruments/maxwell-rsc-instrument/
- QIAcube (Qiagen): https://www.qiagen.com/us/shop/automated-solutions/sample-preparation/qiacube/#orderinginformation
- QIAsymphonie (Qiagen): https://www.qiagen.com/us/shop/automated-solutions/sample-preparation/qiasymphony-spas-instruments/#orderinginformation
- InnuPure (Analytic Jena): https://www.analytik-jena.de/en/life-science/products/cat/cat/automated-extraction.html
- MagMAX (Thermo Fisher Scientific): https://www.thermofisher.com/order/catalog/product/4400074
- truXTRAC (Covaris): http://covarisinc.com/products/ffpe-extraction/
- MagNA Pure (Roche): https://lifescience.roche.com/en_de/brands/magnapure.html#magna-pure-reagents

Quantification of Nucleic Acids

For the different downstream applications, the quality and quantity of the extracted nucleic acids have to be assessed. There are different methods available for DNA or RNA quantification. Nucleic acids can be quantified by UV absorbance, fluorescent dye-based quantification, or quantitative PCR. Currently, there is no gold standard. UV absorbance measures directly all nucleic acids present in the sample. DNA and RNA absorb light at a wavelength of 260 nm, proteins at 280 nm, and organic solvents at 230 nm. The amount of absorbed light is proportional to the amount of DNA, RNA, protein, or organic solvent present. In addition, the purity of a sample can be determined by calculating the ratio of OD260/OD280 and should be between 1.8 and 2.2 and the ratio OD260/OD230, which should be above 1.8. Fluorescent dye-based quantification methods measure the concentration of nucleic acids by fluorescent binding dyes. Depending on whether RNA or DNA is measured, different dyes which bind specifically to either RNA or double-stranded DNA (dsDNA) are used. The quality of nucleic acids cannot be determined with this method. Using quantitative PCR the amount of amplifiable DNA or RNA present in a sample is measured based on a known quantity calculated in a standard curve. After each PCR cycle, the amount of DNA or RNA present is measured by fluorescent probes or dyes, which are incorporated into the nucleic acids. With this method quantity as well as quality can be assessed [221, 224, 229].

Conventional Mutation Analyses

Several methods are currently used to determine genomic variations, including pyrosequencing, Sanger sequencing, real-time PCR-based analysis, fragment length analysis, high-resolution melting (HRM), single-nucleotide probe extension assays (such as SNaPshot), reversed hybridization assays, or shifted termination assays (STA). All of the currently available methods require PCR amplification after DNA extraction (Fig. 27.2).

Each individual method has its advantages and disadvantages concerning factors such as mutation spectrum, costs, sensitivity, and time required for analysis. Pyrosequencing provides a sensitive method for detecting mutations with 5–10% allele frequency and further allows the detection of a wide variation of mutations. However, it may not be economical in all settings due to expensive equipment and reagents. Although Sanger sequencing remains the gold standard in many laboratories, it is time-consuming and relatively less sensitive because it is only able to detect mutations with an allele frequency above 15–20%. On the other hand, it is easy to implement and allows detection of previously unknown mutations.

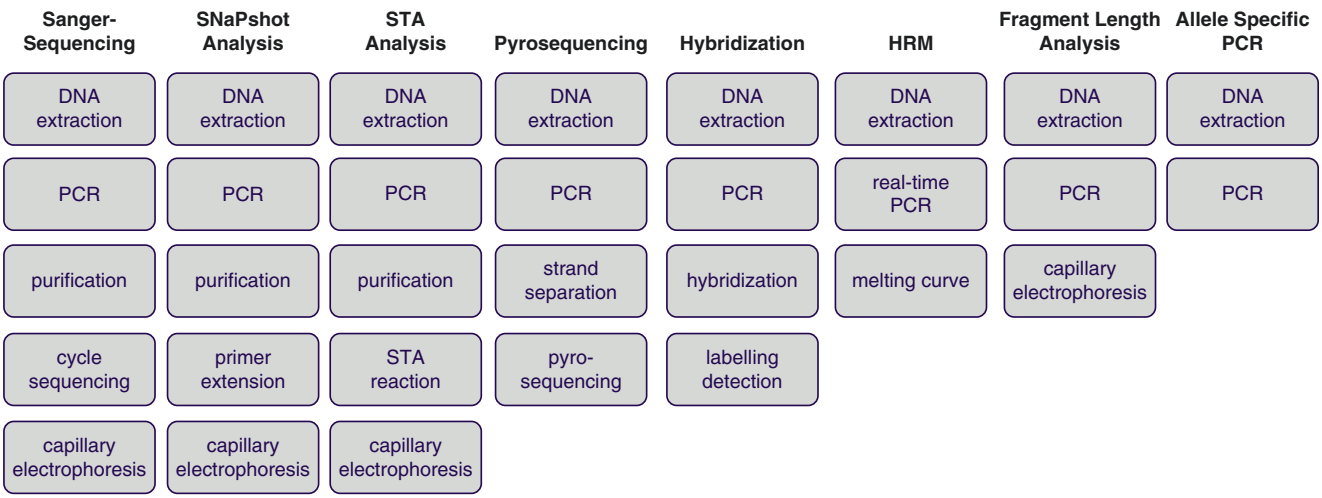


Fig. 27.2 Flowchart outlining the protocol steps for different techniques of mutation analysis. The methods included illustrate the example methods described in this chapter. *SnaPshot* single-nucleotide probe extension assay, *STA* shifted termination assay, *HRM* high-resolution melting

Several methods which are also commercially available are well suited for the detection of described mutations, e.g., real-time PCR-based analyses, single-nucleotide probe extension assays (SNaPshot), reversed hybridization assays, or shifted termination assays (STA). They provide high sensitivity and throughput but may fail to detect additional mutations around hot spot regions.

Two methods that allow sensitive screening without obtaining specific sequence information are high-resolution melting (HRM) and fragment length analysis. HRM can distinguish between wild-type and mutated sequences. It is therefore suited for questions where a high rate of wild type is expected because the wild-type cases may be rapidly excluded from further analyses. Fragment length analysis can only detect mutations changing the length of PCR product, e.g., deletions and insertions.

As several methods are commercially available, the following web-based resources can be used as a basis for laboratory protocols:

- https://www.qiagen.com/sg/resources/technologies/pyro-sequencing-resource-center/technology-overview/
- https://www.thermofisher.com/de/de/home/life-science/sequencing/dna-sequencing/snp-genotyping-variant-detection-sequencing/snp-genotyping-fragment-analysis.html
- https://molecular.roche.com/assays/cobas-egfr-mutation-test-v2-us-ivd/
- http://www.zytomed-systems.de/produkte/molpath/real-time-pcr-assays.html

The basic principles for Sanger sequencing, high-resolution melting, and fragment length analysis are outlined below.

Sanger Sequencing

The dideoxy sequencing method according to Sanger includes several analysis steps. First, when working with DNA from FFPE tissue, the relevant fragments of the DNA have to be amplified. Unbound primers and excess nucleotides have to be separated from the PCR fragments prior to cycle sequencing. This can be done by using exonuclease I and alkaline phosphatase. The cycle sequencing should always be performed as bidirectional sequencing and may be done with different sequencing reagents. As an example, the cycle sequencing protocol working with the BigDye Terminator Cycle Sequencing Ready Reaction Kit (ThermoFisher, Darmstadt, Germany) can be used.

The cycle sequencing products have to be separated from reaction components such as salt ions, dNTPs, and unincorporated dye terminators. The BigDye XTerminator Purification Kit (ThermoFisher, Darmstadt, Germany) provides a convenient purification tool. High-resolution capillary electrophoresis is run according to manufacturer's instructions.

High-Resolution Melting (HRM)

HRM is a molecular technique for high-throughput screening for mutations in a bounded region. Mutation determination using HRM is based on the dissociation of DNA, when exposed to an increasing temperature in the presence of fluorescent dyes intercalating in double-stranded DNA. The presence of a mutation leads to the formation of DNA heteroduplexes followed by a change in melting behavior [230]. Some types of mutations are difficult to detect with this method. Because HRM depends on heteroduplex formation, the rarely occurring hemizygous mutations where the wild-type allele is deleted due to chromosome loss might not be detected.

Protocols can be implemented on different available real-time PCR systems. Primer sequences, amplification conditions, and analysis parameters using the LightCycler 480 system (Roche Diagnostics, Mannheim, Germany) for the relevant exons of *KRAS*, *BRAF*, *PIK3CA*, and *AKT1* were described in detail by Ney et al. [230].

Fragment Length Analysis

Fragment length analysis is a suitable method for the detection of mutations which change the length of amplification products, e.g., insertions and deletions. The most common mutations in exon 19 of *EGFR* are deletions of 15 bp although others do occur. Point mutations occur very rarely, and thus, fragment length analysis is an appropriate method to screen for mutations in this exon.

Amplification of exon 19 is best performed by nested PCR with the reverse primer of the second round labeled with a fluorescent dye. Thereby, PCR products can be easily separated and detected by capillary electrophoresis. To determine the exact mutation status, PCR products of the first round can be analyzed by Sanger sequencing after purification. A detailed protocol is described by Molina-Vila et al. [231].

Massively Parallel Sequencing for Mutation Analysis

In order to investigate the different molecular parameters that might individually affect the therapy of lung cancer in a time- and cost-saving manner, multiplex approaches analyzing many genes of multiple patient samples simultaneously are needed. Thus, methods which allow high-throughput processing of multiple targets are ideally suited for a fast and detailed diagnosis of the mutational status of each lung cancer patient. New parallel sequencing technologies, also known as "next-generation sequencing (NGS)," provide many advantages in genetic tumor characterization such as high sensitivity in detection of somatic mutations, extremely high capacities allowing deep sequence analyses, and the high potential of sample and target multiplexing.

In recent years, targeted massively parallel sequencing approaches of cancer-related genes have become increasingly integrated into daily clinical practice. Already, many institutions use amplicon-based massively parallel sequencing approaches instead of conventional methods like Sanger sequencing, HRM, or fragment length analysis for the analysis of multiple genes [232, 233]. With this method, target regions are enriched by means of a multiplex PCR followed by library preparation and sequencing on a bench-top sequencer. Point mutations, small deletions, insertions, or duplications can be detected on a DNA level. The advantage is that only little amounts of DNA are needed, and even chemically modified and fragmented DNA derived from formalin-fixed and paraffin-embedded (FFPE) tissue can be used [221, 234]. However, on a DNA level, amplicon-based parallel sequencing does not routinely detect chromosomal aberrations such as gene fusions, and the detection of copy-number changes is difficult on FFPE material. Until now, these aberrations are widely analyzed by FISH [171, 235].

Thus, there is an increasing need to develop new technologies that can detect all therapeutically relevant genomic alterations including somatic mutations, rearrangements, and copy-number changes in a single assay to maximize the use of small lung cancer biopsies. Currently, new technologies and assays are being developed and published using either RNA or DNA or total nucleic acids extracted from FFPE material. On a DNA level, hybrid capture-based parallel sequencing enables the simultaneous detection of all genomic alterations. DNA is sheared, libraries are prepared, and target regions are enriched with target-specific biotinylated capture probes, following sequencing on a high-capacity bench-top sequencer [236–239]. The disadvantage of this method is that larger amounts of DNA are needed; thus, smaller samples cannot be analyzed. Further, the data analysis is highly complex, especially the detection of gene fusions (Fig. 27.1), and copy-number changes require customized data analysis pipelines as commercially available software still struggles with the detection of such aberrations [236–239].

On a RNA level, RNA-based parallel sequencing assays have been designed for the detection of known or unknown fusion events (Fig. 27.1). These assays use either RNA or total nucleic acids, and target regions are either enriched by means of a multiplex PCR or target-specific capture probes. Prepared libraries are also sequenced on a bench-top sequencer. Especially some of the multiplex PCR-based assays can be used with as little as 10 ng of RNA. However, as RNA-based parallel sequencing assays can only detect fusions and splicing events reliably, a DNA-based parallel sequencing approach has to be used in conjunction for the detection of somatic gene mutations and copy-number alterations. Moreover, the extraction of intact RNA form FFPE material can be problematic [90, 240].

Massively Parallel Sequencing Operation Workflow in Molecular Pathology Diagnostics

Independently from the specific parallel sequencing technology, first a template library has to be prepared. DNA or RNA templates can be targeted and enriched either by a target-specific multiplex PCR in an amplicon-based parallel sequencing assay or by hybridization of target-specific biotinylated capture probes in a hybrid capture-based parallel sequencing assay [90, 233, 237–239]. For the target

regions of interest, custom multiplex PCR primers or capture probes can be designed, or commercially available lung cancer panels can be used (see protocols a and b). Depending on the supplier, the capture probes are either DNA or RNA probes. Amplicon-based parallel sequencing assays most often use less than 40 ng of DNA or RNA, and hybrid capture-based parallel sequencing assays commonly use at least 50–200 ng of DNA or RNA [90, 233, 237–239].

At the beginning of each library preparation, first the starting material (DNA or RNA) has to be quantified. Here, either UV spectrophotometry, fluorescent dye-based quantification, or quantitative PCR can be used (see protocols c and d). When starting with RNA or total nucleic acids, the RNA has to be first transcribed into complementary DNA (cDNA). In an amplicon-based parallel sequencing assays, the cDNA or DNA of the target region of interest is then enriched by a multiplex PCR. For hybrid capture-based parallel sequencing assays, the DNA has to be fragmented by either mechanical shearing or enzymatic treatment into fragments of 150–200 base pairs. cDNA derived from FFPE material is already highly fragmented and can be used directly. For both parallel sequencing approaches, this initial step is followed by an end-repair step. Next, 3′A-overhangs are generated and sequencing adaptors are ligated. In amplicon-based approaches, adaptors often include an individual DNA sequence of eight to ten nucleotides. These individual DNA sequences are also called indexes or multiplex identifiers (MID) and allow the multiplexing of different patient samples. Thus, not only multiple targets but also up to 384 different tumor samples can be processed at the same time. Next, templates carrying the adapter sequences are selectively amplified by a low cycle adapter-specific PCR and then quantified. During library preparation, different DNA purification and fragment size selection steps are performed, which are often automated using magnetic bead-based procedures (see protocols e).

The library preparation of the amplicon-based sequencing approach is now finished. Quantified libraries are diluted, multiplexed, and sequenced on a bench-top sequencer. In the hybrid capture-based parallel sequencing approach, the target-specific biotinylated capture probes are now hybridized to target sequences of interest to allow for sequence enrichment using streptavidin beads. Libraries containing the target fragments are again selectively amplified by a low cycle adapter-specific PCR and then quantified (see protocols c and d). In the hybrid capture-based approach, the indexes or multiplex identifiers (MID) are incorporated during the first or second low cycle PCR depending on the assays used. Quantified libraries are diluted, multiplexed, and sequenced on a larger bench-top sequencer.

The library preparation workflow of a multiplex PCR-based or a hybrid capture-based parallel sequencing approach is summarized in Fig. 27.3.

Selected Protocols

(d) Multiplex PCR-based protocols (DNA- or RNA-based):
- Ion Ampliseq (Thermo Fisher Scientific): https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html
- GeneReader NGS System (Qiagen): https://www.qiagen.com/us/products/ngs/mdx-ngs-genereader/target-enrichment-and-targeted-panels/
- QIAseq NGS Solutions (Qiagen): https://www.qiagen.com/us/products/ngs/ngs-life-sciences/
- TruSeq Illumina: https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/amplicon-sequencing.html
- TruSight (Illumina): https://www.illumina.com/techniques/sequencing/rna-sequencing/targeted-rna-seq.html
- Multiplicom: http://www.multiplicom.com/products-onco
- Archerdx: http://archerdx.com/

(e) Hybrid capture-based protocols (DNA- or RNA-based):
- SureSelect (Agilent): http://www.genomics.agilent.com/en/SureSelect-DNA-Library-Preps/SureSelectXT-Reagent-Kits/?cid=AG-PT-177&tabId=AG-PR-1302
- Nimblegen (Roche): http://sequencing.roche.com/products/nimblegen-seqcap-target-enrichment.html
- Truseq: https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits.html
- NEB: https://www.neb.com/applications/library-preparation-for-next-generation-sequencing
- QIAseq NGS Solutions (Qiagen): https://www.qiagen.com/us/products/ngs/ngs-life-sciences/whole-genome-and-hybrid-capture/

(f) Fluorescent dye-based quantification
- Qubit Fluorometer (Thermo Fisher Scientific): https://www.thermofisher.com/us/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html
- Quantus Fluorometer (Promega): https://www.promega.de/products/fluorometers-luminometers-multimode-readers/fluorometers/quantus-fluorometer-simple-quantitation/
- PicoGreen quantification (Thermo Fisher Scientific): https://www.thermofisher.com/order/catalog/product/P11496

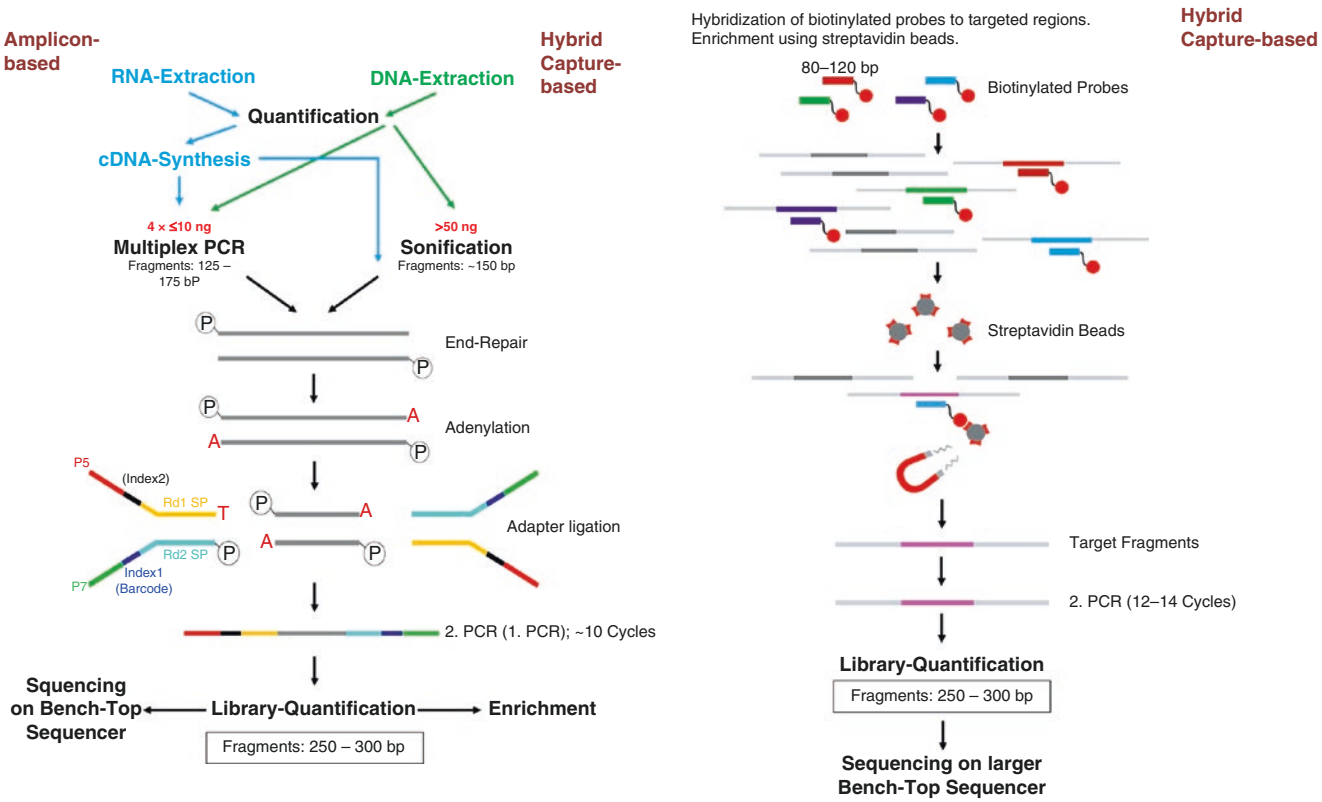


Fig. 27.3 Schematic library preparation workflow for DNA and RNA amplicon-based vs. hybrid capture-based parallel sequencing approaches. Nucleic acids are extracted. RNA is then transcribed into cDNA. In the amplicon-based approach, target regions are enriched by multiplex PCR followed by end repair, adenylation, adapter ligation, second PCR, library quantification, and sequencing on a bench-top sequencer. In the hybrid capture-based approach, prior to the target-specific enrichment, DNA is first fragmented by sonification. cDNA derived from FFPE material does not have to be fragmented. This step is then followed by an end repair, adenylation, adapter ligation, first PCR, and first library quantification step. After the first library quantification, target-specific biotinylated probes are hybridized to the target regions, and target-specific fragments are then captured and enriched by streptavidin beads. Target fragment are amplified by a second PCR; libraries are quantified and sequenced on a larger bench-top sequencer

(g) Quantitative PCR:
- GeneRead DNA QuantiMIZE Kit (Qiagen): https://www.qiagen.com/ch/shop/genes-and-pathways/technology-portals/browse-ngs/next-generation-sequencing/dna-isolation-and-qc/generead-dna-quantimize-kits/
- QIAseq Library Quant System (Qiagen): https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/qiaseq-library-quant-system/#orderinginformation
- KAPA hgDNA Quantification and QC Kit (KapaBiosystems): https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/hgdna-quantification/
- KAPA Library Quantification Kits: https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/library-quantification/
- TruSeq FFPE DNA Library Prep QC Kit (Illumina): https://support.illumina.com/sequencing/sequencing_kits/truseq-ffpe-dna-library-prep-qc-kit.html
- Agilent NGS FFPE QC Kit (Agilent): http://www.genomics.agilent.com/en/NGS-FFPE-QC-Kit-NEW-/NGS-FFPE-QC-Kit/?cid=cat2290001&tabId=prod2380001
- qPCR NGS Library Quantification Kit (Agilent): http://www.genomics.agilent.com/en/Specialty-Master-Mixes-Kits/qPCR-NGS-Library-Quantification-Kit/?cid=AG-PT-176&tabId=AG-PR-1157
- NEB Library Quantification (New England Biolabs): https://www.neb.com/products/sample-preparation-for-next-generation-sequencing/library-quantitation/library-quantitation

(h) Purification and size selection:
- Agencourt AMPure XP (Beckman Coulter): http://uk.beckman.com/nucleic-acid-sample-prep/purification-clean-up/pcr-purification
- SPRIselect Reagent (Beckman Coulter): http://uk.beckman.com/nucleic-acid-sample-prep/purification-clean-up/size-selection

Bench-Top Massively Parallel Sequencing Machines

There are different parallel sequencings platforms on the market. The different bench-top sequencers differ in their sequencing chemistry and their method of sequence detection. However, the underlying principal and steps are almost the same.

One of the first NGS methods was the 454 sequencing approach which was established by 454 Life Science Corporation [241]. Hereby, DNA was amplified by an emulsion PCR. Each emulsion droplet contained a single adapter-linked DNA template, hybridized to a primer-coated bead that then generates a clonal amplicon colony. Subsequent sequencing took place in a picoliter well plate where each well contained only one single bead with the amplified clone [241, 242]. The DNA clones were sequenced in parallel by pyrosequencing. Roche discontinued the production and support of the 454 sequencing machines.

The technology used for the ion semiconductor sequencing system (Table 27.2) is based on ultrasensitive measurements of pH changes. In order to detect the proton that is released during the nucleotide incorporation, sequencing of the clonal DNA templates is performed in picolitre cartridges of a MOSFET (metal-oxide-semiconductor field-effect transistor) flow cell [242, 246]. Prior to semiconductor sequencing, the template clones are also prepared by emulsion PCR as described for the 454 technology.

In contrast to these technologies, single DNA templates are immobilized on the sequencing flow cell of the Illumina clonal DNA clusters that are generated by bridge PCR amplification. Sequencing is performed by DNA strand synthesis using fluorochrome-labeled nucleotides and nucleotide coupling and decapping steps proceeding in tandem [242, 247] (Table 27.2).

In the new Qiagen system, the template clones are also prepared by emulsion PCR as described for the 454 technology. However, the sequencing is then performed on a sequencing flow cell by DNA strand synthesis similar to the Illumina technology [242, 243].

The establishment of bench-top sequencer enabled the adaption of the different technologies in diagnostic laboratories [248–250]. The various NGS systems have different advantages and disadvantages, and molecular pathology laboratories have to decide which method represents best their demands on the diagnostic spectrum, the diagnostic variety, or throughput. Table 27.2 summarizes the essential features of currently available NGS approaches with a critical evaluation of their applicability to the routine diagnostic program in molecular pathology.

Table 27.2 Summary of NGS bench-top sequencers [242–245]

	Illumina	Thermo Fisher Scientific	Qiagen
Principle	Sequencing-by-synthesis	Semiconductor sequencing	Sequencing-by-synthesis
Signal detection	Two- or four-color fluorescence	Measuring of pH value	Four-color fluorescence
Capacity	NextSeq: 16.25–120 Gb[a]	Proton: 10 Gb	N.A.
	MiSeq: 1.2 Mb–15 Gb[a] MiniSeq: 1.65–7.5 Gb[a]	PGM: 30 Mb–2 Gb[a] Ion S5: 0.6–15 Gb[a] Ion S5 XL: 0.6–15 Gb[a]	
Running time	NextSeq: 11–29 h[a]	Proton: 2–4 h	N.A.
	MiSeq: 4–56 h[a] MiniSeq: 17–24 h[a]	PGM: 2.3–7.3 h[a] Ion S5: 2.5–4 h[a] Ion S5 XL: 2.5–4 h[a]	
Generation templates	Bridge PCR	Emulsion PCR	Emulsion PCR
Read length	25–300 bases	200–400 bases	N.A.
Cost per base	Low	Low	Low
Advantages	Suited for high throughput High flexibility in throughput Low costs High flexible in sequencing other companies libraries	Suited for high throughput High flexibility in throughput Low costs Short running time High flexible in sequencing other companies libraries	Flexibility in throughput Low costs
Disadvantages	Reading errors in GC-rich regions Long running time	High number of reading errors in homopolymer stretches	Long running times Only GeneReader panels can be used Reading errors in GC-rich regions

b bases, *h* hours
N.A.: This information is not available
[a]Depending on read length

Quality Assurance

Although NGS opened up tremendous new perspectives for time, cost, and DNA or RNA material saving diagnostics, novel laboratory requirements and quality controls still have to be addressed. Further, the adaption of protocols from the various suppliers might be necessary for FFPE material and the laboratory needs. In the field of NGS, quality controls are not yet defined. According to our experience, the improvement of pre-analytical steps is also very important. Especially, protocols for DNA and RNA extraction have to be optimized to achieve the highest concentration and quality possible [221].

As for all other PCR-based methods, contaminations from one samples or run to the next may occur. During NGS processing amplicon, clones are handled, and the high NGS sensitivity can result in the detection of amplicons remaining from the preceding run. Therefore, careful sample handling as well as cleaning of working places and equipment should be performed. In addition to the conventional control reactions, the following controls are established in our diagnostic laboratory to ensure the detection of contaminations:

- Change of the barcode sets used from run to run.
- Negative controls (no RNA/DNA) run alongside samples through library preparation and sequencing.
- Separation of pre- and post-PCR working areas.
- Pipetting robots are used to avoid human error and contaminations.

Additionally, experience in data interpretation becomes more and more important. Experienced bioinformations are needed, and the field is in need of improvements of commercially available data analysis software. Additionally, threshold settings have to be gathered, and the issue of short- and long-term data storage has to be solved [249].

NanoString nCounter Technology

Another widely used technology for the detection of fusions with known and unknown fusion partners on a RNA level and for the detection of copy-number changes and known somatic mutations on a DNA level is the NanoString nCounter technology. The detection of unknown mutations in a target region is currently not possible. The NanoString nCounter technology multiplexes mixtures of target-specific capture and reporter probes with fluorescently labeled color tags that hybridize to the RNA sequence of the target gene. The reporter tags are directly digitally imaged, counted, and quantified. The number of transcripts present equals the number of reporter tags. Known fusions are detected by junction probes (Figs. 27.1 and 27.4), and unknown fusions are detected by 5′/3′ positional imbalanced score. 5′/3′ expression that diverges from 1 indicating that a fusion event has occurred (Fig. 27.4).

The NanoString nCounter technology does not require an amplification step, thus avoiding PCR artifacts. The disadvantage of the NanoString nCounter technology is that this method currently requires larger amounts of RNA and DNA. NanoString uses their own nCounter system for the read out [88, 91, 242, 251–253].

Fluorescence In Situ Hybridization (FISH)

The method of fluorescence in situ hybridization (FISH) can be used for the detection of chromosomal translocations/inversions or amplifications. The labeled probe is directly hybridized onto the tissue slide.

To detect amplifications, two differently labeled probes are usually used: The locus-specific probe hybridizes to the gene of interest, and a centromere-specific probe binds to the corresponding chromosome. For the detection of a translocation, most commonly the so-called break-apart assays are used: Two differently labeled probes hybridize adjacent to each other to one of the two translocation partners. Currently, the detection of *HER2* and *MET* amplification and *ALK*, *ROS1*, and *RET* translocations or inversions is most commonly done using FISH analysis.

Pretreatment of slides may be performed with the semiautomated VP2000 processor system (Abbott Molecular, Wiesbaden, Germany) or manually using pretreatment reagents from several suppliers. Probes for the different applications are also available from multiple manufacturers.

Assay design and evaluation is described here using the example of *ALK* rearrangement. Chromosomal rearrangements with *ALK* comprise the inversion on chromosome 2 leading to a fusion with *EML4* and the translocation affecting *ALK*, but not *EML4*, such as *ALK-TGF* or *ALK-KIF5B*. Therefore, the use of a break-apart assay is recommended.

A probe system designed to discriminate between *ALK* inversions and translocations is the Zyto*Light*® SPEC ALK/EML4 TriCheck™ (ZytoVision, Bremerhaven, Germany). This assay consists of three differently labeled probes, where two probes hybridize distal and proximal to the *ALK* gene breakpoint region, respectively, and the third probe binds to the *EML4* gene. In an interphase nucleus of a normal cell, two orange/green fusion signals and two blue signals are expected. The *EML4-ALK* inversion is indicated by one separate green signal, one separate orange signal, and an additional blue signal. The separate green and orange signals each co-localize with a blue signal. During inversion, the 5′

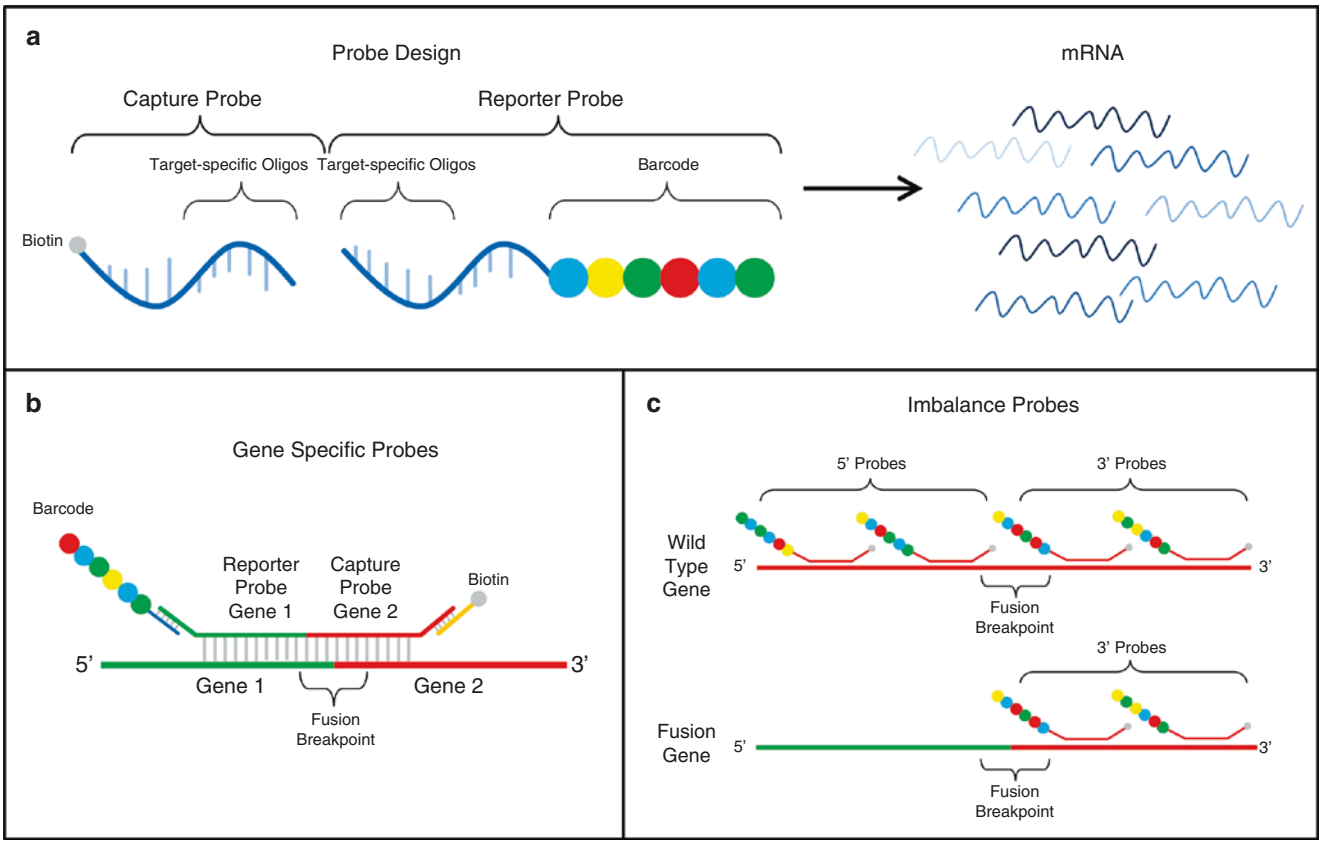


Fig. 27.4 Probe design of the NanoString nCounter technology for the detection of fusions with known and unknown fusion partners. (**a**) Schematic of target-specific capture and reporter probes, which bind to mRNA. The reporter probe includes a colored barcode on the 5′ end, and the capture probe carries a biotin on the 3′ end. (**b**) Gene-specific probes span a fusion breakpoint to detect known fusion events. (3) Imbalance probes compare the expression upstream and downstream of the fusion breakpoint for the detection of unknown fusions. A 5′/3′ expression ratio which diverges from 1 is an indication for a fusion event

part of *ALK* can be deleted, so the separate green signal is lost (Fig. 27.1). A signal pattern consisting of one orange/green fusion signal, one orange signal, and a separate green signal as well as two blue signals indicates an *ALK* translocation without involvement of *EML4*.

Break-apart signals have to be detected in at least 15% of nuclei. One hundred of nuclei are counted to detect the rearrangement. Only nuclei with non-overlapping signals and with the expected number of signals are evaluated. To count signals as two, they have to be separated by at least two signal diameters.

The criteria for *ROS1* FISH interpretation in NSCLC are similar to those created for *ALK* rearrangement with two main patterns: first, the break-apart pattern ("classic" pattern) with one fusion signal and two separated 3′ and 5′ signals and, second, an atypical pattern consisting of an isolated 3′ signal pattern, usually one fusion signal and one isolated 3′ green signal without the corresponding 5′ signal. The green fluorochrome is the probe containing the kinase domain of the *ROS1* gene. The cut-off for positivity is set to 20% of nuclei for *ROS1* FISH.

Conclusion

In the last decade, lung cancer has become the prime example for the success of personalized therapy. Four druggable genomic lesions in *EGFR*, *ALK*, *ROS1*, *BRAF*, and several immunomodulatory antibodies have been approved for first- and second-line therapies. Patients selected by appropriate biomarkers for these therapies show a significant benefit in overall survival and quality of life compared to standard chemotherapy. Also, selective inhibitors overcoming acquired resistance have been introduced into clinical practice requiring molecular reevaluation of relapsed tumors. In this setting, monitoring mutational patterns in peripheral blood ("liquid biopsies") will probably come into focus. Similarly, in squamous cell carcinomas, genetic alterations affecting kinases have been found and are currently evaluated clinically, e.g., *FGFR1* amplifications and *MET* mutations.

A major challenge now is to implement high-quality molecular diagnostics and personalized treatment strategies in routine clinical practice. With conventional molecular methods represented by sequencing technologies, only sequential

analysis of a small number of analytes with low sensitivity is possible. Hence, the development of novel methods that provide sensitive, accurate, and simultaneous detection of the mutation status of many samples and gene loci has been of major interest. Massive parallel sequencing by NGS approaches and analysis of very comprehensive gene sets has replaced conventional sequencing technologies.

References

1. Seidel DZT, Heukamp LC, Peifer M, Bos M, Fernández-Cuesta L, Leenders F, Lu X, Ansén S, Gardizi M, Nguyen C, Berg J, Russell P, Wainer Z, Schildhaus HU, Rogers TM, Solomon B, Pao W, Carter SL, Getz G, Hayes D, Wilkerson MD, Thunnissen E, Travis WD, Perner S, Wright G, Brambilla E, Büttner R, Wolf J, Thomas RK, Gabler F, Wilkening I, Müller C, Dahmen I, Menon R, König K, Albus K, Merkelbach-Bruse S, Fassunke J, Schmitz K, Kuenstlinger H, Kleine MA, Binot E, Querings S, Altmüller J, Bäßmann I, Nürnberg P, Schneider PM, Bogus M, Büttner R, Perner S, Russell P, Thunnissen E, Travis WD, Brambilla E, Soltermann A, Moch H, Brustugun OT, Solberg S, Lund-Iversen M, Helland Å, Muley T, Hoffmann H, Schnabel PA, Chen Y, Groen H, Timens W, Sietsma H, Clement JH, Weder W, Sänger J, Stoelben E, Ludwig C, Engel-Riedel W, Smit E, Heideman DA, Snijders PJ, Nogova L, Sos ML, Mattonet C, Töpelt K, Scheffler M, Goekkurt E, Kappes R, Krüger S, Kambartel K, Behringer D, Schulte W, Galetke W, Randerath W, Heldwein M, Schlesinger A, Serke M, Hekmat K, Frank KF, Schnell R, Reiser M, Hünerlitürkoglu AN, Schmitz S, Meffert L, Ko YD, Litt-Lampe M, Gerigk U, Fricke R, Besse B, Brambilla C, Lantuejoul S, Lorimier P, Moro-Sibilot D, Cappuzzo F, Ligorio C, Damiani S, Field JK, Hyde R, Validire P, Girard P, Muscarella LA, Fazio VM, Hallek M, Soria JC, Carter SL, Getz G, Hayes D, Wilkerson MD, Achter V, Lang U, Seidel D, Zander T, Heukamp LC, Peifer M, Bos M, Pao W, Travis WD, Brambilla E, Büttner R, Wolf J, Thomas RK, Büttner R, Wolf J, Thomas RK, CLCGP NGM. A genomics-based classification of human lung tumors. Sci Transl Med. 2013;5(209):209ra153. https://doi.org/10.1126/scitranslmed.3006802.
2. Buettner R, Wolf J, Thomas RK. Lessons learned from lung cancer genomics: the emerging concept of individualized diagnostics and treatment. J Clin Oncol. 2013;31(15):1858–65. https://doi.org/10.1200/jco.2012.45.9867.
3. Roskoski R Jr. The ErbB/HER receptor protein-tyrosine kinases and cancer. Biochem Biophys Res Commun. 2004;319(1):1–11. https://doi.org/10.1016/j.bbrc.2004.04.150.
4. Akca H, Tani M, Hishida T, Matsumoto S, Yokota J. Activation of the AKT and STAT3 pathways and prolonged survival by a mutant EGFR in human lung cancer cells. Lung Cancer. 2006;54(1):25–33. https://doi.org/10.1016/j.lungcan.2006.06.007.
5. Onn A, Correa AM, Gilcrease M, Isobe T, Massarelli E, Bucana CD, O'Reilly MS, Hong WK, Fidler IJ, Putnam JB, Herbst RS. Synchronous overexpression of epidermal growth factor receptor and HER2-neu protein is a predictor of poor outcome in patients with stage I non-small cell lung cancer. Clin Cancer Res. 2004;10(1 Pt 1):136–43.
6. Harari PM. Epidermal growth factor receptor inhibition strategies in oncology. Endocr Relat Cancer. 2004;11(4):689–708. https://doi.org/10.1677/erc.1.00600.
7. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129–39. https://doi.org/10.1056/NEJMoa040938.
8. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500. https://doi.org/10.1126/science.1099314.
9. Zhang X, Gureasko J, Shen K, PA Cole JK. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell. 2006;125(6):1137–49. https://doi.org/10.1016/j.cell.2006.05.013.
10. Carey KD, Garton AJ, Romero MS, Kahler J, Thomson S, Ross S, Park F, JD Haley NG, Sliwkowski MX. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. Cancer Res. 2006;66(16):8163–71. https://doi.org/10.1158/0008-5472.can-06-0453.
11. Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, Seto T, Satouchi M, Tada H, Hirashima T, Asami K, Katakami N, Takada M, Yoshioka H, Shibata K, Kudoh S, Shimizu E, Saito H, Toyooka S, Nakagawa K, Fukuoka M. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol. 2010;11(2):121–8. https://doi.org/10.1016/s1470-2045(09)70364-x.
12. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Munoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M, Paz-Ares L. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012;13(3):239–46. https://doi.org/10.1016/s1470-2045(11)70393-x.
13. Novello S, Barlesi F, Califano R, Cufer T, Ekman S, MG Levra K, Kerr S, Popat M, Reck SS, Simo GV, Vansteenkiste J, Peters S. Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2016;27(suppl 5):v1–v27. https://doi.org/10.1093/annonc/mdw326.
14. Tan DS, Yom SS, Tsao MS, Pass HI, Kelly K, Peled N, Yung RC, Wistuba II, Yatabe Y, Unger M, Mack PC, Wynes MW, Mitsudomi T, Weder W, Yankelevitz D, Herbst RS, Gandara DR, Carbone DP, Bunn PA Jr, Mok TS, Hirsch FR. The International Association for the Study of Lung Cancer Consensus Statement on optimizing management of EGFR mutation-positive non-small cell lung cancer: status in 2016. J Thorac Oncol. 2016;11(7):946–63. https://doi.org/10.1016/j.jtho.2016.05.008.
15. Brevet M, Arcila M, Ladanyi M. Assessment of EGFR mutation status in lung adenocarcinoma by immunohistochemistry using antibodies specific to the two major forms of mutant EGFR. J Mol Diagn. 2010;12(2):169–76. https://doi.org/10.2353/jmoldx.2010.090140.
16. Huang S-F, Liu H-P, Li L-H, Ku Y-C, Fu Y-N, Tsai H-Y, Chen Y-T, Lin Y-F, Chang W-C, Kuo H-P, Wu Y-C, Chen Y-R, Tsai S-F. High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib

responsiveness in Taiwan. Clin Cancer Res. 2004;10(24):8195–203. https://doi.org/10.1158/1078-0432.ccr-04-1245.
17. Arcila ME, Nafa K, Chaft JE, Rekhtman N, Lau C, Reva BA, Zakowski MF, Kris MG, Ladanyi M. EGFR Exon 20 Insertion Mutations in Lung Adenocarcinomas: Prevalence, Molecular Heterogeneity, and Clinicopathologic Characteristics. Mol Cancer Ther. 2013;12(2):220–9. https://doi.org/10.1158/1535-7163.mct-12-0620.
18. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777–D83. https://doi.org/10.1093/nar/gkw1121.
19. He M, Capelletti M, Nafa K, Yun CH, Arcila ME, Miller VA, Ginsberg MS, Zhao B, Kris MG, Eck MJ, Janne PA, Ladanyi M, Oxnard GR. EGFR exon 19 insertions: a new family of sensitizing EGFR mutations in lung adenocarcinoma. Clin Cancer Res. 2012;18(6):1790–7. https://doi.org/10.1158/1078-0432.ccr-11-2361.
20. Lin YT, Liu YN, Wu SG, Yang JC, Shih JY. Epidermal growth factor receptor tyrosine kinase inhibitor-sensitive exon 19 insertion and exon 20 insertion in patients with advanced non-small-cell lung cancer. Clin Lung Cancer. 2017;18(3):324–32 e1. https://doi.org/10.1016/j.cllc.2016.12.014.
21. Ackerman A, Goldstein MA, Kobayashi S, Costa DB. EGFR delE709_T710insD: a rare but potentially EGFR inhibitor responsive mutation in non-small-cell lung cancer. J Thorac Oncol. 2012;7(10):e19–20. https://doi.org/10.1097/JTO.0b013e3182635ab4.
22. Kobayashi Y, Togashi Y, Yatabe Y, Mizuuchi H, Jangchul P, Kondo C, Shimoji M, Sato K, Suda K, Tomizawa K, Takemoto T, Hida T, Nishio K, Mitsudomi T. EGFR exon 18 mutations in lung cancer: molecular predictors of augmented sensitivity to afatinib or neratinib as compared with first- or third-generation TKIs. Clin Cancer Res. 2015;21(23):5305–13. https://doi.org/10.1158/1078-0432.ccr-15-1046.
23. Yang JC, Sequist LV, Geater SL, Tsai CM, Mok TS, Schuler M, N Yamamoto CJY, Ou SH, Zhou C, Massey D, V Zazulina YLW. Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6. Lancet Oncol. 2015;16(7):830–8. https://doi.org/10.1016/s1470-2045(15)00026-1.
24. Baik CS, D W, Smith C, Martins RG, Pritchard CC. Durable response to tyrosine kinase inhibitor therapy in a lung cancer patient harboring epidermal growth factor receptor tandem kinase domain duplication. J Thorac Oncol. 2015;10(10):e97–9. https://doi.org/10.1097/jto.0000000000000586.
25. Gallant JN, Sheehan JH, Shaver TM, Bailey M, Lipson D, Chandramohan R, Red Brewer M, York SJ, Kris MG, Pietenpol JA, Ladanyi M, Miller VA, Ali SM, Meiler J, Lovly CM. EGFR kinase domain duplication (EGFR-KDD) is a novel oncogenic driver in lung cancer that is clinically responsive to afatinib. Cancer Discov. 2015;5(11):1155–63. https://doi.org/10.1158/2159-8290.cd-15-0654.
26. Konduri K, Gallant JN, Chae YK, Giles FJ, Gitlitz BJ, Gowen K, Ichihara E, Owonikoko TK, Peddareddigari V, Ramalingam SS, Reddy SK, Eaby-Sandy B, Vavala T, Whiteley A, Chen H, Yan Y, Sheehan JH, Meiler J, Morosini D, Ross JS, Stephens PJ, Miller VA, Ali SM, Lovly CM. EGFR fusions as novel therapeutic targets in lung cancer. Cancer Discov. 2016;6(6):601–11. https://doi.org/10.1158/2159-8290.cd-16-0075.
27. Yasuda H, Kobayashi S, Costa DB. EGFR exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. Lancet Oncol. 2012;13(1):e23–31. https://doi.org/10.1016/s1470-2045(11)70129-2.
28. Yasuda H, Park E, Yun CH, Sng NJ, Lucena-Araujo AR, Yeo WL, Huberman MS, Cohen DW, Nakayama S, Ishioka K, Yamaguchi N, Hanna M, Oxnard GR, Lathan CS, Moran T, Sequist LV, Chaft JE, Riely GJ, Arcila ME, Soo RA, Meyerson M, Eck MJ, Kobayashi SS, Costa DB. Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. Sci Transl Med. 2013;5(216):216ra177. https://doi.org/10.1126/scitranslmed.3007205.
29. Jia Y, Juarez J, Li J, Manuia M, Niederst MJ, Tompkins C, Timple N, Vaillancourt MT, Pferdekamper AC, Lockerman EL, Li C, Anderson J, Costa C, Liao D, Murphy E, DiDonato M, Bursulaya B, Lelais G, Barretina J, McNeill M, Epple R, Marsilje TH, Pathan N, Engelman JA, Michellys PY, McNamara P, Harris J, Bender S, Kasibhatla S. EGF816 exerts anticancer effects in non-small cell lung cancer by irreversibly and selectively targeting primary and acquired activating mutations in the EGF receptor. Cancer Res. 2016;76(6):1591–602. https://doi.org/10.1158/0008-5472.can-15-2581.
30. Leventakos K, Kipp BR, Rumilla KM, Winters JL, Yi ES, Mansfield AS. S768I mutation in EGFR in patients with lung cancer. J Thorac Oncol. 2016;11(10):1798–801. https://doi.org/10.1016/j.jtho.2016.05.007.
31. Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med. 2005;352(8):786–92. https://doi.org/10.1056/NEJMoa044238.
32. Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF, Kris MG, Varmus H. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med. 2005;2(3):e73. https://doi.org/10.1371/journal.pmed.0020073.
33. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A. 2008;105(6):2070–5. https://doi.org/10.1073/pnas.0709662105.
34. Cross DA, Ashton SE, Ghiorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, Orme JP, Finlay MR, Ward RA, Mellor MJ, Hughes G, Rahi A, Jacobs VN, Red Brewer M, Ichihara E, Sun J, Jin H, Ballard P, Al-Kadhimi K, Rowlinson R, Klinowska T, Richmond GH, Cantarini M, Kim DW, Ranson MR, Pao W. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. Cancer Discov. 2014;4(9):1046–61. https://doi.org/10.1158/2159-8290.cd-14-0337.
35. Janne PA, Yang JC, Kim DW, Planchard D, Ohe Y, Ramalingam SS, Ahn MJ, Kim SW, Su WC, Horn L, Haggstrom D, Felip E, Kim JH, Frewer P, Cantarini M, Brown KH, Dickinson PA, Ghiorghiu S, Ranson M. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med. 2015;372(18):1689–99. https://doi.org/10.1056/NEJMoa1411817.
36. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, Bergethon K, Shaw AT, Gettinger S, Cosper AK, Akhavanfard S, Heist RS, Temel J, Christensen JG, Wain JC, Lynch TJ, Vernovsky K, Mark EJ, Lanuti M, Iafrate AJ, Mino-Kenudson M, Engelman JA. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med. 2011;3(75):75ra26. https://doi.org/10.1126/scitranslmed.3002003.
37. Takezawa K, Pirazzoli V, Arcila ME, Nebhan CA, Song X, de Stanchina E, Ohashi K, Janjigian YY, Spitzler PJ, Melnick MA, Riely GJ, Kris MG, Miller VA, Ladanyi M, Politi K, Pao W. HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. Cancer Discov. 2012;2(10):922–33. https://doi.org/10.1158/2159-8290.cd-12-0108.

38. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC, Janne PA. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science. 2007;316(5827):1039–43. https://doi.org/10.1126/science.1141478.
39. Bean J, Brennan C, Shih JY, Riely G, Viale A, Wang L, Chitale D, Motoi N, Szoke J, Broderick S, Balak M, Chang WC, Yu CJ, Gazdar A, Pass H, Rusch V, Gerald W, Huang SF, Yang PC, Miller V, Ladanyi M, Yang CH, Pao W. MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. Proc Natl Acad Sci U S A. 2007;104(52):20932–7. https://doi.org/10.1073/pnas.0710370104.
40. Gou LY, Li AN, Yang JJ, Zhang XC, Su J, Yan HH, Xie Z, Lou NN, Liu SY, Dong ZY, Gao HF, Zhou Q, Zhong WZ, Xu CR, Wu YL. The coexistence of MET over-expression and an EGFR T790M mutation is related to acquired resistance to EGFR tyrosine kinase inhibitors in advanced non-small cell lung cancer. Oncotarget. 2016;7(32):51311–9. https://doi.org/10.18632/oncotarget.9697.
41. Gatzemeier U, Groth G, Butts C, Van Zandwijk N, Shepherd F, Ardizzoni A, Barton C, Ghahramani P, Hirsh V. Randomized phase II trial of gemcitabine-cisplatin with or without trastuzumab in HER2-positive non-small-cell lung cancer. Ann Oncol. 2004;15(1):19–27.
42. Ross HJ, Blumenschein GR Jr, Aisner J, Damjanov N, Dowlati A, Garst J, Rigas JR, Smylie M, Hassani H, Allen KE, Leopold L, Zaks TZ, Shepherd FA. Randomized phase II multicenter trial of two schedules of lapatinib as first- or second-line monotherapy in patients with advanced or metastatic non-small cell lung cancer. Clin Cancer Res. 2010;16(6):1938–49. https://doi.org/10.1158/1078-0432.ccr-08-3328.
43. Ohashi K, Sequist LV, Arcila ME, Moran T, Chmielecki J, Lin YL, Pan Y, Wang L, de Stanchina E, Shien K, Aoe K, Toyooka S, Kiura K, Fernandez-Cuesta L, Fidias P, Yang JC, Miller VA, Riely GJ, Kris MG, Engelman JA, Vnencak-Jones CL, Dias-Santagata D, Ladanyi M, Pao W. Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. Proc Natl Acad Sci U S A. 2012;109(31):E2127–33. https://doi.org/10.1073/pnas.1203530109.
44. Wang J, Wang B, Chu H, Yao Y. Intrinsic resistance to EGFR tyrosine kinase inhibitors in advanced non-small-cell lung cancer with activating EGFR mutations. Onco Targets Ther. 2016;9:3711–26. https://doi.org/10.2147/ott.s106399.
45. Ham JS, Kim S, Kim HK, Byeon S, Sun JM, Lee SH, Ahn JS, Park K, Choi YL, Han J, Park W, Ahn MJ. Two cases of small cell lung cancer transformation from EGFR mutant adenocarcinoma during AZD9291 treatment. J Thorac Oncol. 2016;11(1):e1–4. https://doi.org/10.1016/j.jtho.2015.09.013.
46. Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, Kurtz DM, Stehr H, Scherer F, Karlovich CA, Harding TC, Durkin KA, Otterson GA, Purcell WT, Camidge DR, Goldman JW, Sequist LV, Piotrowska Z, Wakelee HA, Neal JW, Alizadeh AA, Diehn M. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun. 2016;7:11815. https://doi.org/10.1038/ncomms11815.
47. Ortiz-Cuaran S, Scheffler M, Plenker D, Dahmen L, Scheel AH, Fernandez-Cuesta L, Meder L, Lovly CM, Persigehl T, Merkelbach-Bruse S, Bos M, Michels S, Fischer R, Albus K, Konig K, Schildhaus HU, Fassunke J, Ihle MA, Pasternack H, Heydt C, Becker C, Altmuller J, Ji H, Muller C, Florin A, Heuckmann JM, Nuernberg P, Ansen S, Heukamp LC, Berg J, Pao W, Peifer M, Buettner R, Wolf J, Thomas RK, Sos ML. Heterogeneous mechanisms of primary and acquired resistance to third-generation EGFR inhibitors. Clin Cancer Res. 2016;22(19):4837–47. https://doi.org/10.1158/1078-0432.ccr-15-1915.
48. Planchard D, Loriot Y, Andre F, Gobert A, Auger N, Lacroix L, Soria JC. EGFR-independent mechanisms of acquired resistance to AZD9291 in EGFR T790M-positive NSCLC patients. Ann Oncol. 2015;26(10):2073–8. https://doi.org/10.1093/annonc/mdv319.
49. Nanjo S, Yamada T, Nishihara H, Takeuchi S, Sano T, Nakagawa T, Ishikawa D, Zhao L, Ebi H, Yasumoto K, Matsumoto K, Yano S. Ability of the Met kinase inhibitor crizotinib and new generation EGFR inhibitors to overcome resistance to EGFR inhibitors. PLoS One. 2013;8(12):e84700. https://doi.org/10.1371/journal.pone.0084700.
50. Normanno N, Denis MG, Thress KS, Ratcliffe M, Reck M. Guide to detecting epidermal growth factor receptor (EGFR) mutations in ctDNA of patients with advanced non-small-cell lung cancer. Oncotarget. 2017;8(7):12501–16. https://doi.org/10.18632/oncotarget.13915.
51. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, Yang JC, Barrett JC, Janne PA. Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. J Clin Oncol. 2016;34(28):3375–82. https://doi.org/10.1200/jco.2016.66.7162.
52. Reck M, Hagiwara K, Han B, Tjulandin S, Grohe C, Yokoi T, Morabito A, Novello S, Arriola E, Molinier O, McCormack R, Ratcliffe M, Normanno N. ctDNA determination of EGFR mutation status in European and Japanese patients with advanced NSCLC: the ASSESS study. J Thorac Oncol. 2016;11(10):1682–9. https://doi.org/10.1016/j.jtho.2016.05.036.
53. Leicht DT, Balan V, Kaplun A, Singh-Gupta V, Kaplun L, Dobson M, Tzivion G. Raf kinases: function, regulation and role in human cancer. Biochim Biophys Acta. 2007;1773(8):1196–212. https://doi.org/10.1016/j.bbamcr.2007.05.001.
54. Marchetti A, Felicioni L, Malatesta S, Grazia Sciarrotta M, Guetti L, Chella A, Viola P, Pullara C, Mucilli F, Buttitta F. Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. J Clin Oncol. 2011;29(26):3574–9. https://doi.org/10.1200/jco.2011.35.9638.
55. Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, Davies H, Cox C, Brignell G, Stephens P, Futreal PA, Wooster R, Stratton MR, Weber BL. BRAF and RAS mutations in human lung cancer and melanoma. Cancer Res. 2002;62(23):6997–7000.
56. Naoki K, Chen TH, Richards WG, Sugarbaker DJ, Meyerson M. Missense mutations of the BRAF gene in human lung adenocarcinoma. Cancer Res. 2002;62(23):7001–3.
57. Tissot C, Couraud S, Tanguy R, Bringuier PP, Girard N, Souquet PJ. Clinical characteristics and outcome of patients with lung cancer harboring BRAF mutations. Lung Cancer. 2016;91:23–8. https://doi.org/10.1016/j.lungcan.2015.11.006.
58. Ross JS, Wang K, Chmielecki J, Gay L, Johnson A, Chudnovsky J, Yelensky R, Lipson D, Ali SM, Elvin JA, Vergilio JA, Roels S, Miller VA, Nakamura BN, Gray A, Wong MK, Stephens PJ. The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. Int J Cancer. 2016;138(4):881–90. https://doi.org/10.1002/ijc.29825.
59. Jang JS, Lee A, Li J, Liyanage H, Yang Y, Guo L, Asmann YW, Li PW, Erickson-Johnson M, Sakai Y, Sun Z, Jeon HS, Hwang H, Bungum AO, Edell ES, Simon VA, Kopp KJ, Eckloff B, Oliveira AM, Wieben E, Aubry MC, Yi E, Wigle D, Diasio RB, Yang P, Jen J. Common oncogene mutations and novel SND1-BRAF transcript fusion in lung adenocarcinoma from never smokers. Sci Rep. 2015;5:9755. https://doi.org/10.1038/srep09755.

60. Peters S, Michielin O, Zimmermann S. Dramatic response induced by vemurafenib in a BRAF V600E-mutated lung adenocarcinoma. J Clin Oncol. 2013;31(20):e341–4. https://doi.org/10.1200/jco.2012.47.6143.
61. Robinson SD, O'Shaughnessy JA, Cowey CL, Konduri K. BRAF V600E-mutated lung adenocarcinoma with metastases to the brain responding to treatment with vemurafenib. Lung Cancer. 2014;85(2):326–30. https://doi.org/10.1016/j.lungcan.2014.05.009.
62. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, Wolf J, Raje NS, Diamond EL, Hollebecque A, Gervais R, Elez-Fernandez ME, Italiano A, Hofheinz RD, Hidalgo M, Chan E, Schuler M, Lasserre SF, Makrutzki M, Sirzen F, Veronese ML, Tabernero J, Baselga J. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. N Engl J Med. 2015;373(8):726–36. https://doi.org/10.1056/NEJMoa1502309.
63. Planchard D, Kim TM, Mazieres J, Quoix E, Riely G, Barlesi F, Souquet PJ, Smit EF, Groen HJ, Kelly RJ, Cho BC, Socinski MA, Pandite L, Nase C, Ma B, D'Amelio A Jr, Mookerjee B, Curtis CM Jr, Johnson BE. Dabrafenib in patients with BRAF(V600E)-positive advanced non-small-cell lung cancer: a single-arm, multicentre, open-label, phase 2 trial. Lancet Oncol. 2016;17(5):642–50. https://doi.org/10.1016/s1470-2045(16)00077-2.
64. Gautschi O, Milia J, Cabarrou B, Bluthgen MV, Besse B, Smit EF, Wolf J, Peters S, Fruh M, Koeberle D, Oulkhouir Y, Schuler M, Curioni-Fontecedro A, Huret B, Kerjouan M, Michels S, Pall G, Rothschild S, Schmid-Bindert G, Scheffler M, Veillon R, Wannesson L, Diebold J, Zalcman G, Filleron T, Mazieres J. Targeted therapy for patients with BRAF-mutant lung cancer: results from the European EURAF Cohort. J Thorac Oncol. 2015;10(10):1451–7. https://doi.org/10.1097/jto.0000000000000625.
65. Planchard D, Besse B, Groen HJ, Souquet PJ, Quoix E, Baik CS, Barlesi F, Kim TM, Mazieres J, Novello S, Rigas JR, Upalawanna A, D'Amelio AM Jr, Zhang P, Mookerjee B, Johnson BE. Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial. Lancet Oncol. 2016;17(7):984–93. https://doi.org/10.1016/s1470-2045(16)30146-2.
66. Yao Z, Torres NM, Tao A, Gao Y, Luo L, Li Q, de Stanchina E, Abdel-Wahab O, Solit DB, Poulikakos PI, Rosen N. BRAF mutants evade ERK-dependent feedback by different mechanisms that determine their sensitivity to pharmacologic inhibition. Cancer Cell. 2015;28(3):370–83. https://doi.org/10.1016/j.ccell.2015.08.001.
67. Joshi M, Rice SJ, Liu X, Miller B, Belani CP. Trametinib with or without vemurafenib in BRAF mutated non-small cell lung cancer. PLoS One. 2015;10(2):e0118210. https://doi.org/10.1371/journal.pone.0118210.
68. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, Look AT. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science. 1994;263(5151):1281–4.
69. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561–6. https://doi.org/10.1038/nature05945.
70. Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, Choi HG, Kim J, Chiang D, Thomas R, Lee J, Richards WG, Sugarbaker DJ, Ducko C, Lindeman N, Marcoux JP, Engelman JA, Gray NS, Lee C, Meyerson M, Janne PA. EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. Clin Cancer Res. 2008;14(13):4275–83. https://doi.org/10.1158/1078-0432.ccr-08-0168.
71. Martelli MP, Sozzi G, Hernandez L, Pettirossi V, Navarro A, Conte D, Gasparini P, Perrone F, Modena P, Pastorino U, Carbone A, Fabbri A, Sidoni A, Nakamura S, Gambacorta M, Fernandez PL, Ramirez J, Chan JK, Grigioni WF, Campo E, Pileri SA, Falini B. EML4-ALK rearrangement in non-small cell lung cancer and non-tumor lung tissues. Am J Pathol. 2009;174(2):661–70. https://doi.org/10.2353/ajpath.2009.080755.
72. Sasaki T, Rodig SJ, Chirieac LR, Janne PA. The biology and treatment of EML4-ALK non-small cell lung cancer. Eur J Cancer. 2010;46(10):1773–80. https://doi.org/10.1016/j.ejca.2010.04.002.
73. Le AT, Varella-Garcia M, Doebele RC. Oncogenic fusions involving exon 19 of ALK. J Thorac Oncol. 2012;7(12):e44.; author reply e. https://doi.org/10.1097/JTO.0b013e31826bb94d.
74. Penzel R, Schirmacher P, Warth A. A novel EML4-ALK variant: exon 6 of EML4 fused to exon 19 of ALK. J Thorac Oncol. 2012;7(7):1198–9. https://doi.org/10.1097/JTO.0b013e3182598af3.
75. Katayama R, Lovly CM, Shaw AT. Therapeutic targeting of anaplastic lymphoma kinase in lung cancer: a paradigm for precision cancer medicine. Clin Cancer Res. 2015;21(10):2227–35. https://doi.org/10.1158/1078-0432.ccr-14-2791.
76. Kim HR, Shim HS, Chung JH, Lee YJ, Hong YK, Rha SY, Kim SH, Ha SJ, Kim SK, Chung KY, Soo R, Kim JH, Cho BC. Distinct clinical features and outcomes in never-smokers with nonsmall cell lung cancer who harbor EGFR or KRAS mutations or ALK rearrangement. Cancer. 2012;118(3):729–39. https://doi.org/10.1002/cncr.26311.
77. Shaw AT, Yeap BY, Mino-Kenudson M, Digumarthy SR, Costa DB, Heist RS, Solomon B, Stubbs H, Admane S, McDermott U, Settleman J, Kobayashi S, Mark EJ, Rodig SJ, Chirieac LR, Kwak EL, Lynch TJ, Iafrate AJ. Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. J Clin Oncol. 2009;27(26):4247–53. https://doi.org/10.1200/jco.2009.22.6993.
78. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Janne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, Fidias P, Stubbs H, Engelman JA, Sequist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Salgia R, Shapiro GI, Clark JW, Iafrate AJ. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med. 2010;363(18):1693–703. https://doi.org/10.1056/NEJMoa1006448.
79. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, De Pas T, Besse B, Solomon BJ, Blackhall F, Wu YL, Thomas M, O'Byrne KJ, Moro-Sibilot D, Camidge DR, Mok T, Hirsh V, Riely GJ, Iyer S, Tassell V, Polli A, Wilner KD, Janne PA. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med. 2013;368(25):2385–94. https://doi.org/10.1056/NEJMoa1214886.
80. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, Felip E, Cappuzzo F, Paolini J, Usari T, Iyer S, Reisman A, Wilner KD, Tursi J, Blackhall F. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. N Engl J Med. 2014;371(23):2167–77. https://doi.org/10.1056/NEJMoa1408440.
81. Takeuchi K, Choi YL, Togashi Y, Soda M, Hatano S, Inamura K, Takada S, Ueno T, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H. KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. Clin Cancer Res. 2009;15(9):3143–9. https://doi.org/10.1158/1078-0432.ccr-08-3248.
82. Fang DD, Zhang B, Gu Q, Lira M, Xu Q, Sun H, Qian M, Sheng W, Ozeck M, Wang Z, Zhang C, Chen X, Chen KX, Li J, Chen SH, Christensen J, Mao M, Chan CC. HIP1-ALK, a novel ALK fusion variant that responds to crizotinib. J Thorac Oncol. 2014;9(3):285–94. https://doi.org/10.1097/jto.0000000000000087.

83. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell. 2007;131(6):1190–203. https://doi.org/10.1016/j.cell.2007.11.025.
84. Yoshida T, Oya Y, Tanaka K, Shimizu J, Horio Y, Kuroda H, Sakao Y, Hida T, Yatabe Y. Differential crizotinib response duration among ALK fusion variants in ALK-positive non-small-cell lung cancer. J Clin Oncol. 2016;34(28):3383–9. https://doi.org/10.1200/jco.2015.65.8732.
85. Marchetti A, Di Lorito A, Pace MV, Iezzi M, Felicioni L, D'Antuono T, Filice G, Guetti L, Mucilli F, Buttitta F. ALK protein analysis by IHC staining after recent regulatory changes: a comparison of two widely used approaches, revision of the literature, and a new testing algorithm. J Thorac Oncol. 2016;11(4):487–95. https://doi.org/10.1016/j.jtho.2015.12.111.
86. Minca EC, Portier BP, Wang Z, Lanigan C, Farver CF, Feng Y, Ma PC, Arrossi VA, Pennell NA, Tubbs RR. ALK status testing in non-small cell lung carcinoma: correlation between ultrasensitive IHC and FISH. J Mol Diagn. 2013;15(3):341–6. https://doi.org/10.1016/j.jmoldx.2013.01.004.
87. von Laffert M, Schirmacher P, Warth A, Weichert W, Buttner R, Huber RM, Wolf J, Griesinger F, Dietel M, Grohe C. ALK-Testing in non-small cell lung cancer (NSCLC): Immunohistochemistry (IHC) and/or fluorescence in-situ Hybridisation (FISH)?: Statement of the Germany Society for Pathology (DGP) and the Working Group Thoracic Oncology (AIO) of the German Cancer Society e.V. (Stellungnahme der Deutschen Gesellschaft fur Pathologie und der AG Thorakale Onkologie der Arbeitsgemeinschaft Onkologie/Deutsche Krebsgesellschaft e.V.). Lung Cancer. 2017;103:1–5. https://doi.org/10.1016/j.lungcan.2016.11.008.
88. Lira ME, Choi YL, Lim SM, Deng S, Huang D, Ozeck M, Han J, Jeong JY, Shim HS, Cho BC, Kim J, Ahn MJ, Mao M. A single-tube multiplexed assay for detecting ALK, ROS1, and RET fusions in lung cancer. J Mol Diagn. 2014;16(2):229–43. https://doi.org/10.1016/j.jmoldx.2013.11.007.
89. Moskalev EA, Frohnauer J, Merkelbach-Bruse S, Schildhaus HU, Dimmler A, Schubert T, Boltze C, Konig H, Fuchs F, Sirbu H, Rieker RJ, Agaimy A, Hartmann A, Haller F. Sensitive and specific detection of EML4-ALK rearrangements in non-small cell lung cancer (NSCLC) specimens by multiplex amplicon RNA massive parallel sequencing. Lung Cancer. 2014;84(3):215–21. https://doi.org/10.1016/j.lungcan.2014.03.002.
90. Pfarr N, Stenzinger A, Penzel R, Warth A, Dienemann H, Schirmacher P, Weichert W, Endris V. High-throughput diagnostic profiling of clinically actionable gene fusions in lung cancer. Genes Chromosomes Cancer. 2016;55(1):30–44. https://doi.org/10.1002/gcc.22297.
91. Reguart N, Teixido C, Gimenez-Capitan A, Pare L, Galvan P, Viteri S, Rodriguez S, Peg V, Aldeguer E, Vinolas N, Remon J, Karachaliou N, Conde E, Lopez-Rios F, Nadal E, Merkelbach-Bruse S, Buttner R, Rosell R, Molina-Vila MA, Prat A. Identification of ALK, ROS1, and RET fusions by a multiplexed mRNA-based assay in formalin-fixed, paraffin-embedded samples from advanced non-small-cell lung cancer patients. Clin Chem. 2017;63(3):751–60. https://doi.org/10.1373/clinchem.2016.265314.
92. Choi YL, Soda M, Yamashita Y, Ueno T, Takashima J, Nakajima T, Yatabe Y, Takeuchi K, Hamada T, Haruta H, Ishikawa Y, Kimura H, Mitsudomi T, Tanio Y, Mano H. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. N Engl J Med. 2010;363(18):1734–9. https://doi.org/10.1056/NEJMoa1007478.
93. Doebele RC, Pilling AB, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, Kondo KL, Linderman DJ, Heasley LE, Franklin WA, Varella-Garcia M, Camidge DR. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clin Cancer Res. 2012;18(5):1472–82. https://doi.org/10.1158/1078-0432.ccr-11-2906.
94. Sasaki T, Koivunen J, Ogino A, Yanagita M, Nikiforow S, Zheng W, Lathan C, Marcoux JP, Du J, Okuda K, Capelletti M, Shimamura T, Ercan D, Stumpfova M, Xiao Y, Weremowicz S, Butaney M, Heon S, Wilner K, Christensen JG, Eck MJ, Wong KK, Lindeman N, Gray NS, Rodig SJ, Janne PA. A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. Cancer Res. 2011;71(18):6051–60. https://doi.org/10.1158/0008-5472.can-11-1340.
95. Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, Jessop NA, Wain JC, Yeo AT, Benes C, Drew L, Saeh JC, Crosby K, Sequist LV, Iafrate AJ, Engelman JA. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. Sci Transl Med. 2012;4(120):120ra17. https://doi.org/10.1126/scitranslmed.3003316.
96. Costa DB, Shaw AT, Ou SH, Solomon BJ, Riely GJ, Ahn MJ, Zhou C, Shreeve SM, Selaru P, Polli A, Schnell P, Wilner KD, Wiltshire R, Camidge DR, Crino L. Clinical experience with crizotinib in patients with advanced ALK-rearranged non-small-cell lung cancer and brain metastases. J Clin Oncol. 2015;33(17):1881–8. https://doi.org/10.1200/jco.2014.59.0539.
97. Crino L, Ahn MJ, De Marinis F, Groen HJ, Wakelee H, Hida T, Mok T, Spigel D, Felip E, Nishio M, Scagliotti G, Branle F, Emeremni C, Quadrigli M, Zhang J, Shaw AT. Multicenter phase II study of whole-body and intracranial activity with ceritinib in patients with ALK-rearranged non-small-cell lung cancer previously treated with chemotherapy and crizotinib: results from ASCEND-2. J Clin Oncol. 2016;34(24):2866–73. https://doi.org/10.1200/jco.2015.65.5936.
98. Shaw AT, Gandhi L, Gadgeel S, Riely GJ, Cetnar J, West H, Camidge DR, Socinski MA, Chiappori A, Mekhail T, Chao BH, Borghaei H, Gold KA, Zeaiter A, Bordogna W, Balas B, Puig O, Henschel V, Ou SH. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. Lancet Oncol. 2016;17(2):234–42. https://doi.org/10.1016/s1470-2045(15)00488-x.
99. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, Dagogo-Jack I, Gadgeel S, Schultz K, Singh M, Chin E, Parks M, Lee D, DiCecca RH, Lockerman E, Huynh T, Logan J, Ritterhouse LL, Le LP, Muniappan A, Digumarthy S, Channick C, Keyes C, Getz G, Dias-Santagata D, Heist RS, Lennerz J, Sequist LV, Benes CH, Iafrate AJ, Mino-Kenudson M, Engelman JA, Shaw AT. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. Cancer Discov. 2016;6(10):1118–33. https://doi.org/10.1158/2159-8290.cd-16-0596.
100. Bergethon K, Shaw AT, Ou SH, Katayama R, Lovly CM, McDonald NT, Massion PP, Siwak-Tapp C, Gonzalez A, Fang R, Mark EJ, Batten JM, Chen H, Wilner KD, Kwak EL, Clark JW, Carbone DP, Ji H, Engelman JA, Mino-Kenudson M, Pao W, Iafrate AJ. ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol. 2012;30(8):863–70. https://doi.org/10.1200/jco.2011.35.6345.
101. Li C, Fang R, Sun Y, Han X, Li F, Gao B, Iafrate AJ, Liu XY, Pao W, Chen H, Ji H. Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. PLoS One. 2011;6(11):e28204. https://doi.org/10.1371/journal.pone.0028204.
102. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H, Ishikawa Y. RET,

ROS1 and ALK fusions in lung cancer. Nat Med. 2012;18(3):378–81. https://doi.org/10.1038/nm.2658.
103. Jun HJ, Johnson H, Bronson RT, de Feraudy S, White F, Charest A. The oncogenic lung cancer fusion kinase CD74-ROS activates a novel invasiveness pathway through E-Syt1 phosphorylation. Cancer Res. 2012;72(15):3764–74. https://doi.org/10.1158/0008-5472.can-11-3990.
104. Rimkunas VM, Crosby KE, Li D, Hu Y, Kelly ME, Gu TL, Mack JS, Silver MR, Zhou X, Haack H. Analysis of receptor tyrosine kinase ROS1-positive tumors in non-small cell lung cancer: identification of a FIG-ROS1 fusion. Clin Cancer Res. 2012;18(16):4449–57. https://doi.org/10.1158/1078-0432.ccr-11-3351.
105. Lin JJ, Ritterhouse LL, Ali SM, Bailey M, Schrock AB, Gainor JF, Ferris LA, Mino-Kenudson M, Miller VA, Iafrate AJ, Lennerz JK, Shaw AT. ROS1 fusions rarely overlap with other oncogenic drivers in non-small cell lung cancer. J Thorac Oncol. 2017;12(5):872–7. https://doi.org/10.1016/j.jtho.2017.01.004.
106. Scheffler M, Schultheis A, Teixido C, Michels S, Morales-Espinosa D, Viteri S, Hartmann W, Merkelbach-Bruse S, Fischer R, Schildhaus HU, Fassunke J, Sebastian M, Serke M, Kaminsky B, Randerath W, Gerigk U, Ko YD, Kruger S, Schnell R, Rothe A, Kropf-Sanchen C, Heukamp L, Rosell R, Buttner R, Wolf J. ROS1 rearrangements in lung adenocarcinoma: prognostic impact, therapeutic options and genetic variability. Oncotarget. 2015;6(12):10577–85. https://doi.org/10.18632/oncotarget.3387.
107. McDermott U, Iafrate AJ, Gray NS, Shioda T, Classon M, Maheswaran S, Zhou W, Choi HG, Smith SL, Dowell L, Ulkus LE, Kuhlmann G, Greninger P, Christensen JG, Haber DA, Settleman J. Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. Cancer Res. 2008;68(9):3389–95. https://doi.org/10.1158/0008-5472.can-07-6186.
108. Shaw AT, Ou SH, Bang YJ, Camidge DR, Solomon BJ, Salgia R, Riely GJ, Varella-Garcia M, Shapiro GI, Costa DB, Doebele RC, Le LP, Zheng Z, Tan W, Stephenson P, Shreeve SM, Tye LM, Christensen JG, Wilner KD, Clark JW, Iafrate AJ. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med. 2014;371(21):1963–71. https://doi.org/10.1056/NEJMoa1406766.
109. Bubendorf L, Buttner R, Al-Dayel F, Dietel M, Elmberger G, Kerr K, Lopez-Rios F, Marchetti A, Oz B, Pauwels P, Penault-Llorca F, Rossi G, Ryska A, Thunnissen E. Testing for ROS1 in non-small cell lung cancer: a review with recommendations. Virchows Arch. 2016;469(5):489–503. https://doi.org/10.1007/s00428-016-2000-3.
110. Sholl LM, Sun H, Butaney M, Zhang C, Lee C, Janne PA, Rodig SJ. ROS1 immunohistochemistry for detection of ROS1-rearranged lung adenocarcinomas. Am J Surg Pathol. 2013;37(9):1441–9. https://doi.org/10.1097/PAS.0b013e3182960fa7.
111. Awad MM, Katayama R, McTigue M, Liu W, Deng YL, Brooun A, Friboulet L, Huang D, Falk MD, Timofeevski S, Wilner KD, Lockerman EL, Khan TM, Mahmood S, Gainor JF, Digumarthy SR, Stone JR, Mino-Kenudson M, Christensen JG, Iafrate AJ, Engelman JA, Shaw AT. Acquired resistance to crizotinib from a mutation in CD74-ROS1. N Engl J Med. 2013;368(25):2395–401. https://doi.org/10.1056/NEJMoa1215530.
112. Davare MA, Vellore NA, Wagner JP, Eide CA, Goodman JR, Drilon A, Deininger MW, O'Hare T, Druker BJ. Structural insight into selectivity and resistance profiles of ROS1 tyrosine kinase inhibitors. Proc Natl Acad Sci U S A. 2015;112(39):E5381–90. https://doi.org/10.1073/pnas.1515281112.
113. Song A, Kim TM, Kim DW, Kim S, Keam B, Lee SH, Heo DS. Molecular changes associated with acquired resistance to crizotinib in ROS1-rearranged non-small cell lung cancer. Clin Cancer Res. 2015;21(10):2379–87. https://doi.org/10.1158/1078-0432.ccr-14-1350.
114. Cargnelutti M, Corso S, Pergolizzi M, Mevellec L, Aisner DL, Dziadziuszko R, Varella-Garcia M, Comoglio PM, Doebele RC, Vialard J, Giordano S. Activation of RAS family members confers resistance to ROS1 targeting drugs. Oncotarget. 2015;6(7):5182–94. https://doi.org/10.18632/oncotarget.3311.
115. Davies KD, Mahale S, Astling DP, Aisner DL, Le AT, Hinz TK, Vaishnavi A, Bunn PA Jr, Heasley LE, Tan AC, Camidge DR, Varella-Garcia M, Doebele RC. Resistance to ROS1 inhibition mediated by EGFR pathway activation in non-small cell lung cancer. PLoS One. 2013;8(12):e82236. https://doi.org/10.1371/journal.pone.0082236.
116. Drilon A, Somwar R, Wagner JP, Vellore NA, Eide CA, Zabriskie MS, Arcila ME, Hechtman JF, Wang L, Smith RS, Kris MG, Riely GJ, Druker BJ, O'Hare T, Ladanyi M, Davare MA. A novel crizotinib-resistant solvent-front mutation responsive to cabozantinib therapy in a patient with ROS1-rearranged lung cancer. Clin Cancer Res. 2016;22(10):2351–8. https://doi.org/10.1158/1078-0432.ccr-15-2013.
117. Dziadziuszko R, Le AT, Wrona A, Jassem J, Camidge DR, Varella-Garcia M, Aisner DL, Doebele RC. An activating KIT mutation induces crizotinib resistance in ROS1-positive lung cancer. J Thorac Oncol. 2016;11(8):1273–81. https://doi.org/10.1016/j.jtho.2016.04.001.
118. Gerlinger M, Norton L, Swanton C. Acquired resistance to crizotinib from a mutation in CD74-ROS1. N Engl J Med. 2013;369(12):1172–3. https://doi.org/10.1056/NEJMc1309091#SA1.
119. Chong CR, Bahcall M, Capelletti M, Kosaka T, Ercan D, Sim T, Sholl LM, Nishino M, Johnson BE, Gray NS, Janne PA. Identification of existing drugs that effectively target NTRK1 and ROS1 rearrangements in lung cancer. Clin Cancer Res. 2017;23(1):204–13. https://doi.org/10.1158/1078-0432.ccr-15-1601.
120. Katayama R, Kobayashi Y, Friboulet L, Lockerman EL, Koike S, Shaw AT, Engelman JA, Fujita N. Cabozantinib overcomes crizotinib resistance in ROS1 fusion-positive cancer. Clin Cancer Res. 2015;21(1):166–74. https://doi.org/10.1158/1078-0432.ccr-14-1385.
121. Solomon BJ, Bauer TM, Felip E, Besse B, James LP, Clancy JS, Klamerus KJ, Martini J-F, Abbattista A, Shaw AT. Safety and efficacy of lorlatinib (PF-06463922) from the dose-escalation component of a study in patients with advanced ALK+ or ROS1+ non-small cell lung cancer (NSCLC). J Clin Oncol. 2016;34(15_suppl):9009. https://doi.org/10.1200/JCO.2016.34.15_suppl.9009.
122. Drilon A, S Siena SIO, Patel M, Ahn MJ, Lee J, Bauer TM, Farago AF, Wheler JJ, Liu SV, Doebele R, Giannetta L, Cerea G, Marrapese G, Schirru M, Amatu A, Bencardino K, Palmeri L, Sartore-Bianchi A, Vanzulli A, Cresta S, Damian S, Duca M, Ardini E, Li G, Christiansen J, Kowalski K, Johnson AD, Patel R, Luo D, Chow-Maneval E, Hornby Z, Multani PS, Shaw AT, De Braud FG. Safety and antitumor activity of the multitargeted pan-TRK, ROS1, and ALK inhibitor entrectinib: combined results from two phase I trials (ALKA-372-001 and STARTRK-1). Cancer Discov. 2017;7(4):400–9. https://doi.org/10.1158/2159-8290.cd-16-1237.
123. Riely GJ, Marks J, Pao W. KRAS mutations in non-small cell lung cancer. Proc Am Thorac Soc. 2009;6(2):201–5. https://doi.org/10.1513/pats.200809-107LC.
124. Karachaliou N, Mayo C, Costa C, Magri I, Gimenez-Capitan A, Molina-Vila MA, Rosell R. KRAS mutations in lung cancer. Clin Lung Cancer. 2013;14(3):205–14. https://doi.org/10.1016/j.cllc.2012.09.007.
125. Martin P, Leighl NB, Tsao MS, Shepherd FA. KRAS mutations as prognostic and predictive markers in non-small cell lung cancer. J Thorac Oncol. 2013;8(5):530–42. https://doi.org/10.1097/JTO.0b013e318283d958.

126. Linardou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S. Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. Lancet Oncol. 2008;9(10):962–72. https://doi.org/10.1016/s1470-2045(08)70206-7.
127. Mao C, Qiu LX, Liao RY, Du FB, Ding H, Yang WC, Li J, Chen Q. KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: a meta-analysis of 22 studies. Lung Cancer. 2010;69(3):272–8. https://doi.org/10.1016/j.lungcan.2009.11.020.
128. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature. 2013;503(7477):548–51. https://doi.org/10.1038/nature12796.
129. Blumenschein GR Jr, Smit EF, Planchard D, Kim DW, Cadranel J, De Pas T, Dunphy F, Udud K, Ahn MJ, Hanna NH, Kim JH, Mazieres J, Kim SW, Baas P, Rappold E, Redhu S, Puski A, Wu FS, Janne PA. A randomized phase II study of the MEK1/MEK2 inhibitor trametinib (GSK1120212) compared with docetaxel in KRAS-mutant advanced non-small-cell lung cancer (NSCLC) dagger. Ann Oncol. 2015;26(5):894–901. https://doi.org/10.1093/annonc/mdv072.
130. Lopez-Chavez A, Thomas A, Rajan A, Raffeld M, Morrow B, Kelly R, Carter CA, Guha U, Killian K, Lau CC, Abdullaev Z, Xi L, Pack S, Meltzer PS, Corless CL, Sandler A, Beadling C, Warrick A, Liewehr DJ, Steinberg SM, Berman A, Doyle A, Szabo E, Wang Y, Giaccone G. Molecular profiling and targeted therapy for advanced thoracic malignancies: a biomarker-derived, multiarm, multihistology phase II basket trial. J Clin Oncol. 2015;33(9):1000–7. https://doi.org/10.1200/jco.2014.58.2007.
131. Holt SV, Logie A, Davies BR, Alferez D, Runswick S, Fenton S, Chresta CM, Gu Y, Zhang J, Wu YL, Wilkinson RW, Guichard SM, Smith PD. Enhanced apoptosis and tumor growth suppression elicited by combination of MEK (selumetinib) and mTOR kinase inhibitors (AZD8055). Cancer Res. 2012;72(7):1804–13. https://doi.org/10.1158/0008-5472.can-11-1780.
132. Sos ML, Fischer S, Ullrich R, Peifer M, Heuckmann JM, Koker M, Heynck S, Stuckrath I, Weiss J, Fischer F, Michel K, Goel A, Regales L, Politi KA, Perera S, Getlik M, Heukamp LC, Ansen S, Zander T, Beroukhim R, Kashkar H, Shokat KM, Sellers WR, Rauh D, Orr C, Hoeflich KP, Friedman L, Wong KK, Pao W, Thomas RK. Identifying genotype-dependent efficacy of single and combined PI3K- and MAPK-pathway inhibition in cancer. Proc Natl Acad Sci U S A. 2009;106(43):18351–6. https://doi.org/10.1073/pnas.0907325106.
133. Clinical Trails. A study of prexasertib (LY2606368) in combination with ralimetinib in participants with advanced or metastatic cancer [Internet] 2017 Mar 17 [cited 2017 Jun 22]. Available from: https://clinicaltrials.gov/ct2/show/NCT02860780.
134. Lung Cancer Group Cologne. Trials [Internet] 2017 [cited 2017 Jun 22]. Available from: http://lungcancergroup.de/en/studienuebersicht/
135. Clinical Trails. Study of LXH254 and LTT462 in NSCLC [Internet] 2017 Jun 18 [cited 2017 Jun 22]. Available from: https://clinicaltrials.gov/ct2/show/NCT02974725
136. Skoulidis F, Byers LA, Diao L, Papadimitrakopoulou VA, Tong P, Izzo J, Behrens C, Kadara H, Parra ER, Canales JR, Zhang J, Giri U, Gudikote J, Cortez MA, Yang C, Fan Y, Peyton M, Girard L, Coombes KR, Toniatti C, Heffernan TP, Choi M, Frampton GM, Miller V, Weinstein JN, Herbst RS, Wong KK, Sharma P, Mills GB, Hong WK, Minna JD, Allison JP, Futreal A, Wang J, Wistuba II, Heymach JV. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. Cancer Discov. 2015;5(8):860–77. https://doi.org/10.1158/2159-8290.cd-14-1236.
137. Heinmoller P, Gross C, Beyser K, Schmidtgen C, Maass G, Pedrocchi M, Ruschoff J. HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. Clin Cancer Res. 2003;9(14):5238–43.
138. Krug LM, Miller VA, Patel J, Crapanzano J, Azzoli CG, Gomez J, Kris MG, Heelan RT, Pizzo B, Tyson L, Sheehan C, Ross JS, Venkatraman E. Randomized phase II study of weekly docetaxel plus trastuzumab versus weekly paclitaxel plus trastuzumab in patients with previously untreated advanced nonsmall cell lung carcinoma. Cancer. 2005;104(10):2149–55. https://doi.org/10.1002/cncr.21428.
139. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, Paz-Ares L. Lung cancer: current therapies and new targeted treatments. Lancet. 2017;389(10066):299–311. https://doi.org/10.1016/s0140-6736(16)30958-8.
140. Suzawa K, Toyooka S, Sakaguchi M, Morita M, Yamamoto H, Tomida S, Ohtsuka T, Watanabe M, Hashida S, Maki Y, Soh J, Asano H, Tsukuda K, Miyoshi S. Antitumor effect of afatinib, as a human epidermal growth factor receptor 2-targeted therapy, in lung cancers harboring HER2 oncogene alterations. Cancer Sci. 2016;107(1):45–52. https://doi.org/10.1111/cas.12845.
141. Shigematsu H, Takahashi T, Nomura M, Majmudar K, Suzuki M, Lee H, Wistuba II, Fong KM, Toyooka S, Shimizu N, Fujisawa T, Minna JD, Gazdar AF. Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. Cancer Res. 2005;65(5):1642–6. https://doi.org/10.1158/0008-5472.can-04-4235.
142. Arcila ME, Chaft JE, Nafa K, Roy-Chowdhuri S, Lau C, Zaidinski M, Paik PK, Zakowski MF, Kris MG, Ladanyi M. Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. Clin Cancer Res. 2012;18(18):4910–8. https://doi.org/10.1158/1078-0432.ccr-12-0912.
143. Bu S, Wang R, Pan Y, S Y, Shen X, Li Y, Sun Y, Chen H. Clinicopathologic characteristics of patients with HER2 insertions in non-small cell lung cancer. Ann Surg Oncol. 2017;24(1):291–7. https://doi.org/10.1245/s10434-015-5044-8.
144. Shimamura T, Ji H, Minami Y, Thomas RK, Lowell AM, Shah K, Greulich H, Glatt KA, Meyerson M, Shapiro GI, Wong KK. Non-small-cell lung cancer and Ba/F3 transformed cells harboring the ERBB2 G776insV_G/C mutation are sensitive to the dual-specific epidermal growth factor receptor and ERBB2 inhibitor HKI-272. Cancer Res. 2006;66(13):6487–91. https://doi.org/10.1158/0008-5472.can-06-0971.
145. Kris MG, Camidge DR, Giaccone G, Hida T, Li BT, O'Connell J, Taylor I, Zhang H, Arcila ME, Goldberg Z, Janne PA. Targeting HER2 aberrations as actionable drivers in lung cancers: phase II trial of the pan-HER tyrosine kinase inhibitor dacomitinib in patients with HER2-mutant or amplified tumors. Ann Oncol. 2015;26(7):1421–7. https://doi.org/10.1093/annonc/mdv186.
146. Kosaka T, Tanizaki J, Paranal RM, Endoh H, Lydon C, Capelletti M, Repellin CE, Choi J, Ogino A, Calles A, Ercan D, Redig AJ, Bahcall M, Oxnard GR, Eck MJ, Janne PA. Response heterogeneity of EGFR and HER2 exon 20 insertions to covalent EGFR and HER2 inhibitors. Cancer Res. 2017;77(10):2712–21. https://doi.org/10.1158/0008-5472.can-16-3404.
147. De Greve J, Teugels E, Geers C, Decoster L, Galdermans D, De Mey J, Everaert H, Umelo I, In't Veld P, Schallier D. Clinical activity of afatinib (BIBW 2992) in patients with lung adenocarcinoma with mutations in the kinase domain of HER2/neu. Lung Cancer. 2012;76(1):123–7. https://doi.org/10.1016/j.lungcan.2012.01.008.
148. Li BT, Lee A, O'Toole S, W Cooper BY, Chaft JE, Arcila ME, Kris MG, Pavlakis N. HER2 insertion YVMA mutant lung cancer: long natural history and response to afatinib.

Lung Cancer. 2015;90(3):617–9. https://doi.org/10.1016/j.lungcan.2015.10.025.
149. Weiler D, Diebold J, Strobel K, Aebi S, Gautschi O. Rapid response to trastuzumab emtansine in a patient with HER2-driven lung cancer. J Thorac Oncol. 2015;10(4):e16–7. https://doi.org/10.1097/jto.0000000000000424.
150. Seki T, Hagiya M, Shimonishi M, Nakamura T, Shimizu S. Organization of the human hepatocyte growth factor-encoding gene. Gene. 1991;102(2):213–9.
151. Awad MM, Oxnard GR, Jackman DM, Savukoski DO, Hall D, Shivdasani P, Heng JC, Dahlberg SE, Janne PA, Verma S, Christensen J, Hammerman PS, Sholl LM. MET exon 14 mutations in non-small-cell lung cancer are associated with advanced age and stage-dependent MET genomic amplification and c-Met overexpression. J Clin Oncol. 2016;34(7):721–30. https://doi.org/10.1200/jco.2015.63.4600.
152. Drilon A. MET exon 14 alterations in lung cancer: exon skipping extends half-life. Clin Cancer Res. 2016;22(12):2832–4. https://doi.org/10.1158/1078-0432.ccr-16-0229.
153. Frampton GM, Ali SM, Rosenzweig M, Chmielecki J, Lu X, Bauer TM, Akimov M, Bufill JA, Lee C, Jentz D, Hoover R, Ou SH, Salgia R, Brennan T, Chalmers ZR, Jaeger S, Huang A, Elvin JA, Erlich R, Fichtenholtz A, Gowen KA, Greenbowe J, Johnson A, Khaira D, McMahon C, Sanford EM, Roels S, White J, Greshock J, Schlegel R, Lipson D, Yelensky R, Morosini D, Ross JS, Collisson E, Peters M, Stephens PJ, Miller VA. Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors. Cancer Discov. 2015;5(8):850–9. https://doi.org/10.1158/2159-8290.cd-15-0285.
154. Kong-Beltran M, Seshagiri S, Zha J, Zhu W, Bhawe K, Mendoza N, Holcomb T, Pujara K, Stinson J, L F, Severin C, Rangell L, Schwall R, Amler L, Wickramasinghe D, Yauch R. Somatic mutations lead to an oncogenic deletion of met in lung cancer. Cancer Res. 2006;66(1):283–9. https://doi.org/10.1158/0008-5472.can-05-2749.
155. Okuda K, Sasaki H, Yukiue H, Yano M, Fujii Y. Met gene copy number predicts the prognosis for completely resected non-small cell lung cancer. Cancer Sci. 2008;99(11):2280–5. https://doi.org/10.1111/j.1349-7006.2008.00916.x.
156. Onozato R, Kosaka T, Kuwano H, Sekido Y, Yatabe Y, Mitsudomi T. Activation of MET by gene amplification or by splice mutations deleting the juxtamembrane domain in primary resected lung cancers. J Thorac Oncol. 2009;4(1):5–11. https://doi.org/10.1097/JTO.0b013e3181913e0e.
157. Krishnaswamy S, Kanteti R, Duke-Cohan JS, Loganathan S, Liu W, Ma PC, Sattler M, Singleton PA, Ramnath N, Innocenti F, Nicolae DL, Ouyang Z, Liang J, Minna J, Kozloff MF, Ferguson MK, Natarajan V, Wang YC, Garcia JG, Vokes EE, Salgia R. Ethnic differences and functional analysis of MET mutations in lung cancer. Clin Cancer Res. 2009;15(18):5714–23. https://doi.org/10.1158/1078-0432.ccr-09-0070.
158. Heist RS, Shim HS, Gingipally S, Mino-Kenudson M, Le L, Gainor JF, Zheng Z, Aryee M, Xia J, Jia P, Jin H, Zhao Z, Pao W, Engelman JA, Iafrate AJ. MET exon 14 skipping in non-small cell lung cancer. Oncologist. 2016;21(4):481–6. https://doi.org/10.1634/theoncologist.2015-0510.
159. Jenkins RW, Oxnard GR, Elkin S, Sullivan EK, Carter JL, Barbie DA. Response to crizotinib in a patient with lung adenocarcinoma harboring a MET splice site mutation. Clin Lung Cancer. 2015;16(5):e101–4. https://doi.org/10.1016/j.cllc.2015.01.009.
160. Paik PK, Shen R, Won H, Rekhtman N, Wang L, Sima CS, Arora A, Seshan V, Ladanyi M, Berger MF, Kris MG. Next-generation sequencing of stage IV squamous cell lung cancers reveals an association of PI3K aberrations and evidence of clonal heterogeneity in patients with brain metastases. Cancer Discov. 2015;5(6):610–21. https://doi.org/10.1158/2159-8290.cd-14-1129.
161. Schrock AB, Frampton GM, Suh J, Chalmers ZR, Rosenzweig M, Erlich RL, Halmos B, Goldman J, Forde P, Leuenberger K, Peled N, Kalemkerian GP, Ross JS, Stephens PJ, Miller VA, Ali SM, Ou SH. Characterization of 298 patients with lung cancer harboring MET exon 14 skipping alterations. J Thorac Oncol. 2016;11(9):1493–502. https://doi.org/10.1016/j.jtho.2016.06.004.
162. Paik PK, Drilon A, PD Fan HY, Rekhtman N, MS Ginsberg L, Borsu NS, Berger MF, CM Rudin ML. Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. Cancer Discov. 2015;5(8):842–9. https://doi.org/10.1158/2159-8290.cd-14-1467.
163. Cappuzzo F, Janne PA, Skokan M, Finocchiaro G, Rossi E, Ligorio C, Zucali PA, Terracciano L, Toschi L, Roncalli M, Destro A, Incarbone M, Alloisio M, Santoro A, Varella-Garcia M. MET increased gene copy number and primary resistance to gefitinib therapy in non-small-cell lung cancer patients. Ann Oncol. 2009;20(2):298–304. https://doi.org/10.1093/annonc/mdn635.
164. Schildhaus HU, Schultheis AM, Ruschoff J, Binot E, Merkelbach-Bruse S, Fassunke J, Schulte W, Ko YD, Schlesinger A, Bos M, Gardizi M, Engel-Riedel W, Brockmann M, Serke M, Gerigk U, Hekmat K, Frank KF, Reiser M, Schulz H, Kruger S, Stoelben E, Zander T, Wolf J, Buettner R. MET amplification status in therapy-naive adeno- and squamous cell carcinomas of the lung. Clin Cancer Res. 2015;21(4):907–15. https://doi.org/10.1158/1078-0432.ccr-14-0450.
165. Ou SH, Kwak EL, Siwak-Tapp C, Dy J, Bergethon K, Clark JW, Camidge DR, Solomon BJ, Maki RG, Bang YJ, Kim DW, Christensen J, Tan W, Wilner KD, Salgia R, Iafrate AJ. Activity of crizotinib (PF02341066), a dual mesenchymal-epithelial transition (MET) and anaplastic lymphoma kinase (ALK) inhibitor, in a non-small cell lung cancer patient with de novo MET amplification. J Thorac Oncol. 2011;6(5):942–6. https://doi.org/10.1097/JTO.0b013e31821528d3.
166. Tanizaki J, Okamoto I, Okamoto K, Takezawa K, Kuwata K, Yamaguchi H, Nakagawa K. MET tyrosine kinase inhibitor crizotinib (PF-02341066) shows differential antitumor effects in non-small cell lung cancer according to MET alterations. J Thorac Oncol. 2011;6(10):1624–31. https://doi.org/10.1097/JTO.0b013e31822591e9.
167. Clinical Trails. Clinical study of oral cMET inhibitor INC280 in adult patients with EGFR wild-type advanced non-small cell lung cancer [Internet] 2017 Mar 29 [cited 2017 Jun 22]. Available from: https://clinicaltrials.gov/ct2/show/NCT02414139.
168. Phay JE, Shah MH. Targeting RET receptor tyrosine kinase activation in cancer. Clin Cancer Res. 2010;16(24):5936–41. https://doi.org/10.1158/1078-0432.ccr-09-0786.
169. Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH, Seo JS. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res. 2012;22(3):436–45. https://doi.org/10.1101/gr.133645.111.
170. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, Shimada Y, Iwakawa R, Ogiwara H, Oike T, Enari M, Schetter AJ, Okayama H, Haugen A, Skaug V, Chiku S, Yamanaka I, Arai Y, Watanabe S, Sekine I, Ogawa S, Harris CC, Tsuda H, Yoshida T, Yokota J, Shibata T. KIF5B-RET fusions in lung adenocarcinoma. Nat Med. 2012;18(3):375–7. https://doi.org/10.1038/nm.2644.
171. Michels S, Scheel AH, Scheffler M, Schultheis AM, Gautschi O, Aebersold F, Diebold J, Pall G, Rothschild S, Bubendorf L, Hartmann W, Heukamp L, Schildhaus HU, Fassunke J, Ihle MA, Kunstlinger H, Heydt C, Fischer R, Nogova L, Mattonet C, Hein R, Adams A, Gerigk U, Schulte W, Luders H, Grohe C, Graeven U, Muller-Naendrup C, Draube A, Kambartel

KO, Kruger S, Schulze-Olden S, Serke M, Engel-Riedel W, Kaminsky B, Randerath W, Merkelbach-Bruse S, Buttner R, Wolf J. Clinicopathological characteristics of RET rearranged lung cancer in European patients. J Thorac Oncol. 2016;11(1):122–7. https://doi.org/10.1016/j.jtho.2015.09.016.

172. Wang R, H H, Pan Y, Li Y, Ye T, Li C, Luo X, Wang L, Li H, Zhang Y, Li F, Y L, Q L, J X, Garfield D, Shen L, Ji H, Pao W, Sun Y, Chen H. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J Clin Oncol. 2012;30(35):4352–9. https://doi.org/10.1200/jco.2012.44.1477.
173. Lee MS, RN Kim HI, Oh DY, Song JY, Noh KW, Kim YJ, Yang JW, Lira ME, Lee CH, Lee MK, Kim YD, Mao M, Han J, Kim J, Choi YL. Identification of a novel partner gene, KIAA1217, fused to RET: Functional characterization and inhibitor sensitivity of two isoforms in lung adenocarcinoma. Oncotarget. 2016;7(24):36101–14. https://doi.org/10.18632/oncotarget.9137.
174. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC, White E, White J, Zwirko Z, Peretz T, Nechushtan H, Soussan-Gutman L, Kim J, Sasaki H, Kim HR, Park SI, Ercan D, Sheehan CE, Ross JS, Cronin MT, Janne PA, Stephens PJ. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012;18(3):382–4. https://doi.org/10.1038/nm.2673.
175. Lee SH, Lee JK, Ahn MJ, Kim DW, Sun JM, Keam B, Kim TM, Heo DS, Ahn JS, Choi YL, Min HS, Jeon YK, Park K. Vandetanib in pretreated patients with advanced non-small cell lung cancer-harboring RET rearrangement: a phase II clinical trial. Ann Oncol. 2017;28(2):292–7. https://doi.org/10.1093/annonc/mdw559.
176. Seto T, Yoh K, Satouchi M, Nishio M, Yamamoto N, Murakami H, Nogami N, Nosaki K, Urata Y, Niho S, Horiike A, Kohno T, Matsumoto S, Nomura S, Kuroda S, Sato A, Ohe Y, Yamanaka T, Ohtsu A, Goto K. A phase II open-label single-arm study of vandetanib in patients with advanced RET-rearranged non-small cell lung cancer (NSCLC): Luret study. J Clin Oncol. 2016;34(15_suppl):9012. https://doi.org/10.1200/JCO.2016.34.15_suppl.9012.
177. Kodama T, Tsukaguchi T, Satoh Y, Yoshida M, Watanabe Y, Kondoh O, Sakamoto H. Alectinib shows potent antitumor activity against RET-rearranged non-small cell lung cancer. Mol Cancer Ther. 2014;13(12):2910–8. https://doi.org/10.1158/1535-7163.mct-14-0274.
178. Lin JJ, Kennedy E, Sequist LV, Brastianos PK, Goodwin KE, Stevens S, Wanat AC, Stober LL, Digumarthy SR, Engelman JA, Shaw AT, Gainor JF. Clinical activity of alectinib in advanced RET-rearranged non-small cell lung cancer. J Thorac Oncol. 2016;11(11):2027–32. https://doi.org/10.1016/j.jtho.2016.08.126.
179. Drilon A, Wang L, Hasanovic A, Suehara Y, Lipson D, Stephens P, Ross J, Miller V, Ginsberg M, Zakowski MF, Kris MG, Ladanyi M, Rizvi N. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. Cancer Discov. 2013;3(6):630–5. https://doi.org/10.1158/2159-8290.cd-13-0035.
180. Drilon A, Rekhtman N, Arcila M, Wang L, Ni A, Albano M, Van Voorthuysen M, Somwar R, Smith RS, Montecalvo J, Plodkowski A, Ginsberg MS, Riely GJ, Rudin CM, Ladanyi M, Kris MG. Cabozantinib in patients with advanced RET-rearranged non-small-cell lung cancer: an open-label, single-centre, phase 2, single-arm trial. Lancet Oncol. 2016;17(12):1653–60. https://doi.org/10.1016/s1470-2045(16)30562-9.
181. Alberti L, Carniti C, Miranda C, Roccato E, Pierotti MA. RET and NTRK1 proto-oncogenes in human diseases. J Cell Physiol. 2003;195(2):168–86. https://doi.org/10.1002/jcp.10252.
182. Sossin WS. Tracing the evolution and function of the Trk superfamily of receptor tyrosine kinases. Brain Behav Evol. 2006;68(3):145–56. https://doi.org/10.1159/000094084.
183. Farago AF, Le LP, Zheng Z, Muzikansky A, Drilon A, Patel M, Bauer TM, Liu SV, Ou SH, Jackman D, Costa DB, Multani PS, Li GG, Hornby Z, Chow-Maneval E, Luo D, Lim JE, Iafrate AJ, Shaw AT. Durable clinical response to entrectinib in NTRK1-rearranged non-small cell lung cancer. J Thorac Oncol. 2015;10(12):1670–4. https://doi.org/10.1097/01.JTO.0000473485.38553.f0.
184. Vaishnavi A, Capelletti M, Le AT, Kako S, Butaney M, Ercan D, Mahale S, Davies KD, Aisner DL, Pilling AB, Berge EM, Kim J, Sasaki H, Park SI, Kryukov G, Garraway LA, Hammerman PS, Haas J, Andrews SW, Lipson D, Stephens PJ, Miller VA, Varella-Garcia M, Janne PA, Doebele RC. Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer. Nat Med. 2013;19(11):1469–72. https://doi.org/10.1038/nm.3352.
185. Ardini E, Menichincheri M, Banfi P, Bosotti R, De Ponti C, Pulci R, Ballinari D, Ciomei M, Texido G, Degrassi A, Avanzi N, Amboldi N, Saccardo MB, Casero D, Orsini P, Bandiera T, Mologni L, Anderson D, Wei G, Harris J, Vernier JM, Li G, Felder E, Donati D, Isacchi A, Pesenti E, Magnaghi P, Galvani A. Entrectinib, a pan-TRK, ROS1, and ALK inhibitor with activity in multiple molecularly defined cancer indications. Mol Cancer Ther. 2016;15(4):628–39. https://doi.org/10.1158/1535-7163.mct-15-0758.
186. Braud FGD, Niger M, Damian S, Bardazza B, Martinetti A, Pelosi G, Marrapese G, Palmeri L, Cerea G, Valtorta E, Veronese S, Sartore-Bianchi A, Ardini E, Isachi A, Martignoni M, Galvani A, Luo D, Yeh L, Senderowicz AM, Siena S. Alka-372-001: first-in-human, phase I study of entrectinib – an oral pan-trk, ROS1, and ALK inhibitor – in patients with advanced solid tumors with relevant molecular alterations. J Clin Oncol. 2015;33(15_suppl):2517. https://doi.org/10.1200/jco.2015.33.15_suppl.2517.
187. Doebele RC, Davis LE, Vaishnavi A, Le AT, Estrada-Bernal A, Keysar S, Jimeno A, Varella-Garcia M, Aisner DL, Li Y, Stephens PJ, Morosini D, Tuch BB, Fernandes M, Nanda N, Low JA. An oncogenic NTRK fusion in a patient with soft-tissue sarcoma with response to the tropomyosin-related kinase inhibitor LOXO-101. Cancer Discov. 2015;5(10):1049–57. https://doi.org/10.1158/2159-8290.cd-15-0443.
188. Derman BA, Mileham KF, Bonomi PD, Batus M, Fidler MJ. Treatment of advanced squamous cell carcinoma of the lung: a review. Transl Lung Cancer Res. 2015;4(5):524–32. https://doi.org/10.3978/j.issn.2218-6751.2015.06.07.
189. Kerr KM, Bubendorf L, Edelman MJ, Marchetti A, Mok T, Novello S, O'Byrne K, Stahel R, Peters S, Felip E. Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer. Ann Oncol. 2014;25(9):1681–90. https://doi.org/10.1093/annonc/mdu145.
190. Coutts JC, Gallagher JT. Receptors for fibroblast growth factors. Immunol Cell Biol. 1995;73(6):584–9. https://doi.org/10.1038/icb.1995.92.
191. Helsten T, Elkin S, Arthur E, Tomson BN, Carter J, Kurzrock R. The FGFR landscape in cancer: analysis of 4,853 tumors by next-generation sequencing. Clin Cancer Res. 2016;22(1):259–67. https://doi.org/10.1158/1078-0432.ccr-14-3212.
192. Liao RG, Jung J, Tchaicha J, Wilkerson MD, Sivachenko A, Beauchamp EM, Liu Q, Pugh TJ, Pedamallu CS, Hayes DN, Gray NS, Getz G, Wong KK, Haddad RI, Meyerson M, Hammerman PS. Inhibitor-sensitive FGFR2 and FGFR3 mutations in lung squamous cell carcinoma. Cancer Res. 2013;73(16):5195–205. https://doi.org/10.1158/0008-5472.can-12-3950.
193. Hibi M, Kaneda H, Tanizaki J, Sakai K, Togashi Y, Terashima M, De Velasco MA, Fujita Y, Banno E, Nakamura Y, Takeda M, Ito A, Mitsudomi T, Nakagawa K, Okamoto I, Nishio K. FGFR gene alterations in lung squamous cell carcinoma are poten-

tial targets for the multikinase inhibitor nintedanib. Cancer Sci. 2016;107(11):1667–76. https://doi.org/10.1111/cas.13071.
194. Dutt A, Ramos AH, Hammerman PS, Mermel C, Cho J, Sharifnia T, Chande A, Tanaka KE, Stransky N, Greulich H, Gray NS, Meyerson M. Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer. PLoS One. 2011;6(6):e20351. https://doi.org/10.1371/journal.pone.0020351.
195. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, Ullrich RT, Menon R, Maier S, Soltermann A, Moch H, Wagener P, Fischer F, Heynck S, Koker M, Schottle J, Leenders F, Gabler F, Dabow I, Querings S, Heukamp LC, Balke-Want H, Ansen S, Rauh D, Baessmann I, Altmuller J, Wainer Z, Conron M, Wright G, Russell P, Solomon B, Brambilla E, Brambilla C, Lorimier P, Sollberg S, Brustugun OT, Engel-Riedel W, Ludwig C, Petersen I, Sanger J, Clement J, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman D, Cappuzzo F, Ligorio C, Damiani S, Hallek M, Beroukhim R, Pao W, Klebl B, Baumann M, Buettner R, Ernestus K, Stoelben E, Wolf J, Nurnberg P, Perner S, Thomas RK. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med. 2010;2(62):62ra93. https://doi.org/10.1126/scitranslmed.3001451.
196. Schildhaus HU, Heukamp LC, Merkelbach-Bruse S, Riesner K, Schmitz K, Binot E, Paggen E, Albus K, Schulte W, Ko YD, Schlesinger A, Ansen S, Engel-Riedel W, Brockmann M, Serke M, Gerigk U, Huss S, Goke F, Perner S, Hekmat K, Frank KF, Reiser M, Schnell R, Bos M, Mattonet C, Sos M, Stoelben E, Wolf J, Zander T, Buettner R. Definition of a fluorescence in-situ hybridization score identifies high- and low-level FGFR1 amplification types in squamous cell lung cancer. Mod Pathol. 2012;25(11):1473–80. https://doi.org/10.1038/modpathol.2012.102.
197. Nogova L, Sequist LV, Garcia JMP, Andre F, Delord J-P, Hidalgo M, Schellens JHM, Cassier PA, Camidge DR, Schuler M, Vaishampayan U, Burris H, Tian GG, Campone M, Wainberg ZA, Lim W-T, LoRusso P, Shapiro GI, Parker K, Chen X, Choudhury S, Ringeisen F, Graus-Porta D, Porter D, Isaacs R, Buettner R, Wolf J. Evaluation of BGJ398, a fibroblast growth factor receptor 1-3 kinase inhibitor, in patients with advanced solid tumors harboring genetic alterations in fibroblast growth factor receptors: results of a global phase I, dose-escalation and dose-expansion study. J Clin Oncol. 2017;35(2):157–65. https://doi.org/10.1200/jco.2016.67.2048.
198. Day E, Waters B, Spiegel K, Alnadaf T, Manley PW, Buchdunger E, Walker C, Jarai G. Inhibition of collagen-induced discoidin domain receptor 1 and 2 activation by imatinib, nilotinib and dasatinib. Eur J Pharmacol. 2008;599(1–3):44–53. https://doi.org/10.1016/j.ejphar.2008.10.014.
199. An SJ, Chen ZH, Su J, Zhang XC, Zhong WZ, Yang JJ, Zhou Q, Yang XN, Huang L, Guan JL, Nie Q, Yan HH, Mok TS, Wu YL. Identification of enriched driver gene alterations in subgroups of non-small cell lung cancer patients based on histology and smoking status. PLoS One. 2012;7(6):e40109. https://doi.org/10.1371/journal.pone.0040109.
200. Hammerman PS, Sos ML, Ramos AH, Xu C, Dutt A, Zhou W, Brace LE, Woods BA, Lin W, Zhang J, Deng X, Lim SM, Heynck S, Peifer M, Simard JR, Lawrence MS, Onofrio RC, Salvesen HB, Seidel D, Zander T, Heuckmann JM, Soltermann A, Moch H, Koker M, Leenders F, Gabler F, Querings S, Ansen S, Brambilla E, Brambilla C, Lorimier P, Brustugun OT, Helland A, Petersen I, Clement JH, Groen H, Timens W, Sietsma H, Stoelben E, Wolf J, Beer DG, Tsao MS, Hanna M, Hatton C, Eck MJ, Janne PA, Johnson BE, Winckler W, Greulich H, Bass AJ, Cho J, Rauh D, Gray NS, Wong KK, Haura EB, Thomas RK, Meyerson M. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. Cancer Discov. 2011;1(1):78–89. https://doi.org/10.1158/2159-8274.cd-11-0005.
201. Xu C, KA Buczkowski Y, Zhang HA, Beauchamp EM, Terai H, YY Li MM, Wong KK, Hammerman PS. NSCLC driven by DDR2 mutation is sensitive to dasatinib and JQ1 combination therapy. Mol Cancer Ther. 2015;14(10):2382–9. https://doi.org/10.1158/1535-7163.mct-15-0077.
202. Karakas B, Bachman KE, Park BH. Mutation of the PIK3CA oncogene in human cancers. Br J Cancer. 2006;94(4):455–9. https://doi.org/10.1038/sj.bjc.6602970.
203. Chaft JE, Arcila ME, Paik PK, Lau C, Riely GJ, Pietanza MC, Zakowski MF, Rusch V, Sima CS, Ladanyi M, Kris MG. Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling. Mol Cancer Ther. 2012;11(2):485–91. https://doi.org/10.1158/1535-7163.mct-11-0692.
204. Yamamoto H, Shigematsu H, Nomura M, Lockwood WW, Sato M, Okumura N, Soh J, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Date H, Lam WL, Minna JD, Gazdar AF. PIK3CA mutations and copy number gains in human lung cancers. Cancer Res. 2008;68(17):6913–21. https://doi.org/10.1158/0008-5472.can-07-5084.
205. Kawano O, Sasaki H, Endo K, Suzuki E, Haneda H, Yukiue H, Kobayashi Y, Yano M, Fujii Y. PIK3CA mutation status in Japanese lung cancer patients. Lung Cancer. 2006;54(2):209–15. https://doi.org/10.1016/j.lungcan.2006.07.006.
206. McGowan M, Hoven AS, Lund-Iversen M, Solberg S, Helland A, Hirsch FR, Brustugun OT. PIK3CA mutations as prognostic factor in squamous cell lung carcinoma. Lung Cancer. 2017;103:52–7. https://doi.org/10.1016/j.lungcan.2016.11.018.
207. Choi M, Kadara H, Zhang J, Parra ER, Rodriguez-Canales J, Gaffney SG, Zhao Z, Behrens C, Fujimoto J, Chow C, Kim K, Kalhor N, Moran C, Rimm D, Swisher S, Gibbons DL, Heymach J, Kaftan E, Townsend JP, Lynch TJ, Schlessinger J, Lee J, Lifton RP, Herbst RS, Wistuba II. Mutation profiles in early-stage lung squamous cell carcinoma with clinical follow-up and correlation with markers of immune function. Ann Oncol. 2017;28(1):83–9. https://doi.org/10.1093/annonc/mdw437.
208. Ji M, Guan H, Gao C, Shi B, Hou P. Highly frequent promoter methylation and PIK3CA amplification in non-small cell lung cancer (NSCLC). BMC Cancer. 2011;11:147. https://doi.org/10.1186/1471-2407-11-147.
209. Yip PY. Phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin (PI3K-Akt-mTOR) signaling pathway in non-small cell lung cancer. Transl Lung Cancer Res. 2015;4(2):165–76. https://doi.org/10.3978/j.issn.2218-6751.2015.01.04.
210. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, Colasacco C, Dacic S, Hirsch FR, Kerr K, Kwiatkowski DJ, Ladanyi M, Nowak JA, Sholl L, Temple-Smolkin R, Solomon B, Souter LH, Thunnissen E, Tsao MS, Ventura CB, Wynes MW, Yatabe Y. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. J Mol Diagn. 2018;20(2):129–59. https://doi.org/10.1016/j.jmoldx.2017.11.004.
211. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol. 2002;161(6):1961–71. https://doi.org/10.1016/s0002-9440(10)64472-0.
212. Dubeau L, Chandler LA, Gralow JR, Nichols PW, Jones PA. Southern blot analysis of DNA extracted from formalin-fixed pathology specimens. Cancer Res. 1986;46(6):2964–9.
213. Greer CE, Wheeler CM, Manos MM. Sample preparation and PCR amplification from paraffin-embedded tissues. PCR Methods Appl. 1994;3(6):S113–22.
214. Marchetti A, Felicioni L, Buttitta F. Assessing EGFR mutations. N Engl J Med. 2006;354(5):526–8.; author reply -8. https://doi.org/10.1056/NEJMc052564.

215. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. Oncotarget. 2012;3(5):546–58. https://doi.org/10.18632/oncotarget.503.
216. Guillou L, Coindre J, Gallagher G, Terrier P, Gebhard S, de Saint Aubain Somerhausen N, Michels J, Jundt G, Vince DR, Collin F, Trassard M, Le Doussal V, Benhattar J. Detection of the synovial sarcoma translocation t(X;18) (SYT;SSX) in paraffin-embedded tissues using reverse transcriptase-polymerase chain reaction: a reliable and powerful diagnostic tool for pathologists. A molecular analysis of 221 mesenchymal tumors fixed in different fixatives. Hum Pathol. 2001;32(1):105–12.
217. Zsikla V, Baumann M, Cathomas G. Effect of buffered formalin on amplification of DNA from paraffin wax embedded small biopsies using real-time PCR. J Clin Pathol. 2004;57(6):654–6.
218. Merkelbach S, Gehlen J, Handt S, Fuzesi L. Novel enzyme immunoassay and optimized DNA extraction for the detection of polymerase-chain-reaction-amplified viral DNA from paraffin-embedded tissue. Am J Pathol. 1997;150(5):1537–46.
219. de Franchis R, Cross NC, Foulkes NS, Cox TM. A potent inhibitor of Taq polymerase copurifies with human genomic DNA. Nucleic Acids Res. 1988;16(21):10355.
220. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J, Thunnissen E, Ladanyi M, C College of American Pathologists International Association for the Study of Lung, P Association for Molecular. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013;15(4):415–53. https://doi.org/10.1016/j.jmoldx.2013.03.001.
221. Heydt C, Fassunke J, Kunstlinger H, Ihle MA, Konig K, Heukamp LC, Schildhaus HU, Odenthal M, Buttner R, Merkelbach-Bruse S. Comparison of pre-analytical FFPE sample preparation methods and their impact on massively parallel sequencing in routine diagnostics. PLoS One. 2014;9(8):e104566. https://doi.org/10.1371/journal.pone.0104566.
222. Huijsmans CJ, Damen J, van der Linden JC, Savelkoul PH, Hermans MH. Comparative analysis of four methods to extract DNA from paraffin-embedded tissues: effect on downstream molecular applications. BMC Res Notes. 2010;3:239. https://doi.org/10.1186/1756-0500-3-239.
223. Kocjan BJ, Hosnjak L, Poljak M. Commercially available kits for manual and automatic extraction of nucleic acids from formalin-fixed, paraffin-embedded (FFPE) tissues. Acta Dermatovenerol Alp Pannonica Adriat. 2015;24(3):47–53.
224. Seiler C, Sharpe A, Barrett JC, Harrington EA, Jones EV, Marshall GB. Nucleic acid extraction from formalin-fixed paraffin-embedded cancer cell line samples: a trade off between quantity and quality? BMC Clin Pathol. 2016;16(1):17. https://doi.org/10.1186/s12907-016-0039-3.
225. Khokhar SK, Mitui M, Leos NK, Rogers BB, Park JY. Evaluation of Maxwell(R) 16 for automated DNA extraction from whole blood and formalin-fixed paraffin embedded (FFPE) tissue. Clin Chem Lab Med. 2012;50 https://doi.org/10.1515/cclm.2011.763.
226. Chen CW, Thomas CA Jr. Recovery of DNA segments from agarose gels. Anal Biochem. 1980;101(2):339–41.
227. Marko MA, Chipperfield R, Birnboim HC. A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. Anal Biochem. 1982;121(2):382–7.
228. Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J. Rapid and simple method for purification of nucleic acids. J Clin Microbiol. 1990;28(3):495–503.
229. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW. Comparison of DNA quantification methods for next generation sequencing. Sci Rep. 2016;6:24067. https://doi.org/10.1038/srep24067.
230. Ney JT, Froehner S, Roesler A, Buettner R, Merkelbach-Bruse S. High-resolution melting analysis as a sensitive prescreening diagnostic tool to detect KRAS, BRAF, PIK3CA, and AKT1 mutations in formalin-fixed, paraffin-embedded tissues. Arch Pathol Lab Med. 2012;136(9):983–92. https://doi.org/10.5858/arpa.2011-0176-OA.
231. Molina-Vila MA, Bertran-Alamillo J, Reguart N, Taron M, Castella E, Llatjos M, Costa C, Mayo C, Pradas A, Queralt C, Botia M, Perez-Cano M, Carrasco E, Tomas M, Mate JL, Moran T, Rosell R. A sensitive method for detecting EGFR mutations in non-small cell lung cancer samples with few tumor cells. J Thorac Oncol. 2008;3(11):1224–35. https://doi.org/10.1097/JTO.0b013e318189f579.
232. Hagemann IS, Devarakonda S, Lockwood CM, Spencer DH, Guebert K, Bredemeyer AJ, Al-Kateb H, Nguyen TT, Duncavage EJ, Cottrell CE, Kulkarni S, Nagarajan R, Seibert K, Baggstrom M, Waqar SN, Pfeifer JD, Morgensztern D, Govindan R. Clinical next-generation sequencing in patients with non-small cell lung cancer. Cancer. 2015;121(4):631–9. https://doi.org/10.1002/cncr.29089.
233. Konig K, Peifer M, Fassunke J, Ihle MA, Kunstlinger H, Heydt C, Stamm K, Ueckeroth F, Vollbrecht C, Bos M, Gardizi M, Scheffler M, Nogova L, Leenders F, Albus K, Meder L, Becker K, Florin A, Rommerscheidt-Fuss U, Altmuller J, Kloth M, Nurnberg P, Henkel T, Bikar SE, Sos ML, Geese WJ, Strauss L, Ko YD, Gerigk U, Odenthal M, Zander T, Wolf J, Merkelbach-Bruse S, Buettner R, Heukamp LC. Implementation of amplicon parallel sequencing leads to improvement of diagnosis and therapy of lung cancer patients. J Thorac Oncol. 2015;10(7):1049–57. https://doi.org/10.1097/jto.0000000000000570.
234. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, Sanford T, Buddavarapu K, Krosting J, Garmire L, Wylie D, Shinde R, Beaudenon S, Alexander EK, Mambo E, Adai AT, Latham GJ. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. J Mol Diagn. 2013;15(2):234–47. https://doi.org/10.1016/j.jmoldx.2012.11.006.
235. Schmitz K, Koeppen H, Binot E, Fassunke J, Kunstlinger H, Ihle MA, Heydt C, Wardelmann E, Buttner R, Merkelbach-Bruse S, Ruschoff J, Schildhaus HU. MET gene copy number alterations and expression of MET and hepatocyte growth factor are potential biomarkers in angiosarcomas and undifferentiated pleomorphic sarcomas. PLoS One. 2015;10(4):e0120079. https://doi.org/10.1371/journal.pone.0120079.
236. Abel HJ, Al-Kateb H, Cottrell CE, Bredemeyer AJ, Pritchard CC, Grossmann AH, Wallander ML, Pfeifer JD, Lockwood CM, Duncavage EJ. Detection of gene rearrangements in targeted clinical next-generation sequencing. J Mol Diagn. 2014;16(4):405–17. https://doi.org/10.1016/j.jmoldx.2014.03.006.
237. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn. 2015;17(3):251–64. https://doi.org/10.1016/j.jmoldx.2014.12.006.
238. Drilon A, Wang L, Arcila ME, Balasubramanian S, Greenbowe JR, Ross JS, Stephens P, Lipson D, Miller VA, Kris MG, Ladanyi M, Rizvi NA. Broad, hybrid capture-based next-generation sequencing identifies actionable genomic alterations in lung adenocarcinomas otherwise negative for such alterations by other genomic testing approaches. Clin Cancer Res. 2015;21(16):3631–9. https://doi.org/10.1158/1078-0432.ccr-14-2683.
239. Suh JH, Johnson A, Albacker L, Wang K, Chmielecki J, Frampton G, Gay L, Elvin JA, Vergilio JA, Ali S, Miller VA, Stephens PJ, Ross JS. Comprehensive genomic profiling facilitates implemen-

tation of the National Comprehensive Cancer Network Guidelines for lung cancer biomarker testing and identifies patients who may benefit from enrollment in mechanism-driven clinical trials. Oncologist. 2016;21(6):684–91. https://doi.org/10.1634/theoncologist.2016-0030.

240. Walther C, Hofvander J, Nilsson J, Magnusson L, Domanski HA, Gisselsson D, Tayebwa J, Doyle LA, Fletcher CD, Mertens F. Gene fusion detection in formalin-fixed paraffin-embedded benign fibrous histiocytomas using fluorescence in situ hybridization and RNA sequencing. Lab Investig. 2015;95(9):1071–6. https://doi.org/10.1038/labinvest.2015.83.
241. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80. https://doi.org/10.1038/nature03959.
242. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51. https://doi.org/10.1038/nrg.2016.49.
243. Koitzsch U, Heydt C, Attig H, Immerschitt I, Merkelbach-Bruse S, Fammartino A, RH Buttner Y, Kong MO. Use of the GeneReader NGS System in a clinical pathology laboratory: a comparative study. J Clin Pathol. 2017; https://doi.org/10.1136/jclinpath-2017-204342.
244. Thermo Fisher Scientific. Ion torrent next-generation sequencing run sequence [Internet]. 2017 [cited 2017 Jun 22]. Available from: https://www.thermofisher.com/de/de/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence.html
245. Illumina. Sequencing systems for every lab [Internet] 2017 [cited 2017 Jun 22]. Available from: https://www.illumina.com/systems.html
246. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348–52. https://doi.org/10.1038/nature10242.
247. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, K Walter XW, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53–9. https://doi.org/10.1038/nature07517.
248. Becker K, Vollbrecht C, Koitzsch U, Koenig K, Fassunke J, Huss S, Nuernberg P, Heukamp LC, Buettner R, Odenthal M, Altmueller J, Merkelbach-Bruse S. Deep ion sequencing of amplicon adapter ligated libraries: a novel tool in molecular diagnostics of formalin fixed and paraffin embedded tissues. J Clin Pathol. 2013;66(9):803–6. https://doi.org/10.1136/jclinpath-2013-201549.
249. Desai AN, Jere A. Next-generation sequencing: ready for the clinics? Clin Genet. 2012;81(6):503–10. https://doi.org/10.1111/j.1399-0004.2012.01865.x.
250. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, van Spaendonck-Zwarts KY, van Tintelen JP, Sijmons RH, Jongbloed JD, Sinke RJ. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat. 2013;34(7):1035–42. https://doi.org/10.1002/humu.22332.
251. Ahn S, Hong M, Van Vrancken M, Lyou YJ, Kim ST, Park SH, Kang WK, Park YS, Jung SH, Woo M, Lee J, Kim KM. A nCounter CNV assay to detect HER2 amplification: a correlation study with immunohistochemistry and in situ hybridization in advanced gastric cancer. Mol Diagn Ther. 2016;20(4):375–83. https://doi.org/10.1007/s40291-016-0205-4.
252. Lira ME, Kim TM, Huang D, Deng S, Koh Y, Jang B, Go H, Lee SH, Chung DH, Kim WH, Schoenmakers EF, Choi YL, Park K, Ahn JS, Sun JM, Ahn MJ, Kim DW, Mao M. Multiplexed gene expression and fusion transcript analysis to detect ALK fusions in lung cancer. J Mol Diagn. 2013;15(1):51–61. https://doi.org/10.1016/j.jmoldx.2012.08.006.
253. Veldman-Jones MH, Brant R, Rooney C, Geh C, Emery H, Harbron CG, Wappett M, Sharpe A, Dymond M, Barrett JC, Harrington EA, Marshall G. Evaluating robustness and sensitivity of the nanostring technologies ncounter platform to enable multiplexed gene expression analysis of clinical samples. Cancer Res. 2015;75(13):2587–93. https://doi.org/10.1158/0008-5472.can-15-0262.

Genomic Applications in Colorectal Carcinomas

28

Lauren L. Ritterhouse and Wade S. Samowitz

Introduction

Genomic testing in colorectal carcinoma has become an integral part of the diagnostic workup of these tumors, as it provides guidance for prognosis, therapy, and inherited susceptibility. The classic molecular categories of colorectal cancer pathogenesis include chromosomal instability (CIN), mismatch repair deficiency (dMMR), and the CpG island methylator phenotype (CIMP). The CIMP pathway is characterized by widespread promoter methylation and is frequently associated with dMMR secondary to *MLH1* promoter methylation. Currently, the main targets of molecular diagnostic testing in colorectal cancer include evaluations for dMMR and EGFR pathway alterations. The following chapter will focus on genomic applications that can be used for colorectal carcinoma, with a particular focus on dMMR assessment for both Lynch syndrome evaluation and therapeutic guidance, potential applications of circulating tumor DNA, and next-generation sequencing (NGS) applications for streamlining the genomic evaluation of colorectal cancers.

Mismatch Repair Deficiency (dMMR)

Evaluation for dMMR, either by microsatellite instability assessment or immunohistochemistry of mismatch repair proteins, has been one of the most commonly employed tests in the molecular diagnostics of colorectal carcinoma. Identifying the dMMR phenotype has many indications, including screening for the presence of Lynch syndrome, obtaining prognostic information, and, more recently, predicting response to immune checkpoint inhibitors.

L. L. Ritterhouse
Department of Pathology, University of Chicago, Chicago, IL, USA

W. S. Samowitz (✉)
Department of Pathology, University of Utah, Salt Lake City, UT, USA
e-mail: Wade.samowitz@aruplab.com

Lynch Syndrome Evaluation

The current model for identification of individuals with Lynch syndrome typically involves universal screening of colorectal and endometrial tumors for dMMR, differentiation of probable sporadic from possible inherited dMMR tumors, and finally evaluation of the germline for mismatch repair gene mutations associated with the possible inherited tumors. Screening for mismatch repair status is usually evaluated by immunohistochemical staining for mismatch repair proteins or by assessing microsatellite instability in the tumor, traditionally with a PCR-based test that interrogates a small panel of mononucleotide repeats (microsatellites) (Fig. 28.1). Probable sporadic mismatch repair-deficient tumors are identified by several characteristics: an immunohistochemical profile indicating loss of both MLH1 and PMS2, the presence of *BRAF* V600E mutations (in colorectal cancers), and *MLH1* promoter methylation (in any mismatch repair-deficient tumor). Acquired *MLH1* promoter methylation is the most common mechanism underlying sporadic mismatch repair-deficient tumors and is associated with the loss of both MLH1 and PMS2 by immunohistochemistry. Once a tumor is identified as mismatch repair-deficient, the IHC profile can be helpful in guiding the subsequent germline analysis, as specific protein losses are usually associated with mutations in specific genes [1]. Genomic evaluation has already rendered this feature of IHC somewhat obsolete, however, as typically all four mismatch repair genes, and many other cancer predisposition genes, are evaluated by NGS panel testing once the determination has been made that a mismatch repair-deficient tumor may be Lynch-associated.

As colorectal carcinomas are more routinely being tested by multigene targeted NGS panels, genomic techniques can also be used in the first step of screening for mismatch repair deficiency. Mismatch repair deficiency and subsequent microsatellite instability can be detected by several different strategies using a targeted NGS panel. As many microsatellites are already being captured as part of routine NGS targeted cancer panels, the data is often readily available and only needs to be analyzed correctly. Several groups have published bioinformatics algorithms to detect microsatellite

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_28

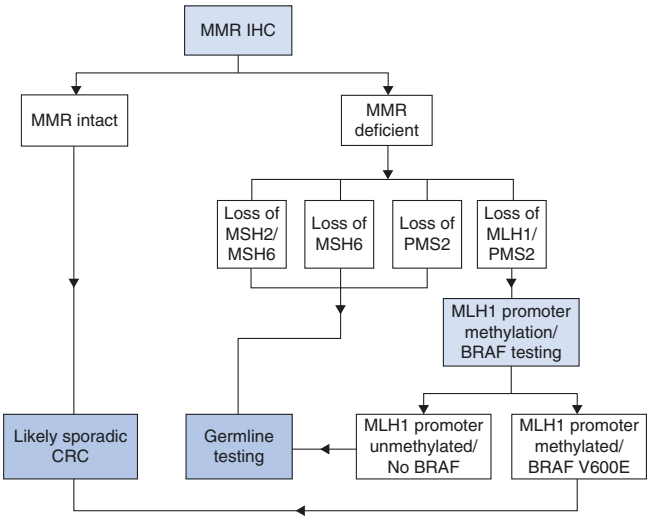


Fig. 28.1 Lynch screening algorithm based on the use of mismatch repair immunohistochemistry. MMR mismatch repair, IHC immunohistochemistry, CRC colorectal carcinoma

instability at numerous (ranging from 10s to 1000s depending on panel size) loci, using similar concepts that are employed in PCR-based microsatellite instability testing. While each of the strategies differs slightly, the basic premise involves evaluating the number and size of repeat length alleles at particular microsatellite loci, which can either be compared to a matched normal sample or to a previously determined large number of normal, microsatellite-stable samples [2–5]. Although several of these algorithms have been made publicly available, this strategy still requires proper bioinformatic implementation. Other groups have taken a more simplistic approach to identifying possible mismatch repair deficiency on NGS targeted panels that is determined by tumor mutational burden (TMB) and/or high frequency of insertion/deletion mutations occurring in repeat regions [6, 7]. TMB constitutes the somatic mutational rate present in a tumor, commonly denoted as either the number of mutations occurring per megabase (Mb) or the number of mutations present within an exome. However, TMB calculations have not been standardized, and care must be taken to evaluate how each individual lab is calculating and reporting this rate. Tumors with dMMR typically have a high TMB, and so this feature can be used to identify them. For example, in the Stadler et al. study, 100% of dMMR tumors (100% sensitivity) were detected by identifying cases that had a TMB above a certain threshold (20 mutations for a 341-gene panel). The three false positives that were identified were found to be the result of another hypermutated phenotype, one due to *POLE* mutations, which can be recognized by their exceptionally high mutation rate (>150 mutations in this panel), the presence of hotspot mutations in the *POLE* gene (P286R or V411 L), and their specific mutational signature (prevalence of C > A transversions) [7, 8].

The second step of determining whether a mismatch repair-deficient tumor is Lynch-associated or sporadic can also be evaluated by genomic techniques. The point mutation responsible for the *BRAF* V600E mutation is present on almost all targeted tumor NGS panels, and the presence of this mutation is a good indicator that a mismatch repair-deficient colorectal cancer is sporadic. A more obvious application, however, is the NGS evaluation of mismatch repair genes in the tumor and germline for mutations. This has the potential to directly identify Lynch syndrome in individuals with colorectal cancer, thus disrupting the stepwise approach shown in Fig. 28.1. With the increasing use of NGS evaluation of colorectal cancer (and its decreasing cost), it may be difficult to justify the current algorithm of sequential tissue tests followed by germline NGS.

In addition to identifying Lynch syndrome, NGS evaluation of germline and tumor can also identify tumors with acquired mismatch repair gene mutations. At one time, this was thought to be a very uncommon mechanism for the development of sporadic mismatch repair deficiency. Indeed, individuals with dMMR colorectal tumors without *MLH1* methylation and without detectable germline mismatch repair gene mutations were assumed to have Lynch syndrome in which the mutation was missed, perhaps for technical reasons. However, recent studies using NGS evaluation of the tumor uncovered acquired mismatch repair gene mutations in approximately 70% of MMR-deficient tumors without *MLH1* methylation, establishing this as an important mechanism for sporadic mismatch repair deficiency [9]. This is a substantial contribution of genomics, as identification of this subset of sporadic MMR-deficient tumors avoids the financial and psychological costs associated with the labeling and screening associated with an incorrect designation as "probable" Lynch for the individual and family members [9]. Such individuals have been termed "Lynch-like," a somewhat unfortunate choice of terminology given the non-inherited nature of the disease.

There are several challenges that must be addressed when using targeted NGS panels for the evaluation of genomic alterations involved in Lynch syndrome. First, many labs do not routinely sequence matched normal or germline samples, which can make distinguishing between somatic and germline alterations difficult, in addition to creating up to 15% false-positive variants when calculating TMB after filtering from germline population databases [10]. Second, many standard NGS panels designed for tumor sequencing may not cover all genomic regions required for a full Lynch syndrome evaluation, such as the 3′ end of *EPCAM*.

Due to the presence of at least 16 pseudogenes that share homology with *PMS2*, analyzing this gene for mutations can be technically difficult and may have led to the underreporting of Lynch syndrome patients with *PMS2* variants [11];

however, several techniques can be used to address this issue including modifying the target enrichment to include long-range PCR [12–14]. Other ways of identifying *PMS2* variants that can be applied to tumor FFPE tissues include the use of amplicon-based panels with carefully designed primers that utilize sequence differences between *PMS2* and its pseudogenes to avoid co-amplification and sequencing of the pseudogenes, as well as the isolation of RNA with subsequent RT-PCR and cDNA sequencing of the entire *PMS2* transcript that utilizes primers anchored within exon 10, an exon free of homology with pseudogenes, ensuring specific amplification of *PMS2* [15]. Success with germline Lynch syndrome evaluation by NGS has been demonstrated with >99% accuracy, including the detection of large copy number alterations and mutations in *PMS2* [16]. However, limitations still exist when examining tumor FFPE tissue by DNA, particularly at the 3′ end of the gene, as the pseudogene *PMS2CL* has strong similarity to *PMS2* exons 9 and 11–15, due to extensive gene conversion [17], and this is a potential limitation when detecting a "Lynch-like" tumor due to acquired *PMS2* mutations.

Additional considerations must be taken into account when approaching Lynch syndrome evaluation with broadly based tumor and germline NGS panels. A well-thought procedure must be in place for appropriate counseling and consent process beyond what is usually in place for typical somatic tumor NGS profiling test. There is the possibility of discovering "incidental" findings that may be related or unrelated to the patient's colorectal carcinoma, such as pathogenic or likely pathogenic variants in other genes associated with hereditary cancer predisposition such as *CDH1* or *BRCA2*.

Finally, genomic approaches will be necessary if we shift from a paradigm of identifying Lynch syndrome in those who already have a colorectal (or endometrial) cancer to one in which we identify Lynch in unaffected individuals. The current paradigm has its greatest impact on mutation-carrying relatives of affected individuals. Evaluation of young adults or newborns for mismatch repair gene mutations has the potential for decreasing cancer incidence in all gene carriers. If such an approach is taken, NGS approaches will be the most economical and efficient way to evaluate mismatch repair genes associated with Lynch syndrome [18].

MSI/dMMR for Stage II Colorectal Cancer

Adjuvant chemotherapy for stage II colon cancer offers only a small overall survival benefit and therefore is not uniformly recommended [19, 20]. Given that dMMR confers an improved prognosis in colorectal cancer patients, dMMR status has been proposed as a biomarker to identify patients that are unlikely to benefit from adjuvant chemotherapy [21, 22]. In support of this notion is a study showing that adjuvant chemotherapy with fluoropyrimidine plus oxaliplatin is not beneficial in stage II dMMR colorectal carcinoma, although it does appear to be beneficial for stage III dMMR tumors [23]. Some data suggests that the utility of dMMR biomarker in this setting is dependent on other factors, such as the location of the tumor and the sporadic vs. inherited nature of the dMMR [23, 24]. There is some evidence, however, that dMMR stage II tumors which are high risk, as defined by pathologic and clinical parameters, may benefit from oxaliplatin-based adjuvant chemotherapy [23]. Additional studies with larger numbers of patients will be necessary to definitively establish the utility of the dMMR phenotype as a biomarker for the treatment of stage II cancer.

MSI/dMMR and TMB for Immunotherapy

In addition to identifying patients of Lynch syndrome, colorectal cancer MSI status is now recognized as a biomarker to predict response to immune checkpoint inhibitor therapy [25, 26]. The sensitivity of dMMR tumors to immunotherapy is thought to be due to the stimulation of the immune system by numerous neoantigens arising from the high mutation burden of these tumors. In a landmark ruling, the FDA granted pembrolizumab accelerated approval as the first FDA-approved drug for a genetic biomarker, rather than for a specific tumor type. Pembrolizumab is approved for patients with unresectable or metastatic, MSI-H or dMMR solid tumors that have progressed following prior treatment and who have no satisfactory alternative treatment options, in addition to MSI-H or dMMR colorectal cancers that have progressed following treatment with a fluoropyrimidine, oxaliplatin, and irinotecan (FDA.gov).

In addition to dMMR tumors, colonic and extra-colonic tumors with a high TMB have also been shown to be responsive to immunotherapy [27–30]. The mechanism of response to immune checkpoint inhibitors is thought to be similar to that of dMMR tumors, in that a large number of mutations generate neoantigens which stimulate antitumor immune response [27, 28]. In fact, in colorectal cancer, some data suggest that high TMB, as can readily be determined by NGS, may be a superior (more comprehensive) biomarker of response to immune checkpoint therapy to dMMR status alone [31]. For example, colorectal cancers with hotspot *POLE* mutations (*POLE* P286R and V411L) have been shown to have a very high mutation rate and are potential candidates for immunotherapies [32].

EGFR Pathway

Signaling through the EGFR pathway plays an important role in the initiation and progression of colorectal cancer. Anti-EGFR monoclonal antibodies such as cetuximab and panitumumab are directed against the extracellular domain of the

EGFR tyrosine kinase to block ligand binding and disrupt downstream signaling of the RAS/RAF/MEK/ERK pathway [33–36]. Therefore, mutations in the downstream effectors of the EGFR signaling pathway are associated with resistance to EGFR inhibition. Activating mutations in *KRAS*, found in approximately 30–50% of colorectal cancers, are a well-established mechanism of resistance to EGFR-targeted therapy in colorectal cancer, as these mutations cause constitutive activation of downstream MAPK pathway [37–40]. *NRAS*, which is closely related to *KRAS* and mutated in approximately 3–5% of colorectal cancers, is also known to be associated with poor responses to EGFR therapy [41].

In addition to *KRAS* and *NRAS*, other genetic changes have been inconsistently associated with lack of response to EGFR therapies, including mutations in *BRAF* and *PIK3CA* [42–44]. Similar to *KRAS*-activating mutations, *BRAF* mutation promotes cancer growth through constitutive activation of the MAPK pathway in 5–10% of colorectal cancers. Studies of *PIK3CA* mutations have shown conflicting results with respect to EGFR therapy but may provide value in selecting patients for trials targeting the PI3K/AKT/mTOR pathway. Additionally, *PIK3CA* mutations have been associated with improved survival in patients who use aspirin following diagnosis [45, 46]. According to recent recommendation guidelines adopted in collaboration between ASCP, CAP, AMP, and ASCO, there is insufficient evidence to recommend *PIK3CA* or *PTEN* testing outside clinical trial setting. These same recommendation guidelines support the evaluation of "extended" *RAS* (which includes the evaluation of the following codons in *KRAS* and *NRAS* 12, 13, 59, 61, 117, and 146). Of interest, *BRAF* V600E testing is recommended for prognosis and Lynch syndrome workup, but not for decisions regarding EGFR-related therapy [47].

The list of genes and targets that should be evaluated prior to initiation of anti-EGFR therapy (and clinical trials of other therapies) in colorectal cancer will likely continue to expand. This will be increasingly difficult to deal with using single-gene assays; evaluation of extended *RAS* alone would typically require six separate assays. A comprehensive genomic strategy which at once evaluates all of extended *RAS* and numerous other genes in anticipation of the changing landscape of therapeutics and clinical trials is arguably more efficient and cost-effective than single-gene or single-exon assays.

ctDNA

With the advent of molecular techniques that are able to detect very low levels of tumor mutations (1% mutant allele fraction or better), various avenues for evaluating ctDNA in plasma have been explored. The use of ctDNA for monitoring disease has many advantages. Peripheral blood specimens are readily available without the need for an invasive biopsy to obtain a tumor sample. ctDNA has the potential for detecting genomic heterogeneity by representing a source of tumor DNA from multiple metastatic tumor sites. Molecular techniques that are able to achieve the high levels of sensitivity required for ctDNA detection include PCR-based assays as well as next-generation sequencing technologies utilizing molecular barcodes. Advantages of using a more broad-based approach such as NGS include the ability to detect multiple mutation targets at once, negating the need for multiple single amplicon-based tests or patient and mutation-specific design.

Potential applications of ctDNA testing in colorectal carcinoma include monitoring for the development of genetic alterations of therapy resistance and the evaluation of minimal residual disease. Several studies have used ctDNA to assess for the development of resistance to EGFR-targeted therapy. Low-frequency *KRAS* alterations have been identified in the plasma of patients with metastatic colorectal cancer undergoing anti-EGFR therapy as a potential mechanism of resistance [48–51]. Additional mechanisms of resistance have also been detected in ctDNA of patients who have become resistant to cetuximab or panitumumab therapy, which include *HER2*, *MET*, and *ERBB2* amplification, as well as mutations in *NRAS*, *EGFR*, and *MAP2K1*. For example, mutant *RAS* clones have been shown to rise in the peripheral blood during EGFR blockade and subsequently decline upon withdrawal of anti-EGFR antibodies [52, 53].

Evaluation of ctDNA for assessment of minimal residual disease following surgery can potentially stratify patients, based on outcome, to help guide the need for adjuvant therapy. In one study, approximately 80% of stage II colon cancer patients in whom ctDNA was detected postoperatively had a recurrence compared to only 10% of patients in whom ctDNA was undetectable [54]. Therefore, ctDNA following surgery could be a potential biomarker to stratify patients with stage II colon cancer to determine who is at a high risk for recurrence and could possibly benefit from adjuvant therapy. In another study, serial analysis of ctDNA in colorectal cancer patients following therapy was able to detect recurrence on average 9 months earlier than CT scans [55], therefore offering the potential to augment or replace imaging for follow-up.

Utility of Universal NGS Testing at Diagnosis: Upfront NGS

In the era of increasing number of tests required on single pathology specimens (dMMR, tumor mutation burden, predictors of EGFR therapy response, and biomarkers for novel clinical trials), the possibility of streamlining genomic analyses into a single testing modality is very attractive. This has the potential to preserve valuable tissue, reduce turnaround times,

and reduce overall costs. If a single, comprehensive NGS-based panel test was to be performed on every colorectal carcinoma at initial resection (referred to here as "upfront NGS"), this would allow for the efficient detection of several key biomarkers necessary to achieve a highly personalized therapy. First, this comprehensive testing strategy would cover both the extended *RAS* requirements and the numerous potential targets in ongoing and future clinical trials. Second, Lynch syndrome evaluation could be performed as described above, which would be facilitated by the sequencing of matched normal or germline tissues. Third, the identification of a dMMR tumor would be feasible to help in prognostication and potential stratification of stage II patients, as well as identify patients that are likely to be responsive to immune checkpoint inhibitors. Tumor mutation burden, another biomarker associated with response to immunotherapy, would also be obtainable. Finally, the knowledge of the comprehensive genomics of a colorectal tumor would allow for the successful monitoring of patients with ctDNA during and following their treatment.

Besides providing immediate results for decision-making at the time of initial surgical resection, results from upfront NGS testing would also be readily available should the need for additional information arise in the future. For example, if EGFR therapy is contemplated a year after resection because of a new hepatic metastasis, it would be extremely helpful to have the extended *RAS* mutation status at hand. Compared to this, the current paradigm of searching for the relevant block and performing and interpreting the test when needed is time-consuming and inefficient.

A disadvantage of the upfront testing approach is that it does not take into account the future development of tumor clones with additional mutations which are not present in the original tumor. For example, if the hepatic metastasis in the above example harbored a new *NRAS* mutation, then this would not be detected by the upfront analysis of the primary tumor. One way to address this is to combine upfront testing to see the original mutation spectrum of the tumor with ctDNA evaluation to detect any clones with different mutational profiles.

Summary

Genomic testing is a major disruptive force in the molecular diagnostics of colorectal cancer. Detection of dMMR deficiency, tumor mutational burden, Lynch syndrome, and biomarkers for EGFR therapy response and other therapeutic targets can be efficiently and cost-effectively achieved by next-generation sequencing. This will likely replace the traditional Lynch syndrome algorithmic tissue workup and the panoply of single-gene/single-exon assays commonly used today. In the future, "upfront NGS" on all colorectal cancers (and germline) at time of diagnosis or resection could provide a mutational base template for subsequent ctDNA testing (Fig. 28.2).

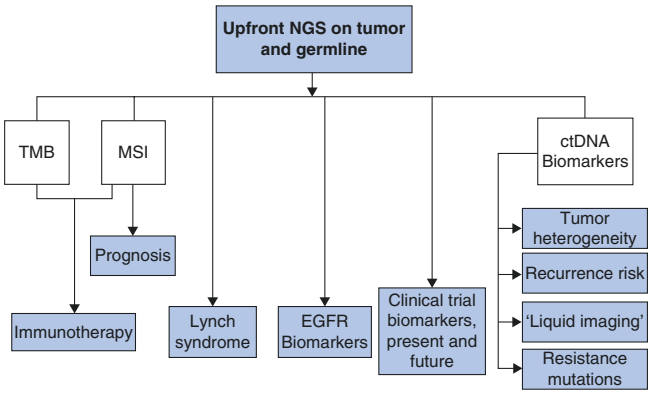


Fig. 28.2 The utility of upfront NGS testing on all colorectal carcinomas. TMB tumor mutational burden, MSI microsatellite instability, ctDNA circulating tumor DNA

References

1. Furtado LV, Samowitz WS. Colorectal cancer molecular profiling: from IHC to NGS in search of optimal algorithm. Virchows Arch. 2017;471:235.
2. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014;30(7):1015–6.
3. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. Clin Chem. 2014;60(9):1192–9.
4. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. Oncotarget. 2017;8(5):7452–63.
5. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. Nat Med. 2016;22(11):1342–50.
6. Nowak JA, Yurgelun MB, Bruce JL, Rojas-Rudilla V, Hall DL, Shivdasani P, et al. Detection of mismatch repair deficiency and microsatellite instability in colorectal adenocarcinoma by targeted next-generation sequencing. J Mol Diagn. 2017;19(1):84–91.
7. Stadler ZK, Battaglin F, Middha S, Hechtman JF, Tran C, Cercek A, et al. Reliable detection of mismatch repair deficiency in colorectal cancers using mutational load in next-generation sequencing panels. J Clin Oncol. 2016;34(18):2141–7.
8. Church DN, Briggs SE, Palles C, Domingo E, Kearsey SJ, Grimes JM, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. Hum Mol Genet. 2013;22(14):2820–8.
9. Haraldsdottir S, Hampel H, Tomsic J, Frankel WL, Pearlman R, de la Chapelle A, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology. 2014;147(6):1308–16. e1
10. Garofalo A, Sholl L, Reardon B, Taylor-Weiner A, Amin-Mansour A, Miao D, et al. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. Genome Med. 2016;8(1):79.
11. van der Klift HM, Mensenkamp AR, Drost M, Bik EC, Vos YJ, Gille HJ, et al. Comprehensive mutation analysis of PMS2 in a large cohort of Probands suspected of Lynch syndrome or constitutional mismatch repair deficiency syndrome. Hum Mutat. 2016;37(11):1162–79.
12. Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, et al. Identification of a variety of mutations in cancer

predisposition genes in patients with suspected Lynch syndrome. Gastroenterology. 2015;149(3):604–13. e20
13. Clendenning M, Hampel H, LaJeunesse J, Lindblom A, Lockman J, Nilbert M, et al. Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat. 2006;27(5):490–5.
14. Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, et al. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum Mutat. 2010;31(5):588–93.
15. Hansen MF, Neckmann U, Lavik LA, Vold T, Gilde B, Toft RK, et al. A massive parallel sequencing workflow for diagnostic genetic testing of mismatch repair genes. Mol Genet Genomic Med. 2014;2(2):186–200.
16. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn. 2012;14(4):357–66.
17. Vaughn CP, Hart KJ, Samowitz WS, Swensen JJ. Avoidance of pseudogene interference in the detection of 3′ deletions in PMS2. Hum Mutat. 2011;32(9):1063–71.
18. Hampel H, de la Chapelle A. How do we approach the goal of identifying everybody with Lynch syndrome? Familial Cancer. 2013;12(2):313–7.
19. Pahlman LA, Hohenberger WM, Matzel K, Sugihara K, Quirke P, Glimelius B. Should the benefit of adjuvant chemotherapy in colon cancer be re-evaluated? J Clin Oncol. 2016;34(12):1297–9.
20. Yothers G, O'Connell MJ, Allegra CJ, Kuebler JP, Colangelo LH, Petrelli NJ, et al. Oxaliplatin as adjuvant therapy for colon cancer: updated results of NSABP C-07 trial, including survival and subset analyses. J Clin Oncol. 2011;29(28):3768–74.
21. Lanza G, Gafa R, Santini A, Maestri I, Guerzoni L, Cavazzini L. Immunohistochemical test for MLH1 and MSH2 expression predicts clinical outcome in stage II and III colorectal cancer patients. J Clin Oncol. 2006;24(15):2359–67.
22. Grant M, Haydon A, Au L, Wilkins S, Oliva K, Segelov E, et al. Immunohistochemistry testing for mismatch repair deficiency in Stage 2 colon cancer: a cohort study of two cancer centres. Int J Surg. 2018;51:71–5.
23. Tougeron D, Mouillet G, Trouilloud I, Lecomte T, Coriat R, Aparicio T, et al. Efficacy of adjuvant chemotherapy in colon cancer with microsatellite instability: a large multicenter AGEO Study. J Natl Cancer Inst. 2016;108(7):1–9.
24. Yang L, He W, Yang Q, Kong P, Xie Q, Jiang C, et al. Combination of primary tumor location and mismatch repair status guides adjuvant chemotherapy in stage II colon cancer. Oncotarget. 2017;8(58):99136–49.
25. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015;372(26):2509–20.
26. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357(6349):409–13.
27. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371(23):2189–99.
28. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348(6230):124–8.
29. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther. 2017;16(11):2598–608.
30. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. 2017;9(1):34.
31. George TJ, Frampton GM, Sun J, Gowen K, Kennedy M, Greenbowe JR, et al. Tumor mutational burden as a potential biomarker for PD1/PD-L1 therapy in colorectal cancer. J Clin Oncol. 2016;34(15_suppl):3587.
32. Gong J, Wang C, Lee PP, Chu P, Fakih M. Response to PD-1 blockade in microsatellite stable metastatic colorectal cancer harboring a POLE mutation. J Natl Compr Cancer Netw. 2017;15(2):142–7.
33. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med. 2004;351(4):337–45.
34. Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyns B, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. J Clin Oncol. 2007;25(13):1658–64.
35. Heinemann V, von Weikersthal LF, Decker T, Kiani A, Vehling-Kaiser U, Al-Batran SE, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. Lancet Oncol. 2014;15(10):1065–75.
36. Schwartzberg LS, Rivera F, Karthaus M, Fasola G, Canon JL, Hecht JR, et al. PEAK: a randomized, multicenter phase II study of panitumumab plus modified fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) or bevacizumab plus mFOLFOX6 in patients with previously untreated, unresectable, wild-type KRAS exon 2 metastatic colorectal cancer. J Clin Oncol. 2014;32(21):2240–7.
37. Bokemeyer C, Bondarenko I, Makhson A, Hartmann JT, Aparicio J, de Braud F, et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. J Clin Oncol. 2009;27(5):663–71.
38. Tol J, Koopman M, Cats A, Rodenburg CJ, Creemers GJ, Schrama JG, et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. N Engl J Med. 2009;360(6):563–72.
39. Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. J Clin Oncol. 2010;28(31):4697–705.
40. Peeters M, Price TJ, Cervantes A, Sobrero AF, Ducreux M, Hotko Y, et al. Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. J Clin Oncol. 2010;28(31):4706–13.
41. Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. N Engl J Med. 2013;369(11):1023–34.
42. Di Nicolantonio F, Martini M, Molinari F, Sartore-Bianchi A, Arena S, Saletti P, et al. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. J Clin Oncol. 2008;26(35):5705–12.
43. Laurent-Puig P, Cayre A, Manceau G, Buc E, Bachet JB, Lecomte T, et al. Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. J Clin Oncol. 2009;27(35):5924–30.
44. Sartore-Bianchi A, Martini M, Molinari F, Veronese S, Nichelatti M, Artale S, et al. PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. Cancer Res. 2009;69(5):1851–7.
45. Liao X, Lochhead P, Nishihara R, Morikawa T, Kuchiba A, Yamauchi M, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. N Engl J Med. 2012;367(17):1596–606.

46. Gu M, Nishihara R, Chen Y, Li W, Shi Y, Masugi Y, et al. Aspirin exerts high anti-cancer activity in PIK3CA-mutant colon cancer cells. Oncotarget. 2017;8(50):87379–89.
47. Sepulveda AR, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK, et al. Molecular Biomarkers for the Evaluation of Colorectal Cancer: guideline from the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. J Mol Diagn. 2017;19(2):187–225.
48. Morelli MP, Overman MJ, Dasari A, Kazmi SM, Mazard T, Vilar E, et al. Characterizing the patterns of clonal selection in circulating tumor DNA from patients with colorectal cancer refractory to anti-EGFR treatment. Ann Oncol. 2015;26(4):731–6.
49. Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature. 2012;486(7404):537–40.
50. Taly V, Pekin D, Benhaim L, Kotsopoulos SK, Le Corre D, Li X, et al. Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. Clin Chem. 2013;59(12):1722–31.
51. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
52. Takegawa N, Yonesaka K, Sakai K, Ueda H, Watanabe S, Nonagase Y, et al. HER2 genomic amplification in circulating tumor DNA from patients with cetuximab-resistant colorectal cancer. Oncotarget. 2016;7(3):3453–60.
53. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med. 2015;21(7):795–801.
54. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med. 2016;8(346):346ra92.
55. Scholer LV, Reinert T, Orntoft MW, Kassentoft CG, Arnadottir SS, Vang S, et al. Clinical implications of monitoring circulating tumor DNA in patients with colorectal cancer. Clin Cancer Res. 2017;23(18):5437–45.

Genomic Applications in Pancreatic and Gastric Tumors

29

Fátima Carneiro and Ralph H. Hruban

Introduction

As beautifully articulated in this book, cancer is fundamentally a genetic disease [112]. The origins of neoplasia, the drivers of growth and heterogeneity of established cancers, and the response of certain tumors to targeted therapy can best be understood through the lens of genetics. An understanding of the central importance of genetics in oncogenesis was the basis for the first cancer exome sequencing efforts led by Bert Vogelstein and his colleagues [53, 124]. Vogelstein and colleagues showed that cancer exomes can be sequenced and that sequencing can uncover new oncogenic pathways and new therapeutic targets. Based on these early successes, several large-scale efforts to sequence all major cancer types were launched, including The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) and the International Cancer Genome Consortium (ICGC, http://icgc.org/). Through these efforts the exomes and genomes of all of the major cancer types have now been sequenced. While the initial hope was that major therapeutic targets would be identified in each of the cancers sequenced, the reality is much more complex. Pathologists are in a unique position to unlock and harness this complexity.

Pancreatic Neoplasms

In the decades after the *KRAS* oncogene was found to be a major driver of ductal adenocarcinoma of the pancreas (pancreatic cancer), researchers painstakingly identified, one by one, the major tumor suppressor genes targeted in pancreatic cancer [45, 101]. *TP53* and *p16/CDKN2A* were first discovered in other tumor types and then shown to be inactivated in a significant fraction of pancreatic cancers, while *SMAD4* was discovered by Scott Kern in pancreatic cancer (Table 29.1) [18, 40, 93]. More recently, whole-exome and whole-genome sequencing efforts have been applied to all of the major tumor types in the pancreas (Tables 29.1 and 29.2)

Table 29.1 Genes somatically targeted in ductal adenocarcinomas of the pancreas

Gene	Percentage of cases
KRAS	90–95
P16/CDKN2A	90
TP53	75
SMAD4	55
ARID1A, *BRAF*, *ERBB2*, *GATA 6*, *GNAS*, *MET*, *MLL3*, *MYC*, *PBRM1*,*RNF43*, *ROBO2*, *SF3B1*, *SLIT2*, *SMARCA2*, *SMARCA4*, *TGFβR2*,	Each <20

Table 29.2 Genes somatically targeted in other neoplasms of the pancreas

Tumor type	Genes
Acinar carcinoma	*SMAD4*, *TP53*, *JAK1*, *BRAF* (including fusions), *RB1*, *RAF1* (including fusions), *BRCA2*, *BRCA1*, *and ATM* (mostly germline), *CDKN2A*, *CTNNB1, APC, PRKAR1A, RB1*
Intraductal papillary mucinous neoplasm	*KRAS*, *GNAS*, *RNF43*, *TP53*, *SMAD4*, *CDKN2A*, *PIK3CA*
Mucinous cystic neoplasm	*KRAS*, *RNF43*, *TP53*, *CDKN2A*, *PIK3CA*
Pancreatic neuroendocrine tumor	*MEN1*, DAXX, *ATRX*, *TSC2*, *PTEN*, *PIK3CA*
Pancreatic neuroendocrine carcinoma	*TP53*, *RB1*
Serous cystic neoplasm	*VHL*
Solid pseudopapillary neoplasm	*CTNNB1* (beta-catenin)

F. Carneiro
Ipatimup/i3S, Faculty of Medicine of the University of Porto and Centro Hospitalar São João, Porto, Portugal

R. H. Hruban (✉)
The Sol Goldman Pancreatic Cancer Research Center, Departments of Pathology and Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
e-mail: rhruban@jhmi.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_29

[7, 11, 50–53, 95, 115, 121, 126, 127] [add TCGA reference here when available]. The sequencing of ductal adenocarcinomas confirmed the four previously discovered "mountains" (*KRAS*, *TP53*, *p16/CDKN2A*, and *SMAD4*) and revealed a long tail of less frequently mutated genes (Table 29.1). The sequencing of other tumor types in the pancreas, such as pancreatic neuroendocrine tumors, revealed new pathways and novel therapeutic targets [51].

When the genetic alterations discovered in various types of tumors of the pancreas are compared with tumor histopathology, there is a remarkable congruence between genomics and histopathology [123]. An integrated histologic-genetic classification therefore does not discard the well-tested morphologic classification system that has been developed and tested over many decades by expert pathologists; instead, this classification system embraces the old and the new and integrates morphology and molecular findings into a cohesive system with prognostic and therapeutic implications. Here we present a few examples that highlight the value of integrating molecular findings with the existing morphologic classification of neoplasms of the pancreas.

Ductal Adenocarcinoma of the Pancreas

Infiltrating ductal adenocarcinomas of the pancreas ("pancreatic cancers") are genetically remarkably homogeneous. The *KRAS* oncogene is activated by point mutation and/or amplification in virtually all of these cancers (95%) [11, 45, 53]. In addition, as noted earlier, *TP53*, *p16/CDKN2A*, and *SMAD4* are each inactivated in >50% of ductal adenocarcinomas [11, 53, 93, 121] [add TCGA reference here when available]. Remarkably, of all of the genes targeted in ductal adenocarcinomas, only *SMAD4* and the chromatin remodeling genes *ARID1A* and *MLL*, *MLL2*, and *MLL3* have been consistently shown to have prognostic significance [13, 98, 121]. *SMAD4* loss also appears to influence the patterns of spread of the disease as ductal adenocarcinomas with *SMAD4* loss are more likely to metastasize widely than are ductal adenocarcinomas with intact *SMAD4* [47].

The genetic alterations in infiltrating ductal adenocarcinomas also provide insight into the causes of the disease. For example, cigarette smoking is estimated to cause 20% of pancreatic cancers, and pancreatic cancers that arise in smokers have more somatic mutations than do pancreatic cancers in non-smokers [12]. A strong family history of pancreatic cancer increases risk, and a number of familial pancreatic cancer genes have been discovered (Table 29.3). These genes include *BRCA1*, *BRCA2*, *PALB2*, *p16/CDKN2A*, *ATM*, *STK11*, and the DNA mismatch repair genes (*hMLH1*, *hMSH2*, etc. [52, 94, 95, 134]). These genes are important to recognize because not only are germline alterations in these genes associated with an increased risk of pancreatic and extrapancreatic neoplasms, but some also are targetable. For example, some infiltrating ductal adenocarcinomas with biallelic inactivation of *BRCA2* are exquisitely sensitive to PARP (poly ADP-ribose polymerase) inhibitors [79]. The DNA mismatch repair genes will be described in greater detail below in the section on medullary carcinomas.

Table 29.3 Genes targeted in the germline that predispose to pancreatic cancer

Gene
ATM
BRCA1
BRCA2
CDKN2A
MLH1, *MSH2*, *MSH6*, *PMS2*
PALB2
PRSS1
STK11
TP53

We should note that there have also been attempts to classify infiltrating ductal adenocarcinomas based solely on gene expression patterns. Collisson and colleagues described three subtypes (classical, quasi-mesenchymal, and exocrine-like) and Moffitt and colleagues two subtypes (basal-like which closely matches quasi-mesenchymal) and classical (which closely matches Collison's classical subtype), while Bailey et al. reported four subtypes (squamous, pancreatic progenitor, immunogenic, and aberrantly differentiated endocrine/exocrine) [7, 23, 85] [add TCGA reference here when available]. While some of these subtypes, such as the squamous subtype, have a known histologic correlate, others do not. One concern of any classification system that relies on gene expression and ignores histopathology is that some subtypes may be confounded by nonneoplastic components of the tumor. Pancreatic cancers are, in particular, associated with low neoplastic cellularity, and gene expression patterns, even from microdissected tumors, can reflect gene expression from nonneoplastic stromal and contaminating acinar cells. Further validation including more careful correlation of gene expression with cell type is needed before these gene expression subtypes can be accepted.

Although not the focus of this review, protein expression is obviously also extremely important (see http://www.pancreaticcancerdatabase.org). Targeted therapies can take advantage of the unique patterns of protein expression in ductal adenocarcinoma of the pancreas. For example, pancreatic cancers elicit an intense desmoplastic reaction, and albumin-bound paclitaxel (nab-paclitaxel), which is believed to bind to SPARC in this desmoplastic stroma helping to localize the antineoplastic agent to the tumor, has been shown to be effective in treating patients with pancreatic cancer [48, 113].

Pancreatic Intraepithelial Neoplasia

One of the greatest potential benefits of our improved understanding of the genetic drivers of pancreatic neoplasia is the opportunity it presents to develop new tools for the early detection of pancreatic cancer. Pathologists are in a unique position to aid these efforts, because pathologists have unique insight into the earliest forms of pancreatic neoplasia.

There are currently three types of precursor lesions. Intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNs) are macroscopic neoplasms that form cysts, and IPMNs and MCNs will be discussed later in the section on cystic neoplasms of the pancreas. Pancreatic intraepithelial neoplasia (PanIN) is a microscopic lesion [10]. Although the data are obviously incomplete because PanINs are not well visualized with currently available imaging technologies, histologic, genetic, and clinical findings all suggest that most infiltrating ductal adenocarcinomas arise from PanIN lesions [10]. PanINs have been reported to have many of the same genetic alterations as are found in infiltrating ductal adenocarcinomas, including activation of *KRAS* and inactivation of *p16*/*CDKN2A*, *TP53*, and *SMAD4* [10]. An understanding of this timing of genetic alterations is critical, because PanINs are common in the aging population, and the early genetic events may be too common in the population to be useful for gene-based screening efforts. For example, activating point mutations in the *KRAS* gene may be too prevalent to be useful to form the basis of an early detection test [71]. Attempts to define the timing of genetic alterations in PanINs have been confounded by the propensity of invasive pancreatic cancer to invade the nonneoplastic duct system and then grow within the nonneoplastic ducts, thereby mimicking PanINs [10]. In order to overcome this, Laura Wood and colleagues recently sequenced the exomes of PanIN-3 lesions that occurred in pancreata without an invasive carcinoma [43]. In so doing, they were able to show that *KRAS* and *p16*/*CDKN2A* alterations are common but that *SMAD4* and *TP53* inactivation is rare in PanINs and instead is more suggestive of an invasive carcinoma.

With the understanding that *KRAS* is a common and early mutation, while *SMAD4* and *TP53* are late and significant events, Goggins and colleagues sequenced pancreatic juice samples collected at the time of endoscopy [56, 57, 133]. They showed that mutations in key cancer drivers, including *TP53*, can be detected in endoscopically obtained pancreatic secretions [56, 57, 133]. Of note, in some patients these mutations were detected >1 year before their cancer was detectable on imaging [56, 57, 133].

Variants of Ductal Adenocarcinoma of the Pancreas

Adenosquamous Carcinoma

Adenosquamous carcinomas are malignant epithelial neoplasms of the pancreas with significant components of both glandular and squamous differentiation [44]. By definition, the squamous component should comprise at least 30% of the neoplasm [44]. Adenosquamous carcinomas are rare, accounting for only 2–4% of the malignancies of the exocrine pancreas. Adenosquamous carcinomas are important to recognize for two reasons. First, they are associated with a very poor prognosis. The median survival for patients with an adenosquamous carcinoma is only 6 months [15, 114]. Second, some adenosquamous carcinomas respond to therapy that includes a platinum-based agent [114]. Inclusion of a platinum agent in the adjuvant regimen improves survival (HR = 2.4; 95% CI, 1.0–5.8; $P = 0.04$) [114].

Similar to ductal adenocarcinomas, the *KRAS*, *SMAD4*, and *TP53* genes are frequently targeted in adenosquamous carcinomas [17]. However, Liu and colleagues recently reported that unlike ductal adenocarcinomas, 80% of adenosquamous carcinomas harbor inactivating mutations in the *UPF1* gene [76]. The mechanism by which *UPF1* inactivation promotes tumor growth is not well understood, but the *UPF1* gene codes for a component in the nonsense-mediated RNA decay pathway.

The squamous component of adenosquamous carcinoma overexpresses markers of squamous differentiation including p63, and as noted above, one of the gene expression subtypes described by Bailey and colleagues is the squamous subtype which has some overlap with the quasi-mesenchymal subtype of Collison et al. [7, 23].

Colloid Carcinoma

Colloid carcinomas (mucinous non-cystic adenocarcinomas) are infiltrating gland-forming epithelial malignancies characterized by abundant mucin production by the neoplastic cells which are suspended ("floating") in large pools of extracellular mucin [44]. By definition, the colloid component should comprise at least 80% of the neoplasm [44]. Again, these are rare neoplasms, comprising only 1–3% of the malignancies of the exocrine pancreas. Colloid carcinomas almost always arise in association with an intestinal-type IPMN [99]. Colloid carcinomas are important to recognize because they are associated with an excellent prognosis with 5-year survival rates approaching 60% [2].

Since almost all colloid carcinomas arise from IPMNs, we can glean a lot about the genetic alterations in colloid carcinomas by looking at those in IPMNs [127]. Scarpa and colleagues conducted targeted sequencing of >50 genes in a series of 48 intraductal papillary neoplasms and found *GNAS* and/or *KRAS*

mutations in 44/48 (92%) of the IPMNs [6]. *RNF43* was the third most commonly mutated gene. *TP53* and *SMAD4* inactivation was limited to IPMNs with high-grade dysplasia. As predicted, matched invasive colloid carcinomas had the same alterations as their associated IPMNs. Gloria Su and colleagues reported that *PIK3CA* is mutated in 10% of IPMNs [34].

Hepatoid Carcinoma

Hepatoid carcinomas are malignant epithelial neoplasms with significant hepatocellular differentiation [44]. Only a handful of these rare carcinomas have been reported, too few to draw any conclusions on the genetic alterations that drive these neoplasms.

Medullary Carcinoma

Medullary carcinomas of the pancreas, although rare, are a great example of the power of genetic-histologic correlations. Medullary carcinomas are characterized by poor differentiation, pushing borders, and a syncytial growth pattern [9, 35, 44, 119]. As is true for medullary carcinomas of the colon, medullary carcinomas of the pancreas often have prominent intratumoral lymphocytes. Despite their poor differentiation, medullary carcinomas are associated with a significantly better prognosis than ductal adenocarcinomas [9, 35, 44, 86, 119].

Many, but not all, carcinomas with medullary histology have microsatellite instability (MSI), and not surprisingly one of the DNA mismatch repair genes (*hMLH1* or *hMSH2*) is often inactivated in medullary cancers. While medullary carcinomas of pancreas are usually *KRAS* wild-type, 33% are associated with *BRAF* mutations and a handful with Epstein-Barr virus [35, 46, 97, 119]. Because microsatellite instability will lead to the generation of hundreds of additional mutations, MSI-high pancreatic cancers appear to be particularly sensitive to immunotherapy [66].

Signet Ring Carcinoma

Signet ring carcinomas are composed of round non-cohesive cells which contain prominent intracytoplasmic mucin [44]. Since signet ring carcinomas can arise in the stomach and lobular carcinoma in the breast, a gastric or mammary primary needs to be excluded before establishing the diagnosis of a primary in the pancreas. Little is known of the genetics of these rare neoplasms, but a case of signet ring cell carcinoma with mismatch repair deficiency has been reported [46].

Undifferentiated Carcinoma

Undifferentiated carcinomas of the pancreas range from neoplasms composed of pleomorphic mononuclear cells to those composed of relatively monomorphic spindle cells [44]. At the molecular level, loss of E-cadherin expression is seen in 95% of undifferentiated carcinomas [120]. As a result, the neoplastic cells of anaplastic carcinomas are poorly cohesive, very infiltrative, and associated with an extremely poor prognosis (mean survival of 5.2 months after diagnosis) [120].

Undifferentiated Carcinoma with Osteoclast-Like Giant Cells

Undifferentiated carcinomas with osteoclast-like giant cells (UCOCGCs) composed of large benign appearing multinucleated giant cells admixed with atypical neoplastic mononuclear epithelial cells [44]. The atypical mononuclear cells variably express markers of epithelial differentiation, and the osteoclast-like giant cells express markers of histiocytic differentiation. Molecular analyses demonstrate that the osteoclast-like giant cells are reactive [118]. A case of an UCOCGC with microsatellite instability has been reported, as has a patient with the familial atypical multiple mole melanoma (FAMMM) syndrome [62, 88].

UCOCGCs are highly aggressive carcinomas with a mean survival of only 12 months [44].

Thus, molecular analyses have helped separate carcinomas of the pancreas into distinct groups with discrete biologies, separate prognoses, and different susceptibilities to various therapies. This molecular classification has clinical implications.

Acinar Cell Carcinoma

Acinar cell carcinomas are rare epithelial neoplasms of the pancreas defined by the significant production of exocrine enzymes [44]. The genetic alterations present in acinar cell carcinomas differ significantly from those of ductal adenocarcinomas. First, on average, acinar cell carcinomas have more mutations per cancer than do ductal adenocarcinomas [50]. In a whole-exome sequencing study reported by Wood and colleagues, acinar cell carcinomas harbored an average of 119 somatic mutations per carcinoma as compared to 63 reported in ductal adenocarcinomas [50, 53]. Second, in contrast to ductal adenocarcinomas, *KRAS* is not targeted in acinar carcinomas [50]. Third, Klimstra and colleagues reported that *RAF* fusions are common in acinar cell carcinomas [22]. Finally, one-third of acinar cell carcinomas have potentially targetable genetic alterations, including mutations in DNA repair *BRCA2*, *PALB2*, *ATM*, *BAP1*, *BRAF*, and *JAK1* [22, 50, 77].

Less common genetic alterations in acinar cell carcinomas include mutations in *SMAD4* (25%), *TP53* (15%), *GNAS* (10%), *RNF43* (5%), and *MEN1* (5%) [50].

Cystic Neoplasms

In the case of cystic neoplasms of the pancreas, the molecular findings beautifully complement the existing morphologic classification system [1, 126, 127]. There are four main types

of neoplastic cysts of the pancreas: intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm (MCN), serous cystic neoplasm, and solid pseudopapillary neoplasm (SPN). Each of these neoplasms has been well-described morphologically, and each has well-characterized clinical features. The exomes of all four of these cystic neoplasms have been sequenced, and a distinct set of genes appears to be altered in each. Virtually all SPNs have a beta-catenin gene mutations; the *VHL* gene is targeted in serous cystadenomas; the *RNF43*, *GNAS*, *KRAS*, *TPS3*, *CDKN2A*, and *SMAD4* genes are targeted in IPMNs; and the *RNF43*, *KRAS*, *TPS3*, *CDKN2A*, and *SMAD4* genes are targeted in MCNs [1, 126, 127]. The molecular classification therefore almost perfectly matches the morphologic. This not only validates the classification system but also has immediate clinical implications because it suggests that the type of cystic neoplasm in the pancreas can be determined simply by sequencing cyst fluid samples obtained endoscopically [57, 104].

Neuroendocrine Neoplasms

The current morphologic classification system for neuroendocrine neoplasms of the pancreas lumps together small cell carcinomas of the pancreas and neuroendocrine neoplasms with classic well-differentiated neuroendocrine morphology ("salt and pepper nuclei") and a Ki-67 labeling index of >20% (or >20 mitoses per 10 high power fields) under the designation of neuroendocrine carcinoma [44]. Recent sequencing has, however, demonstrated that these two neoplasms are, in fact, genetically completely distinct. Whole-exome sequencing has identified three pathways that are commonly targeted in well-differentiated pancreatic neuroendocrine tumors (PanNETs). These include the *MEN1* gene, the *DAXX* and *ATRX* genes, and genes coding for members of the mammalian target of rapamycin (mTOR) pathway [42, 51]. By contrast, small cell carcinomas of the pancreas lack these signature mutations. Instead, the *TP53* and *RB* genes are targeted in small cell carcinomas [128]. These molecular analyses make it clear that small cell carcinomas and PanNETs with classic neuroendocrine morphology and a Ki-67 labeling index of >20% should not be grouped together [128].

Molecular analyses also suggest an additional way to classify grade 1 and grade 2 PanNETs. Approximately one in six PanNETs has a mutation in a gene coding for a member of the mTOR pathway [51]. Whereas morphologically indistinguishable from PanNETs without a mTOR pathway mutation, PanNETs with a mutation in a mTOR pathway gene are important to recognize because they are predicted to be sensitive to mTOR pathway inhibitors such as everolimus [63, 131].

Conclusions

The exomes of all of the major types of neoplasms of the pancreas have been sequenced, and neoplasms of the pancreas are now among the best characterized of all neoplasms. An integration of this molecular understanding with the existing morphology-based classification system has helped define new tumor subtypes, discover new markers of cyst type, and has defined therapeutic targets.

Gastric Cancer

Gastric cancer (GC) is the fifth most common cancer worldwide and the third leading cause of cancer mortality worldwide [33]. Despite a declining incidence in many countries in the developed world, there is an increase in global mortality from the disease due to population growth and increasing longevity in developing countries [4, 16, 19, 33]. There is a remarkable tenfold international variation in stomach cancer incidence. In regions of high incidence, cancers of the antrum and pylorus are most common, whereas cancers of the proximal stomach and of the esophagogastric junction (EGJ) are more common in low-incidence countries [92].

Gastric carcinogenesis is a multistep and multifactorial process that, in many cases, appears to involve a progression from normal mucosa through chronic gastritis, atrophic gastritis, and intestinal metaplasia, to dysplasia and invasive carcinoma, a sequence of events that has been designated as the Correa cascade of multistep gastric carcinogenesis [24, 25]. However, the Correa model does not explain all carcinogenic steps of GC. Actually, a proportion of gastric adenocarcinomas arises in non-intestinalized mucosa, and this portion retains a gastric phenotype (and gastric differentiation is also observed in gastric dysplasia, the ultimate precursor lesion of gastric adenocarcinoma) [19, 108, 110].

Invasive GC is highly heterogeneous from the morphological and molecular standpoints. Noteworthy, GC heterogeneity encompasses not only inter-patient variability (intertumoral heterogeneity) but also variations within the same tumor (intratumoral heterogeneity). The latter includes spatial heterogeneity in different tumor areas, and temporal heterogeneity, along progression from primary to recurrent and/or metastatic disease [5].

The large number of histopathological classifications proposed along the years attests to the great variability of GC morphology. Moreover, it is worth pointing out that morphological inratumoral heterogeneity, i.e., the coexistence of different morphological components within the same tumor, is frequent, adding complexity to histological classifications

and raising concerns about accuracy and reproducibility. Among many classifications of GC that have been proposed, those most commonly used are the WHO classification [65], the Laurén classification [64], and, in the East, the classification of the Japanese Gastric Cancer Association [49].

WHO Classification

This classification is based on the recognition of specific histological patterns, also observed in other segments of the gastrointestinal tract, providing a unifying approach. However, this classification does not show direct relationships with epidemiological, clinicopathological, prognostic, or molecular characteristics. WHO classification describes tubular and papillary adenocarcinomas (corresponding to Laurén's intestinal subtype), mucinous adenocarcinoma, poorly cohesive carcinoma (corresponding to Laurén's diffuse subtype), mixed carcinoma, and other rare histological variants, the latter accounting for a small percentage of cases (about 5%).

Tubular Adenocarcinoma

This type consists of tubular structures, branching glands, or acinar structures surrounded by various degrees of desmoplasia. A poorly differentiated variant, composed of compact infiltrative sheets of tumor cells, has been called solid carcinoma. At the other end of the spectrum, (extremely) well-differentiated gastric adenocarcinoma has been described [132], either with gastric or intestinal cell differentiation [54, 68, 109], and some of these mimic complete-type intestinal metaplasia in the stomach [32]. Tubular adenocarcinomas develop mainly in the antrum and body of the stomach and are strongly linked to chronic *H. pylori* infection, atrophic gastritis, and intestinal metaplasia.

Papillary Adenocarcinoma

Typically, this type grows as an exophytic polypoid mass with sharply demarcated invading edge; these adenocarcinomas are composed of pointed or blunted papillary epithelial processes with fibrovascular cores. Papillary adenocarcinomas occur mainly in the proximal stomach and are frequently associated with liver metastases.

Mucinous Adenocarcinoma

Mucinous adenocarcinomas (also described as mucoid or colloid carcinomas) are composed of malignant epithelium mixed with extracellular mucinous pools.

Poorly Cohesive Carcinomas, Including Signet Ring Cell Carcinoma

Discohesive tumors were previously included into a general category of signet ring cell carcinoma even in cases in which signet ring cells were not identified. The current WHO classification recognizes that a general category of poorly cohesive tumors that may display various morphologies: signet ring cell type (composed predominantly or exclusively of signet ring cells characterized by a central optically clear, globoid droplet of cytoplasmic mucin with an eccentrically placed nucleus) and other cell variants (resembling histiocytes or lymphocytes; mimicking plasma cells [39] and bizarre/atypical neoplastic cells). Finally, a mixture of the different cell types can be seen, including signet ring cells.

Mixed Carcinoma

These adenocarcinomas show a mixture of morphologically identifiable glandular (tubular/papillary) and poorly cohesive cellular histological components. Mixed carcinomas have been shown to be clonal [21, 135], and the phenotypic divergence has been attributed to somatic mutation in E-cadherin gene (*CDH1*), restricted to the poorly cohesive component [82]. Enhanced promoter CpG island hypermethylation also has been implicated in the histogenesis of mixed carcinoma [91].

Laurén Classification

Despite dating back to 1965, Laurén classification still remains widely accepted and used. It is based on a dichotomous scheme that distinguishes two main types with distinct epidemiologic profiles, histogenesis, clinicopathological and molecular features, as well as biological behavior:

(i) Intestinal type, forming tubular or papillary structures, most commonly occurs in elderly patients, mainly males. Intestinal GC is more common in high-risk regions and is steadily decreasing in incidence and tends to metastasize hematogenously to the liver.
(ii) Diffuse type, characterized by poorly cohesive and infiltrative tumor cells, which may have, or not, a signet ring cell (SRC) morphology, is more common in young patients, mainly females. Diffuse GC is more common in low-risk areas and has a relatively stable incidence and usually disseminates through peritoneal surfaces.

Both subtypes share environmental risk factors (e.g., *Helicobacter pylori* and Epstein-Barr virus infection, dietary habits, smoking); however, the pathogenesis of diffuse GC is less well understood and encompasses a hereditary component [19, 20].

Moving Toward Molecular Subtyping and Precision Medicine

Many genes have been reported as differentially expressed in different histological types of GC. Some oncogenes are preferentially altered in a specific type of GC, such as *ERBB2/*

HER2 in intestinal-type GC [4, 36, 89]. *HER2* overexpression and/or amplification is present in about 20% of GCs [96]. There is current interest in the immunohistochemical (IHC) expression and in situ hybridization detection of HER2 in GC, scored as 0, 1+, 2+, or 3+ (Fig. 29.1). There is evidence that the tumors with *HER2* overexpression/amplification may respond to therapy with the humanized monoclonal antibody Trastuzumab (Herceptin®), as shown in ToGA trial [8]. Compared to breast carcinoma, HER2 positivity in gastric cancer is frequently heterogenous, and there is a less stringent correlation between *HER2* amplification and protein overexpression [96].

Some oncogenes are altered preferentially in diffuse carcinoma, such as *BCL2* and *FGFR2* (formerly K-*sam*). Other oncogenes are altered in both intestinal-type and diffuse carcinomas, including *CTNNB1* (encoding β-catenin), *MET*, and *MYC*. Many tumor suppressor genes have been implicated in gastric carcinoma development, including *APC* and *DCC* in intestinal-type carcinomas and *CDH1* and *RB1* in diffuse carcinomas. Other tumor suppressor genes are altered in both types of gastric carcinoma, such as *PTEN* and *TP53*, although these are more common in intestinal-type carcinoma.

In recent years, several studies analyzed molecular alterations present in GC at high resolution using various high-throughput platforms [26, 70, 105, 106]. These studies attempted to achieve an integrated molecular classification scheme, clustering the comprehensive molecular data obtained into subgroups with different molecular signatures and clinical phenotypes.

Transcriptomic technologies have been used to identify gene expression signatures with clinical relevance. One study, by Tan IB et al. [105], assessed GC cell lines to identify two "intrinsic" GC subtypes based on gene expression (G-INT and G-DIF). When applied in resected GC samples, the cell line findings correlated with Laurén classification, survival, and response to chemotherapy. These signatures were further analyzed in 248 GC specimens by Lei Z et al. (Singapore-Duke group) [70], leading to the identification of three subtypes that show differences in molecular/genomic features, morphology, "carcinogenic" pathways, and response to therapy: (1) mesenchymal subtype, enriched with diffuse GCs and characterized by cancer stem cell (CSC) and epithelial-to-mesenchymal transition (EMT) gene expression patterns; (2) proliferative subtype, with high levels of *TP53* mutations, genomic instability, and activation of

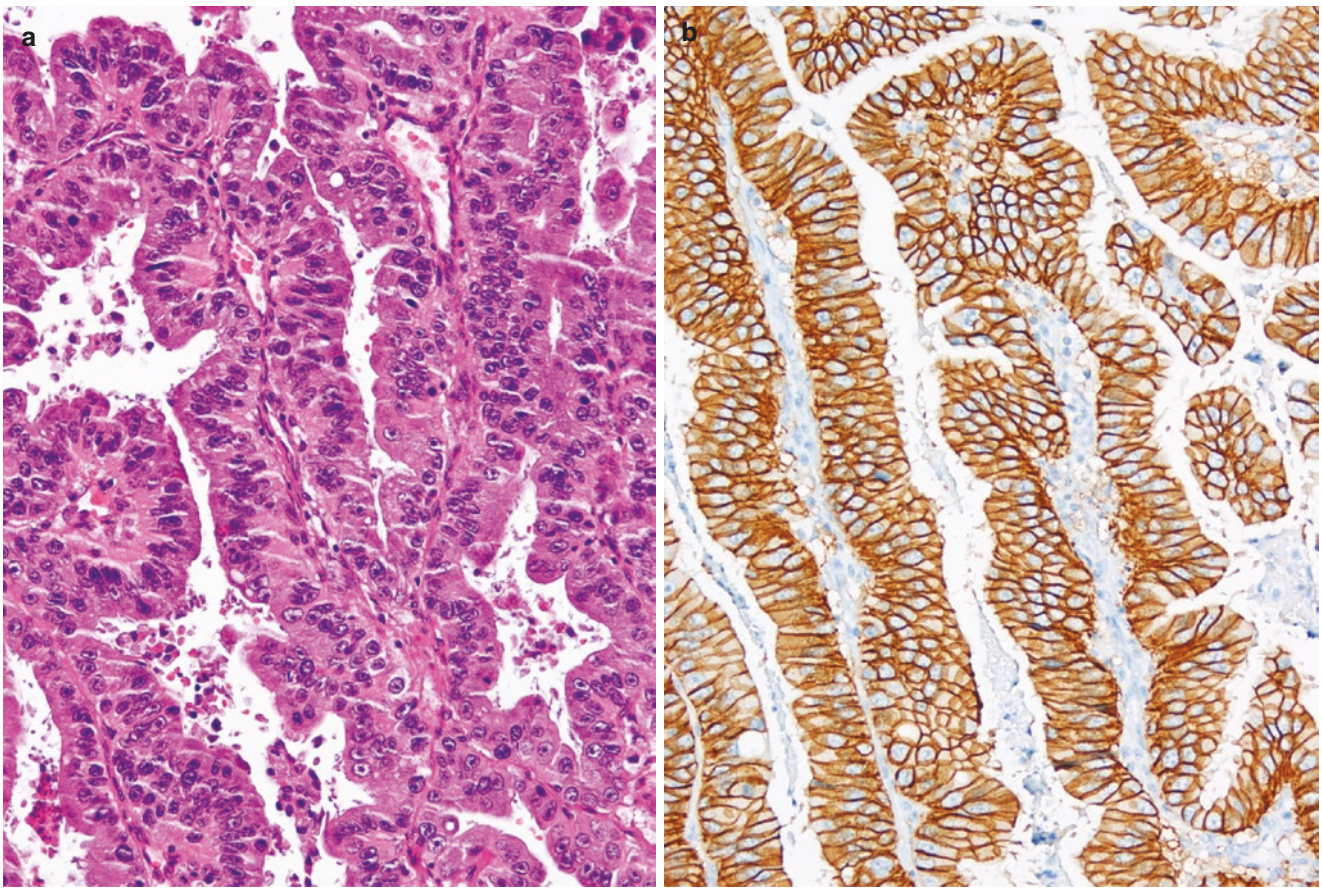


Fig. 29.1 HER2 expression: (**a**) Gastric adenocarcinoma, papillary type (WHO classification) (HE; original magnification 400×); (**b**) immunohistochemical expression of HER2, displaying strong, complete basolateral reactivity, scored as 3+ (IHC-HER2; original magnification 400×)

oncogenic pathways; and (3) metabolic subtype, overexpressing genes that are normally expressed in gastric mucosa. Proliferative and metabolic subtypes frequently correspond to Laurén intestinal type. More recently, Das K et al. confirmed similar results using NanoString nCounter technology in 54 formalin-fixed and paraffin-embedded (FFPE) GC tissues. This study further characterized the expression profiles of mesenchymal, proliferative, and metabolic subtypes and identified new biomarkers of response to mTOR inhibitors (RAD001), namely, overexpression of *MMP9* and *BRCA2* [27].

Deng et al. [28] performed a comprehensive survey of genomic alterations in GC and found systematic patterns of molecular exclusivity of genes related to receptor tyrosine kinase (RTK)/RAS signaling: *FGFR2* (in 9% of tumors), *KRAS* (9%), *EGFR* (8%), *ERBB2* (7%), and *MET* (4%). These genes were frequently amplified in GC in a mutually exclusive manner [28]. However, these results have not been confirmed in recent studies from Korea, using immunohistochemistry and in situ hybridization [61, 89]. In one of these studies [61], RTKs were overexpressed in 218 patients; EGFR was most commonly overexpressed (40%), followed by HER2 (14%) and MET (12%). Furthermore, 2.5% and 11% of cases had simultaneous overexpression of three and two RTKs, respectively [61]. In one study, RTK-amplified GCs (RA-GCs) were observed in 10.5% of 993 consecutive advanced gastric cancer patients who underwent radical gastrectomy, not previously submitted to neoadjuvant chemotherapy, and it was observed that the RA-GC status correlated with older age ($P < 0.001$), differentiated histology ($P = 0.001$), intestinal or mixed type by Laurén classification ($P < 0.001$), lymphovascular invasion ($P = 0.026$), and mutant pattern of *TP53* ($P < 0.001$). Altogether, these studies suggest that a proportion of GC patients may be potentially treatable by RTK/RAS-directed therapies [28, 61].

The landmark study of GC molecular-based stratification was carried out by The Cancer Genome Atlas (TCGA) research network [106]. The TGCA proposed a four-tiered molecular classification that identifies: (1) Epstein-Barr virus-positive (EBV+) GC, characterized by EBV positivity (Fig. 29.2), a stable genome, lack of *TP53* mutations, prevalent *ARID1A* mutation, recurrent *PIK3CA* mutations, frequent *JAK2* and *PD-L1* amplification, and a high level of DNA hypermethylation, as previously reported [58]; (2) GC with microsatellite instability (MSI-high), characterized by DNA hypermethylation, *MLH1* silencing (Fig. 29.3), and mutation in druggable target genes such as *RNF43* and *ERBB2*; (3) genomically stable (GS) GC, associated with diffuse histologic type and recurrent *CDH1* and *RHOA* events, as confirmed by previous studies [55, 69, 117]; and (4) GC with chromosomal instability (CIN), exhibiting intestinal morphology, high number of *TP53* mutations, and amplifications of tyrosine kinase receptors (TKR). EBV+ and MSI-high GCs are particularly interesting. Morphologically, they are characterized by prominent immune infiltrate (Figs. 29.2 and 29.3) [102] and frequently display the features of gastric cancer with lymphoid stroma (GCLS) [37, 75, 80, 90, 116]. There is growing evidence about the possibility of using PD-1/PD-L1 immune checkpoint inhibitors in these two molecular subtypes of GC [67]. In GC, as in other tumor models, PD-L1 overexpression is associated with high densities of CD3+ and CD8+ tumor-infiltrating lymphocytes [59, 74, 107], GCLS morphology [38], and both EBV+ and MSI-high *status* [14, 29, 30, 59, 81]. However, the correlation with *PD-L1* amplification, prognosis, and response to targeted immunotherapies is still debated and deserves further studies.

The Asian Cancer Research Group (ACRG) described four molecular subtypes with distinct prognostic implications [26]: (1) MSI-high tumors, with intestinal morphology and the best prognosis as previously described [60]; (2) epithelial-to-mesenchymal transition (MSS/EMT) GC, with diffuse morphology and the worst prognosis; and (3 and 4) microsatellite-stable adenocarcinomas (MSS), with no EMT signature, either *TP53*-active (MSS/*TP53*+) or *TP53*-inactive (MSS/*TP53*-), and with intermediate prognosis. The

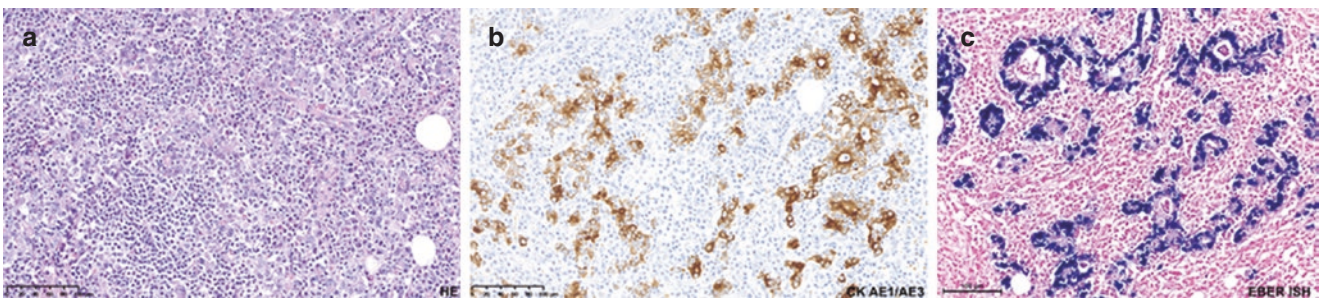


Fig. 29.2 EBV infection in a gastric carcinoma with lymphoid stroma (GCLS). (**a**) Histological features – the inflammatory infiltrate of GCLS is so abundant that it obscures tumor epithelial cells (HE; original magnification 200×). (**b**) Neoplastic cells are highlighted by cytokeratin immunostaining (IHC-AE1/AE3; original magnification 200×). (**c**) EBV infection is demonstrated by EBER-ISH that allows the direct visualization of positive probe signals in cancer-infected cells (ISH; original magnification 200×)

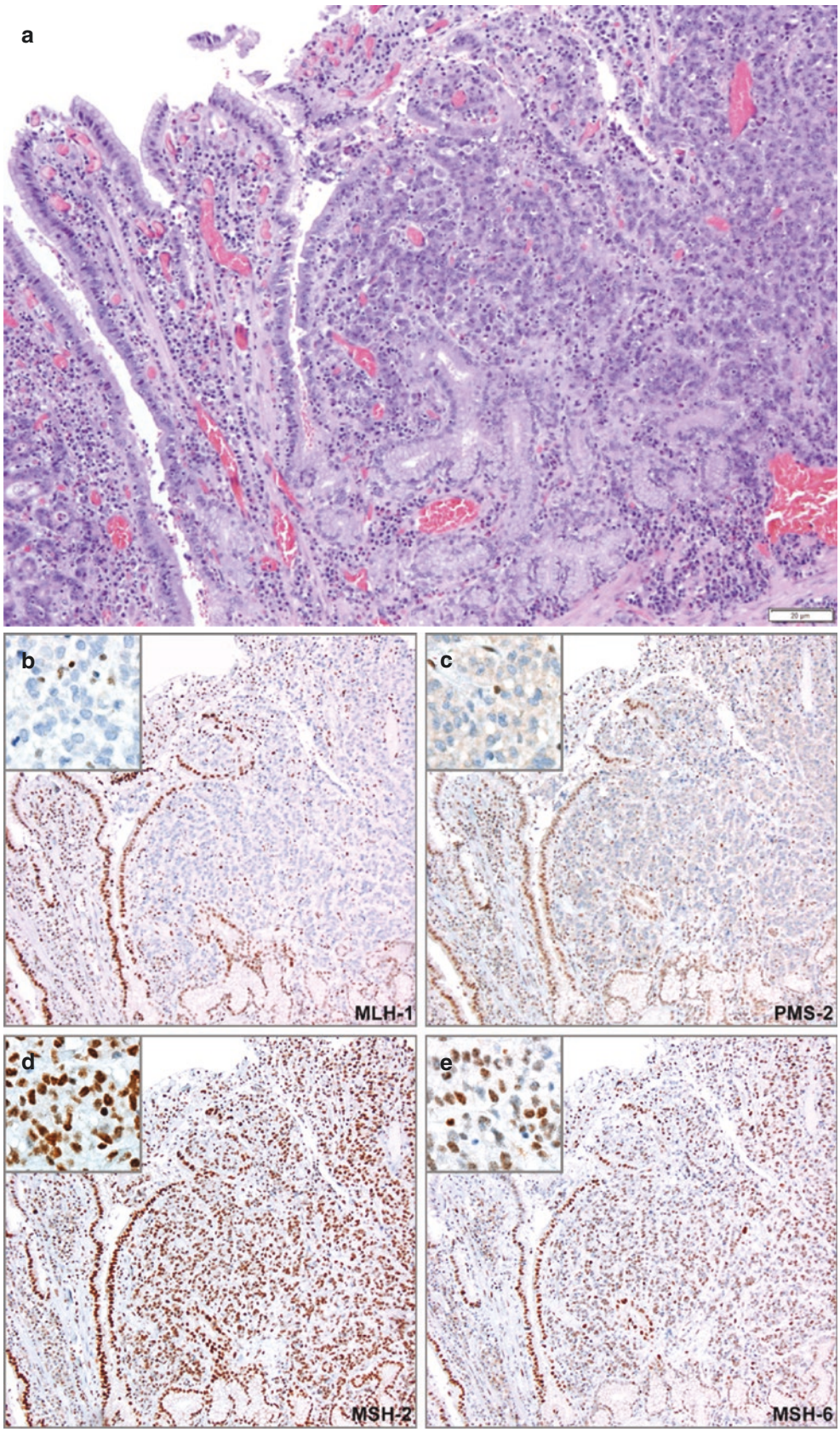


Fig. 29.3 Mismatch repair deficiency (MMRd) evaluated by immunohistochemistry (IHC) in a gastric adenocarcinoma with microsatellite instability (MSI-high). (**a**) Histological features – the stroma of the tumor displays abundant inflammatory infiltrate (HE; original magnification 200×). (**b**) Loss of expression of MLH-1(IHC; original magnification 200×). (**c**) Loss of expression of PMS-2 (IHC; original magnification 200×). (**d**) Preserved nuclear expression of MSH-2 (IHC; original magnification 200×). (**e**) Preserved nuclear expression of MSH-6 (IHC; original magnification 200×)

MSS/*TP53*- subtype (roughly corresponds to the proliferative and CIN subtypes) is prevalent (36–50% of GCs) and harbors genomic amplification of TKR and/or *RAS*, which are in-use or potential therapeutic targets. In keeping with a previous study by Deng N et al. [28], the ACRG found that RTK and *RAS* amplifications tended to be mutually exclusive, emphasizing GC intertumor heterogeneity and the importance of investigating molecular alterations for targeted therapy.

As shown in Table 29.4, the molecular classifications overlap only partially, highlighting the need of a consensual classification that may serve as a roadmap for patient stratification for prognostic evaluation and targeted therapy, in the era of precision medicine.

To date, few correlation studies of morphologic classification and molecular profiles have been carried out [103, 122]. In-depth studies combining molecular features with histopathological and IHC profiles may reveal interesting associations and are currently feasible, thanks to new molecular profiling technologies, generating accurate molecular information from FFPE tissues [27, 83]. On this note, Setia et al. [100] have recently proposed a practical algorithm based on IHC and in situ hybridization techniques currently available in routine diagnostic practice. In this study the authors translated different molecular subgroups into specific immunophenotypes with prognostic and predictive significance. The study of Setia et al. was conducted in a sample of 149 GC cases in a Western population [100]. More recently a validation study was performed in a large-scale Asian cohort (n = 349), providing similar results [3]. Other studies from Korea, encompassing 438 patients that underwent palliative chemotherapy [61] and 993 patients submitted to radical gastrectomy, without neoadjuvant chemotherapy [89], also using immunohistochemistry and in situ hybridization assay, identified a molecular spectrum of distinct GC subtypes.

In an attempt to refine the molecular classification of GC and identify mutational signatures with prognostic capability, Li et al. [73] analyzed the mutation burden of the tumors that were classified into regular (86.8%) and hypermutated (13.2%) subtypes. Major findings derived from this study include (i) there are ubiquitous and specific mutational processes underlying the pathogenesis of different subtypes of GC with varying mutation burdens; (ii) several previously unreported significantly mutated genes (SMGs) were identified; (iii) *CDH1* mutation is an independent prognostic factor of worse survival in patients with diffuse-type GC; and (iv) regular-mutated GC can be further stratified into two subtypes (C1 and C2) with distinct clinical outcomes (C1 is enriched in *TP53*, *XIRP2*, and *APC* mutations and is associated with a significantly better prognostic outcome, whereas C2 is overrepresented by mutations in *ARID1A*, *CDH1*, *PIK3CA*, *ERBB2*, and *RHOA*). Because *ARID1A* is frequently mutated in both EBV and MSI subtypes, its mutation alone is not likely to constitute an alternative GC pathway.

Recently, Yamazawa et al. [129] assessed the expression a panel of primitive phenotypic markers, including embryonic stem cell markers (OCT4, NANOG, SALL4, CLDN6, and LIN28) and known oncofetal proteins (AFP and GPC3), using tissue microarray on 386 GCs. On the basis of the expression profiles, the tumors were clustered into three groups: group 1 (primitive phenotype), AFP, CLDN6, GPC3, or diffuse SALL4 positive; group 2 (SALL4-focal), only focal SALL4 positive; and group 3 (negative), all markers negative. Groups 1 and 2 predominantly consisted of intestinal-type adenocarcinoma, including 13 fetal gut-like adenocarcinomas exclusively identified in group 1. Group 1 was significantly associated with higher T-stage, presence of vascular invasion and nodal metastasis, and poor prognosis compared with groups 2 and 3. Furthermore, group1 phenotype was an independent risk factor for disease-free survival. Group 1 showed frequent *TP53* overexpression and little association with EBV infection or mismatch repair deficiency (MMRd). Further analysis of the Cancer Genome Atlas data set validated these observations and revealed that tumors with primitive phenotypes were mostly classified as "chromosomal instability" (CIN) in the TCGA molecular classification of GC (Table 29.4). On the basis of these findings, Yamazawa et al. [129] claimed that GC with primitive enterocyte phenotypes is an aggressive subgroup of intestinal-type/chromosomal instability GC and suggested that therapeutic strategies targeting primitive markers, such as GPC3, CLDN6, and SALL4, should be considered.

Hereditary Gastric Cancer

Hereditary cancer syndromes linked to 1–3% of GC consist of two principal syndromes: hereditary diffuse gastric cancer (HDGC) and familial intestinal GC [87]. Germline mutations of the E-cadherin (*CDH1*) gene (OMIM# 192090) are the genetic cause of HDGC in about 30–40% of the affected families [87, 111]. Germline mutations in *CTNNA1* have also been identified in HDGC [84]. More recently, exome and multiplexed targeted sequencing led to the identification of new candidate HDGC susceptibility genes, some of which are associated with other hereditary cancer predisposition syndromes, namely, *BRCA2*, *STK11*, *SDHB*, *PRSS1*, *ATM*, *MSR1*, *PALB2*, *INSR*, *FBXO24*, *and DOT1L* [31, 41].

GAPPS (gastric adenocarcinoma and proximal polyposis of the stomach) syndrome, first described in 2012, is an autosomal-dominant cancer predisposition syndrome with a significant risk of gastric, but not colorectal, adenocarcinoma, which is characterized by carpeting fundic gland polyposis restricted to the proximal stomach [125, 130]. GAPPS syndrome is now recognized as a variant of familial adenomatous polyposis, caused by mutations in the promoter region of *APC* gene (exon 1B) [72].

Table 29.4 Molecular/immunohistochemical classifications of GC and their therapeutic implications

Tan IB et al. [105] (n = 270)			G-DIF (44%)	G-INT (56%)	
			Worst prognosis Cisplatin sensitive Diffuse GC	Best prognosis 5-FU and oxaliplatin sensitive Intestinal GC	
Lei Z et al. [70] (n = 248)			Mesenchymal	Proliferative	Metabolic
			Low *TP53* mutations Low E-cadherin mRNA CSC/EMT proprieties mTOR inhibitors Diffuse GC	High *TP53* mutations Genomic instability Oncogene amplification[a] DNA hypomethylation Intestinal GC (intestinal phenotype)	Low *TP53* mutations Normal gastric mucosa gene expression 5-FU sensitive Intestinal GC (gastric phenotype)
Bass AJ et al. [106] The Cancer genome atlas (TCGA) (n = 295)	EBV (9%)	MSI (22%)	GS (20%)	CIN (50%)	
	EBV-CIMP *CDKN2A* silencing *PIK3CA* mutations *PD-L1/2* amplification *JAK2* amplification	Gastric-CIMP *MLH1* silencing *PIK3CA* mutations *HER2/3* mutations *EGFR* mutations	*CDH1* mutations *RHOA* mutations *CLDN18-ARHGAP* fusion (RhoA-GTPase) Diffuse GC	High *TP53* mutations *TKR-RAS* amplification Amplification of cell-cycle Mediators Intestinal GC	
Cristescu R et al. [26] Asian Cancer research group (ACRG) (n = 251		MSI (23%)	MSS/EMT (15%)	MSS/*TP53*- (inactive) (36%)	MSS/*TP53*+ (active) 26%
	EBV+ cases included in MSS/*TP53*+	*MLH1* loss Hypermutation (*KRAS, ARID1A, PIK3CA*) Best prognosis Intestinal GC	*CDH1* loss Worst prognosis Diffuse GC	High *TP53* mutations Genomic instability Oncogene amplification[b] Intermediate prognosis Intestinal GC	Intermediate prognosis Intestinal GC
Yamazawa S et al. [129] Primitive enterocyte phenotype (PEP) (n = 386)				Group 1 (24%) and group 2 (14.5%)	
	Group 3 (61.4%)				
	Group 1 – Primitive phenotype: Expression of AFP, CLDN6, GPC3, or diffuse-SALL4 Higher frequency of advanced staging (T2–T4), vascular invasion, and lymph node metastasis than those in group 2 or 3 TP53 overexpression Poor prognosis Intestinal GC (fetal gut-like pattern noted exclusively in group 1)			Group 2 – SALL4 focal expression Intestinal GC	Group 3 – Negative for all markers Encompasses all TCGA molecular subtypes EBV+ and MSI cases mostly classified in this group Intestinal and diffuse GC

GC gastric cancer, *TCGA* The Cancer Genome Atlas, *ACRG* Asian Cancer Research Group, *5-FU* fluorouracil, *CSC* cancer stem cell, *EMT* epithelial-to-mesenchymal transition, *EBV* Epstein-Barr virus, *CIMP* CpG island methylation phenotype, *MSI* microsatellite instability, *GS* genomically stable, *CIN* chromosomal instability, *TKR* tyrosine kinase receptors, *MSS* microsatellite stable, *PEP* primitive enterocyte phenotype

[a]*ERBB2*, *CCNE1*, *MYC*, and *KRAS*

[b]*ERBB2*, *EGFR*, *CCNE1*, *CCND1*, *MDM2*, *ROBO2*, *GATA6*, and *MYC*

GC is also increased in other heritable syndromes, such as Li-Fraumeni syndrome with germline mutation of *TP53*, Peutz-Jeghers syndrome with frameshift mutation in *STK11*, Lynch syndrome with germline DNA mismatch repair gene mutation, and familial adenomatous polyposis (FAP) with germline *APC* mutation [4, 87].

Conclusions

GC remains the third most common cause of cancer death worldwide, with limited therapeutic strategies available. With the advent of next-generation sequencing, our understanding of its pathogenesis and molecular alterations continues to be revolutionized. These advances are making it feasible to integrate clinical, genome-based, and phenotype-based diagnostic and therapeutic methods and apply them to individual GC patients in the era of precision medicine [78].

Conflict of Interest Ralph Hruban receives royalty payments from Myriad Genetics for the *PALB2* invention.

Fátima Carneiro has no conflicts of interest.

Glossary

MLH1 MUT L homologue 1
MSH MUT S homologue
MSI-H High-grade microsatellite instability
TERT Telomerase RT component

References

1. Abraham SC, Klimstra DS, Wilentz RE, Wu TT, Cameron JL, Yeo CJ, Hruban RH. Solid-pseudopapillary tumors of the pancreas are genetically distinct from pancreatic ductal adenocarcinomas and almost always harbor beta-catenin mutations. Am J Pathol. 2002;160:1361–9.
2. Adsay NV, Merati K, Andea A, Sarkar F, Hruban RH, Wilentz RE, Goggins M, Iocobuzio-Donahue C, Longnecker DS, Klimstra DS. The dichotomy in the preinvasive neoplasia to invasive carcinoma sequence in the pancreas: differential expression of MUC1 and MUC2 supports the existence of two separate pathways of carcinogenesis. Mod Pathol. 2002;15:1087–95.
3. Ahn S, Lee SJ, Kim Y, Kim A, Shin N, Choi KU, Lee CH, Huh GY, Kim KM, Setia N, Lauwers GY, Park DY. High-throughput protein and mRNA expression-based classification of gastric cancers can identify clinically distinct subtypes, concordant with recent molecular classifications. Am J Surg Pathol. 2017;41:106–15.
4. Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric adenocarcinoma. Nat Rev Dis Primers. 2017;3:17036.
5. Alsina M, Gullo I, Carneiro F. Intratumoral heterogeneity in gastric cancer: a new challenge to face. Ann Oncol. 2017;28:912–3.
6. Amato E, Molin MD, Mafficini A, Yu J, Malleo G, Rusev B, Fassan M, Antonello D, Sadakari Y, Castelli P, Zamboni G, Maitra A, Salvia R, Hruban RH, Bassi C, Capelli P, Lawlor RT, Goggins M, Scarpa A. Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas. J Pathol. 2014;233:217–27.
7. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, Nourse C, Murtaugh LC, Harliwong I, Idrisoglu S, Manning S, Nourbakhsh E, Wani S, Fink L, Holmes O, Chin V, Anderson MJ, Kazakoff S, Leonard C, Newell F, Waddell N, Wood S, Xu Q, Wilson PJ, Cloonan N, Kassahn KS, Taylor D, Quek K, Robertson A, Pantano L, Mincarelli L, Sanchez LN, Evers L, Wu J, Pinese M, Cowley MJ, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chantrill LA, Mawson A, Humphris J, Chou A, Pajic M, Scarlett CJ, Pinho AV, Giry-Laterriere M, Rooman I, Samra JS, Kench JG, Lovell JA, Merrett ND, Toon CW, Epari K, Nguyen NQ, Barbour A, Zeps N, Moran-Jones K, Jamieson NB, Graham JS, Duthie F, Oien K, Hair J, Grutzmann R, Maitra A, Iacobuzio-Donahue CA, Wolfgang CL, Morgan RA, Lawlor RT, Corbo V, Bassi C, Rusev B, Capelli P, Salvia R, Tortora G, Mukhopadhyay D, Petersen GM, I. Australian Pancreatic Cancer Genome, Munzy DM, Fisher WE, Karim SA, Eshleman JR, Hruban RH, Pilarsky C, Morton JP, Sansom OJ, Scarpa A, Musgrove EA, Bailey UH, Hofmann O, Sutherland RL, Wheeler DA, Gill AJ, Gibbs RA, Pearson JV, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531(7592):47–52.
8. Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, Lordick F, Ohtsu A, Omuro Y, Satoh T, Aprile G, Kulikov E, Hill J, Lehle M, Ruschoff J, Kang YK. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet (London, England). 2010;376:687–97.
9. Banville N, Geraghty R, Fox E, Leahy DT, Green A, Keegan D, Geoghegan J, O'Donoghue D, Hyland J, Sheahan K. Medullary carcinoma of the pancreas in a man with hereditary nonpolyposis colorectal cancer due to a mutation of the MSH2 mismatch repair gene. Hum Pathol. 2006;37:1498–502.
10. Basturk O, Hong SM, Wood LD, Adsay NV, Albores-Saavedra J, Biankin AV, Brosens L, Fukushima N, Goggins M, Hruban RH, Kato Y, Klimstra D, Klöppel G, Krasinskas AM, Longnecker D, Matthaei H, Offerhaus GJ, Shimizu M, Takaori K, Terris B, Yachida S, Esposito I, Furukawa T. A revised classification system and recommendations from the Baltimore consensus meeting for neoplastic precursor lesions in the pancreas. Am J Surg Pathol. 2015;39(12):1730–41.
11. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, Chang DK, Cowley MJ, Gardiner BB, Song S, Harliwong I, Idrisoglu S, Nourse C, Nourbakhsh E, Manning S, Wani S, Gongora M, Pajic M, Scarlett CJ, Gill AJ, Pinho AV, Rooman I, Anderson M, Holmes O, Leonard C, Taylor D, Wood S, Xu Q, Nones K, Fink JL, Christ A, Bruxner T, Cloonan N, Kolle G, Newell F, Pinese M, Mead RS, Humphris JL, Kaplan W, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chou A, Chin VT, Chantrill LA, Mawson A, Samra JS, Kench JG, Lovell JA, Daly RJ, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N, Kakkar N, Zhao F, Wu YQ, Wang M, Muzny DM, Fisher WE, Brunicardi FC, Hodges SE, Reid JG, Drummond J, Chang K, Han Y, Lewis LR, Dinh H, Buhay CJ, Beck T, Timms L, Sam M, Begley K, Brown A, Pai D, Panchal A, Buchner N, De Borja R, Denroche RE, Yung CK, Serra S, Onetto N, Mukhopadhyay D, Tsao MS, Shaw PA, Petersen GM, Gallinger S, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Schulick RD, Wolfgang CL, Morgan RA, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012;491:399–405.

12. Blackford A, Parmigiani G, Kensler TW, Wolfgang C, Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Eshleman JR, Goggins M, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Klein A, Cameron JL, Olino K, Schulick R, Winter J, Vogelstein B, Velculescu VE, Kinzler KW, Hruban RH. Genetic mutations associated with cigarette smoking in pancreatic cancer. Cancer Res. 2009a;69:3681–8.
13. Blackford A, Serrano OK, Wolfgang CL, Parmigiani G, Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Eshleman JR, Goggins M, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Cameron JL, Olino K, Schulick R, Winter J, Herman JM, Laheru D, Klein AP, Vogelstein B, Kinzler KW, Velculescu VE, Hruban RH. SMAD4 gene mutations are associated with poor prognosis in pancreatic cancer. Clin Cancer Res. 2009b;15:4674–9.
14. Boger C, Behrens HM, Mathiak M, Kruger S, Kalthoff H, Rocken C. PD-L1 is an independent prognostic predictor in gastric cancer of Western patients. Oncotarget. 2016;7:24269–83.
15. Boyd CA, Benarroch-Gampel J, Sheffield KM, Cooksley CD, Riall TS. 415 patients with adenosquamous carcinoma of the pancreas: a population-based analysis of prognosis and survival. J Surg Res. 2012;174:12–9.
16. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the human development index (2008-2030): a population-based study. Lancet Oncol. 2012;13:790–801.
17. Brody JR, Costantino CL, Potoczek M, Cozzitorto J, McCue P, Yeo CJ, Hruban RH, Witkiewicz AK. Adenosquamous carcinoma of the pancreas harbors KRAS2, DPC4 and TP53 molecular alterations similar to pancreatic ductal adenocarcinoma. Mod Pathol. 2009;22(5):651–9.
18. Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, Seymour AB, Weinstein CL, Hruban RH, Yeo CJ, Kern SE. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nature Genet. 1994;8:27–32.
19. Carneiro F. Stomach cancer, pp. 383–391. In: Stewart BW, Wild CP, editors. World cancer report 2014. Lyon: International Agency for Research on Cancer; 2014.
20. Carneiro F, Grabsch HI. Pathogenesis of gastric cancer, pp. 61–72. In: Kukar M, Hochwald SN, editors. Minimally invasive foregut surgery for malignancy: principles and practice. Cham: Springer International Publishing; 2015.
21. Carvalho B, Buffart TE, Reis RM, Mons T, Moutinho C, Silva P, van Grieken NC, Grabsch H, van de Velde CJ, Ylstra B, Meijer GA, Carneiro F. Mixed gastric carcinomas show similar chromosomal aberrations in both their diffuse and glandular components. Cell Oncol. 2006;28:283–94.
22. Chmielecki J, Hutchinson KE, Frampton GM, Chalmers ZR, Johnson A, Shi C, Elvin J, Ali SM, Ross JS, Basturk O, Balasubramanian S, Lipson D, Yelensky R, Pao W, Miller VA, Klimstra DS, Stephens PJ. Comprehensive genomic profiling of pancreatic acinar cell carcinomas identifies recurrent RAF fusions and frequent inactivation of DNA repair genes. Cancer Discov. 2014;4:1398–405.
23. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, Danenberg KL, Tempero MA, Spellman PT, Hanahan D, Gray JW. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011;17:500–3.
24. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process–first American Cancer Society award lecture on cancer epidemiology and prevention. Cancer Res. 1992;52:6735–40.
25. Correa P. The biological model of gastric carcinogenesis. IARC Sci Publ. 2004;150:301–10.
26. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung SH, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S, Aggarwal A. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21:449–56.
27. Das K, Chan XB, Epstein D, Teh BT, Kim KM, Kim ST, Park SH, Kang WK, Rozen S, Lee J, Tan P. NanoString expression profiling identifies candidate biomarkers of RAD001 response in metastatic gastric cancer. ESMO Open. 2016;1:e000009.
28. Deng N, Goh LK, Wang H, Das K, Tao J, Tan IB, Zhang S, Lee M, Wu J, Lim KH, Lei Z, Goh G, Lim QY, Tan AL, Sin Poh DY, Riahi S, Bell S, Shi MM, Linnartz R, Zhu F, Yeoh KG, Toh HC, Yong WP, Cheong HC, Rha SY, Boussioutas A, Grabsch H, Rozen S, Tan P. A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. Gut. 2012;61:673–84.
29. Derks S, Liao X, Chiaravalli AM, Xu X, Camargo MC, Solcia E, Sessa F, Fleitas T, Freeman GJ, Rodig SJ, Rabkin CS, Bass AJ. Abundant PD-L1 expression in Epstein-Barr virus-infected gastric cancers. Oncotarget. 2016;7:32925–32.
30. Dong M, Wang HY, Zhao XX, Chen JN, Zhang YW, Huang Y, Xue L, Li HG, Du H, Wu XY, Shao CK. Expression and prognostic roles of PIK3CA, JAK2, PD-L1, and PD-L2 in Epstein-Barr virus-associated gastric carcinoma. Hum Pathol. 2016;53:25–34.
31. Donner I, Kiviluoto T, Ristimaki A, Aaltonen LA, Vahteristo P. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. Familial Cancer. 2015;14:241–6.
32. Endoh Y, Tamura G, Motoyama T, Ajioka Y, Watanabe H. Well-differentiated adenocarcinoma mimicking complete-type intestinal metaplasia in the stomach. Hum Pathol. 1999;30:826–32.
33. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136:E359–86.
34. Garcia-Carracedo D, Chen ZM, Qiu W, Huang AS, Tang SM, Hruban RH, Su GH. PIK3CA mutations in mucinous cystic neoplasms of the pancreas. Pancreas. 2014;43:245–9.
35. Goggins M, Offerhaus GJ, Hilgers W, Griffin CA, Shekher M, Tang D, Sohn TA, Yeo CJ, Kern SE, Hruban RH. Pancreatic adenocarcinomas with DNA replication errors (RER+) are associated with wild-type K-ras and characteristic histopathology. Poor differentiation, a syncytial growth pattern, and pushing borders suggest RER+. Am J Pathol. 1998;152:1501–7.
36. Grabsch H, Sivakumar S, Gray S, Gabbert HE, Muller W. HER2 expression in gastric cancer: rare, heterogeneous and of no prognostic value – conclusions from 924 cases of two independent series. Cell Oncol. 2010;32:57–65.
37. Grogg KL, Lohse CM, Pankratz VS, Halling KC, Smyrk TC. Lymphocyte-rich gastric cancer: associations with Epstein-Barr virus, microsatellite instability, histology, and survival. Mod Pathol. 2003;16:641–51.
38. Gullo I, Gonçalves G, Pinto ML, Almeida GM, Oliveira C, Carneiro F. Gastric carcinoma with lymphoid stroma: a study of Epstein-Barr virus, microsatellite instability, tumor immune microenvironment and PD-L1 expression. Virchows Archiv (An International Journal of Pathology). 2016;469(Suppl 1):S3–4.
39. Gupta R, Arora R, Das P, Singh MK. Deeply eosinophilic cell variant of signet-ring type of gastric carcinoma: a diagnostic dilemma. Int J Clin Oncol. 2008;13:181–4.
40. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science (New York, NY). 1996;271:350–3.
41. Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, Schrader KA, Schaeffer DF, Shumansky K, Zogopoulos G, Santos TA, Claro I, Carvalho J, Nielsen C, Padilla S, Lum A, Talhouk A, Baker-Lange K, Richardson S, Lewis I, Lindor NM, Pennell E,

MacMillan A, Fernandez B, Keller G, Lynch H, Shah SP, Guilford P, Gallinger S, Corso G, Roviello F, Caldas C, Oliveira C, Pharoah PD, Huntsman DG. Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. JAMA Oncol. 2015;1:23–32.

42. Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, Bettegowda C, Rodriguez FJ, Eberhart CG, Hebbar S, Offerhaus GJ, McLendon R, Rasheed BA, He Y, Yan H, Bigner DD, Oba-Shinjo SM, Marie SK, Riggins GJ, Kinzler KW, Vogelstein B, Hruban RH, Maitra A, Papadopoulos N, Meeker AK. Altered telomeres in tumors with ATRX and DAXX mutations. Science (New York, NY). 2011;333:425.
43. Hosoda W, Chianchiano P, Griffin JF, Pittman ME, Brosens LA, Noe M, Yu J, Shindo K, Suenaga M, Rezaee N, Yonescu R, Ning Y, Albores-Saavedra J, Yoshizawa N, Harada K, Yoshizawa A, Hanada K, Yonehara S, Shimizu M, Uehara T, Samra JS, Gill AJ, Wolfgang CL, Goggins MG, Hruban RH, Wood LD. Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4. J Pathol. 2017;242:16–23.
44. Hruban RH, Pitman MB, Klimstra DS. Tumors of the pancreas. Atlas of tumor pathology, vol. Fourth Series, Fascicle 6. Washington, DC: American Registry of Pathology and Armed Forces Institute of Pathology; 2007.
45. Hruban RH, van Mansfeld ADM, Offerhaus GJ, van Weering DHJ, Allison DC, Goodman SN, Kensler TW, Bose KK, Cameron JL, Bos JL. K- ras oncogene activation in adenocarcinoma of the human pancreas. A study of 82 carcinomas using a combination of mutant-enriched polymerase chain reaction analysis and allele-specific oligonucleotide hybridization. Am J Pathol. 1993;143:545–54.
46. Humphris JL, Patch AM, Nones K, Bailey PJ, Johns AL, McKay S, Chang DK, Miller DK, Pajic M, Kassahn KS, Quinn MC, Bruxner TJ, Christ AN, Harliwong I, Idrisoglu S, Manning S, Nourse C, Nourbakhsh E, Stone A, Wilson PJ, Anderson M, Fink JL, Holmes O, Kazakoff S, Leonard C, Newell F, Waddell N, Wood S, Mead RS, Xu Q, Wu J, Pinese M, Cowley MJ, Jones MD, Nagrial AM, Chin VT, Chantrill LA, Mawson A, Chou A, Scarlett CJ, Pinho AV, Rooman I, Giry-Laterriere M, Samra JS, Kench JG, Merrett ND, Toon CW, Epari K, Nguyen NQ, Barbour A, Zeps N, Jamieson NB, McKay CJ, Carter CR, Dickson EJ, Graham JS, Duthie F, Oien K, Hair J, Morton JP, Sansom OJ, Grutzmann R, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Schulick RD, Wolfgang CL, Morgan RA, Lawlor RT, Rusev B, Corbo V, Salvia R, Cataldo I, Tortora G, Tempero MA, I. Australian Pancreatic Cancer Genome, Hofmann O, Eshleman JR, Pilarsky C, Scarpa A, Musgrove EA, Gill AJ, Pearson JV, Grimmond SM, Waddell N, Biankin AV. Hypermutation in pancreatic cancer. Gastroenterology. 2017;152:68–74.e62.
47. Iacobuzio-Donahue CA, Fu B, Yachida S, Luo M, Abe H, Henderson CM, Vilardell F, Wang Z, Keller JW, Banerjee P, Herman JM, Cameron JL, Yeo CJ, Halushka MK, Eshleman JR, Raben M, Klein AP, Hruban RH, Hidalgo M, Laheru D. DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. J Clin Oncol. 2009;27:1806–13.
48. Infante MH, Sato N, Tonascia J, Klein AP, Riall TS, Yeo CJ, Iacobuzio-Donahue CA, Goggins M. Peritumoral fibroblast SPARC expression and patient outcome with resectable pancreatic adenocarcinoma. J Clin Oncol. 2007;25:319–25.
49. JGCA. Japanese classification of gastric carcinoma: 3rd English edition. Gastric Cancer. 2011;14:101–12.
50. Jiao Y, Yonescu R, Offerhaus GJ, Klimstra DS, Maitra A, Eshleman JR, Herman JG, Poh W, Pelosof L, Wolfgang CL, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N, Wood LD. Whole-exome sequencing of pancreatic neoplasms with acinar differentiation. J Pathol. 2014;232:428–35.
51. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, Diaz LA Jr, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science (New York, NY). 2011;331:1199–203.
52. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, Parmigiani G, Kern SE, Velculescu VE, Kinzler KW, Vogelstein B, Eshleman JR, Goggins M, Klein AP. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science (New York, NY). 2009;324:217.
53. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science (New York, NY). 2008;321:1801–6.
54. Joo M, Han SH. Gastric-type extremely well-differentiated adenocarcinoma of the stomach: a challenge for preoperative diagnosis. J Pathol Transl Med. 2016;50:71–4.
55. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y, Ichimura T, Ushiku T, Funahashi S, Tateishi K, Wada I, Shimizu N, Nomura S, Koike K, Seto Y, Fukayama M, Aburatani H, Ishikawa S. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet. 2014;46:583–7.
56. Kanda M, Sadakari Y, Borges M, Topazian M, Farrell J, Syngal S, Lee J, Kamel I, Lennon AM, Knight S, Fujiwara S, Hruban RH, Canto MI, Goggins M. Mutant TP53 in duodenal samples of pancreatic juice from patients with pancreatic cancer or high-grade dysplasia. Clin Gastroenterol Hepatol. 2013a;11:719–730.e715.
57. Kanda M, Knight S, Topazian M, Syngal S, Farrell J, Lee J, Kamel I, Lennon AM, Borges M, Young A, Fujiwara S, Seike J, Eshleman J, Hruban RH, Canto MI, Goggins M. Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts. Gut. 2013b;62:1024–33.
58. Kaneda A, Matsusaka K, Aburatani H, Fukayama M. Epstein-Barr virus infection as an epigenetic driver of tumorigenesis. Cancer Res. 2012;72:3445–50.
59. Kawazoe A, Kuwata T, Kuboki Y, Shitara K, Nagatsuma AK, Aizawa M, Yoshino T, Doi T, Ohtsu A, Ochiai A. Clinicopathological features of programmed death ligand 1 expression with tumor-infiltrating lymphocyte, mismatch repair, and Epstein-Barr virus status in a large cohort of gastric cancer patients. Gastric Cancer. 2017;20:407–15.
60. Kim H, An JY, Noh SH, Shin SK, Lee YC, Kim H. High microsatellite instability predicts good prognosis in intestinal-type gastric cancers. J Gastroenterol Hepatol. 2011;26:585–92.
61. Kim HS, Shin SJ, Beom SH, Jung M, Choi YY, Son T, Kim HI, Cheong JH, Hyung WJ, Noh SH, Chung H, Park JC, Shin SK, Lee SK, Lee YC, Koom WS, Lim JS, Chung HC, Rha SY, Kim H. Comprehensive expression profiles of gastric cancer molecular subtypes by immunohistochemistry: implications for individualized therapy. Oncotarget. 2016;7:44608–20.
62. Koorstra JB, Maitra A, Morsink FH, Drillenburg P, Ten Kate FJ, Hruban RH, Offerhaus JA. Undifferentiated Carcinoma with Osteoclastic Giant Cells (UCOCGC) of the pancreas associated with the Familial Atypical Multiple Mole Melanoma Syndrome (FAMMM). Am J Surg Pathol. 2008;32:1905–9.
63. Krueger DA, Care MM, Holland K, Agricola K, Tudor C, Mangeshkar P, Wilson KA, Byars A, Sahmoud T, Franz

DN. Everolimus for subependymal giant-cell astrocytomas in tuberous sclerosis. N Engl J Med. 2010;363:1801–11.
64. Laurén P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand. 1965;64:31–49.
65. Lauwers GY, Carneiro F, Graham DY, Curado M-P, Franceschi S, Montgomery E, Tatematsu M, Hattori T. Gastric Cancer. 4th ed. Lyon: IARC; 2010.
66. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhaijee F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardoll DM, Papadopoulos N, Kinzler KW, Zhou S, Cornish TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B, Diaz LA Jr. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015;372:2509–20.
67. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Greten TF, Duffy AG, Ciombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdhoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bever KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA, Diaz LA Jr. Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade. Science (New York, NY). 2017;357(6349):409–13.
68. Lee WA. Gastric extremely well differentiated adenocarcinoma of gastric phenotype: as a gastric counterpart of adenoma malignum of the uterine cervix. World J Surg Oncol. 2005;3:28.
69. Lee YS, Cho YS, Lee GK, Lee S, Kim YW, Jho S, Kim HM, Hong SH, Hwang JA, Kim SY, Hong D, Choi IJ, Kim BC, Kim BC, Kim CH, Choi H, Kim Y, Kim KW, Kong G, Kim HL, Bhak J, Lee SH, Lee JS. Genomic profile analysis of diffuse-type gastric cancers. Genome Biol. 2014;15(4):R55.
70. Lei Z, Tan IB, Das K, Deng N, Zouridis H, Pattison S, Chua C, Feng Z, Guan YK, Ooi CH, Ivanova T, Zhang S, Lee M, Wu J, Ngo A, Manesh S, Tan E, Teh BT, So JB, Goh LK, Boussioutas A, Lim TK, Flotow H, Tan P, Rozen SG. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. Gastroenterology. 2013;145:554–65.
71. Lennon AM, Wolfgang CL, Canto MI, Klein AP, Herman JM, Goggins M, Fishman EK, Kamel I, Weiss MJ, Diaz LA, Papadopoulos N, Kinzler KW, Vogelstein B, Hruban RH. The early detection of pancreatic cancer: what will it take to diagnose and treat curable pancreatic neoplasia? Cancer Res. 2014;74:3381–9.
72. Li J, Woods SL, Healey S, Beesley J, Chen X, Lee JS, Sivakumaran H, Wayte N, Nones K, Waterfall JJ, Pearson J, Patch AM, Senz J, Ferreira MA, Kaurah P, Mackenzie R, Heravi-Moussavi A, Hansford S, Lannagan TR, Spurdle AB, Simpson PT, da Silva L, Lakhani SR, Clouston AD, Bettington M, Grimpen F, Busuttil RA, Di Costanzo N, Boussioutas A, Jeanjean M, Chong G, Fabre A, Olschwang S, Faulkner GJ, Bellos E, Coin L, Rioux K, Bathe OF, Wen X, Martin HC, Neklason DW, Davis SR, Walker RL, Calzone KA, Avital I, Heller T, Koh C, Pineda M, Rudloff U, Quezado M, Pichurin PN, Hulick PJ, Weissman SM, Newlin A, Rubinstein WS, Sampson JE, Hamman K, Goldgar D, Poplawski N, Phillips K, Schofield L, Armstrong J, Kiraly-Borri C, Suthers GK, Huntsman DG, Foulkes WD, Carneiro F, Lindor NM, Edwards SL, French JD, Waddell N, Meltzer PS, Worthley DL, Schrader KA, Chenevix-Trench G. Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. Am J Hum Genet. 2016a;98:830–42.
73. Li X, Wu WK, Xing R, Wong SH, Liu Y, Fang X, Zhang Y, Wang M, Wang J, Li L, Zhou Y, Tang S, Peng S, Qiu K, Chen L, Chen K, Yang H, Zhang W, Chan MT, Lu Y, Sung JJ, Yu J. Distinct subtypes of gastric Cancer defined by molecular characterization include novel mutational signatures with prognostic capability. Cancer Res. 2016b;76:1724–32.
74. Li Z, Lai Y, Sun L, Zhang X, Liu R, Feng G, Zhou L, Jia L, Huang X, Kang Q, Lin D, Gao J, Shen L. PD-L1 expression is associated with massive lymphocyte infiltration and histology in gastric cancer. Hum Pathol. 2016c;55:182–9.
75. Lim H, Park YS, Lee JH, Son DH, Ahn JY, Choi KS, Kim DH, Choi KD, Song HJ, Lee GH, Jung HY, Kim JH, Yook JH, Kim BS. Features of gastric carcinoma with lymphoid stroma associated with Epstein-Barr virus. Clinical Gastroenterol Hepatol. 2015;13:1738–1744.e1732.
76. Liu C, Karam R, Zhou Y, Su F, Ji Y, Li G, Xu G, Lu L, Wang C, Song M, Zhu J, Wang Y, Zhao Y, Foo WC, Zuo M, Valasek MA, Javle M, Wilkinson MF, Lu Y. The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. Nat Med. 2014a;20:596–8.
77. Liu W, Shia J, Gonen M, Lowery MA, O'Reilly EM, Klimstra DS. DNA mismatch repair abnormalities in acinar cell carcinoma of the pancreas: frequency and clinical significance. Pancreas. 2014b;43:1264–70.
78. Liu X, Meltzer SJ. Gastric Cancer in the era of precision medicine. Cell Mol Gastroenterol Hepatol. 2017;3:348–58.
79. Lowery MA, Kelsen DP, Stadler ZK, Yu KH, Janjigian YY, Ludwig E, D'Adamo DR, Salo-Mullen E, Robson ME, Allen PJ, Kurtz RC, O'Reilly EM. An emerging entity: pancreatic adenocarcinoma associated with a known BRCA mutation: clinical descriptors, treatment implications, and future directions. Oncologist. 2011;16:1397–402.
80. Lu BJ, Lai M, Cheng L, Xu JY, Huang Q. Gastric medullary carcinoma, a distinct entity associated with microsatellite instability-H, prominent intraepithelial lymphocytes and improved prognosis. Histopathology. 2004;45:485–92.
81. Ma C, Patel K, Singhi AD, Ren B, Zhu B, Shaikh F, Sun W. Programmed death-ligand 1 expression is common in gastric cancer associated with Epstein-Barr virus or microsatellite instability. Am J Surg Pathol. 2016;40:1496–506.
82. Machado JC, Soares P, Carneiro F, Rocha A, Beck S, Blin N, Berx G, Sobrinho-Simoes M. E-cadherin gene mutations provide a genetic basis for the phenotypic divergence of mixed gastric carcinomas. Lab Investig. 1999;79:459–65.
83. Mafficini A, Amato E, Fassan M, Simbolo M, Antonello D, Vicentini C, Scardoni M, Bersani S, Gottardi M, Rusev B, Malpeli G, Corbo V, Barbi S, Sikora KO, Lawlor RT, Tortora G, Scarpa A. Reporting tumor molecular heterogeneity in histopathological diagnosis. PLoS One. 2014;9:e104979.
84. Majewski IJ, Kluijt I, Cats A, Scerri TS, de Jong D, Kluin RJ, Hansford S, Hogervorst FB, Bosma AJ, Hofland I, Winter M, Huntsman D, Jonkers J, Bahlo M, Bernards R. An alpha-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. J Pathol. 2013;229:621–9.
85. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, Bentrem DJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA, Yeh JJ. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78.
86. Nakata B, Wang YQ, Yashiro M, Ohira M, Ishikawa T, Nishino H, Seki S, Hirakawa K. Negative hMSH2 protein expression in pancreatic carcinoma may predict a better prognosis of patients. Oncol Rep. 2003;10:997–1000.

87. Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015;16:e60–70.
88. Osaka H, Yashiro M, Nishino H, Nakata B, Ohira M, Hirakawa K. A case of osteoclast-type giant cell tumor of the pancreas with high-frequency microsatellite instability. Pancreas. 2004;29:239–41.
89. Park CK, Park JS, Kim HS, Rha SY, Hyung WJ, Cheong JH, Noh SH, Lee SK, Lee YC, Huh YM, Kim H. Receptor tyrosine kinase amplified gastric cancer: clinicopathologic characteristics and proposed screening algorithm. Oncotarget. 2016;7:72099–112.
90. Park S, Choi MG, Kim KM, Kim HS, Jung SH, Lee JH, Noh JH, Sohn TS, Bae JM, Kim S. Lymphoepithelioma-like carcinoma: a distinct type of gastric cancer. J Surg Res. 2015;194:458–63.
91. Park SY, Kook MC, Kim YW, Cho NY, Kim TY, Kang GH. Mixed-type gastric cancer and its association with high-frequency CpG island hypermethylation. Virchows Arch. 2010;456:625–33.
92. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer. 2006;118:3030–44.
93. Redston MS, Caldas C, Seymour AB, Hruban RH, da Costa L, Yeo CJ, Kern SE. p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. Cancer Res. 1994;54:3025–33.
94. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu VE, Goggins M, Vogelstein B, Papadopoulos N, Hruban RH, Kinzler KW, Klein AP. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2012;2:41–6.
95. Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, Kurtz RC, Olson SH, Rustgi AK, Schwartz AG, Stoffel E, Syngal S, Zogopoulos G, Ali SZ, Axilbund J, Chaffee KG, Chen YC, Cote ML, Childs EJ, Douville C, Goes FS, Herman JM, Iacobuzio-Donahue C, Kramer M, Makohon-Moore A, McCombie RW, McMahon KW, Niknafs N, Parla J, Pirooznia M, Potash JB, Rhim AD, Smith AL, Wang Y, Wolfgang CL, Wood LD, Zandi PP, Goggins M, Karchin R, Eshleman JR, Papadopoulos N, Kinzler KW, Vogelstein B, Hruban RH, Klein AP. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. Cancer Discov. 2016;6:166–75.
96. Ruschoff J, Dietel M, Baretton G, Arbogast S, Walch A, Monges G, Chenard MP, Penault-Llorca F, Nagelmeier I, Schlake W, Hofler H, Kreipe HH. HER2 diagnostics in gastric cancer-guideline validation and development of standardized immunohistochemical testing. Virchows Arch. 2010;457:299–307.
97. Samdani RT, Hechtman JF, O'Reilly E, DeMatteo R, Sigel CS. EBV-associated lymphoepithelioma-like carcinoma of the pancreas: case report with targeted sequencing analysis. Pancreatology. 2015;15:302–4.
98. Sausen M, Phallen J, Adleff V, Jones S, Leary RJ, Barrett MT, Anagnostou V, Parpart-Li S, Murphy D, Kay Li Q, Hruban CA, Scharpf R, White JR, O'Dwyer PJ, Allen PJ, Eshleman JR, Thompson CB, Klimstra DS, Linehan DC, Maitra A, Hruban RH, Diaz LA Jr, Von Hoff DD, Johansen JS, Drebin JA, Velculescu VE. Clinical implications of genomic alterations in the tumor and circulation of pancreatic cancer patients. Nat Commun. 2015;6:7686.
99. Seidel G, Zahurak M, Iacobuzio-Donahue CA, Sohn TA, Adsay NV, Yeo CJ, Lillemoe KD, Cameron JL, Hruban RH, Wilentz RE. Almost all infiltrating colloid carcinomas of the pancreas and periampullary region arise from in situ papillary neoplasms: a study of 39 cases. Am J Surg Pathol. 2002;26:56–63.
100. Setia N, Agoston AT, Han HS, Mullen JT, Duda DG, Clark JW, Deshpande V, Mino-Kenudson M, Srivastava A, Lennerz JK, Hong TS, Kwak EL, Lauwers GY. A protein and mRNA expression-based classification of gastric cancer. Modern Pathol. 2016;29:772–84.
101. Smit VTHBM, Boot AJM, Smits AMM, Fleuren GJ, Cornelisse CJ, Bos JL. K- ras codon 12 mutations occur very frequently in pancreatic adenocarcinomas. Nucleic Acids Res. 1988;16:7773–82.
102. Solcia E, Klersy C, Mastracci L, Alberizzi P, Candusso ME, Diegoli M, Tava F, Riboni R, Manca R, Luinetti O. A combined histologic and molecular approach identifies three groups of gastric cancer with different prognosis. Virchows Arch. 2009;455:197–211.
103. Speck O, Tang W, Morgan DR, Kuan PF, Meyers MO, Dominguez RL, Martinez E, Gulley ML. Three molecular subtypes of gastric adenocarcinoma have distinct histochemical features reflecting Epstein-Barr virus infection status and neuroendocrine differentiation. Appl Immunohistochem Mol Morphol. 2015;23:633–45.
104. Springer S, Wang Y, Molin MD, Masica DL, Jiao Y, Kinde I, Blackford A, Raman SP, Wolfgang CL, Tomita T, Niknafs N, Douville C, Ptak J, Dobbyn L, Allen PJ, Klimstra DS, Schattner MA, Schmidt MC, Yip-Schneider M, Cummings OW, Brand RE, Zeh HJ, Singhi AD, Scarpa A, Salvia R, Malleo G, Zamboni G, Falconi M, Jang JY, Kim SW, Kwon W, Hong SM, Song KB, Kim SC, Swan N, Murphy J, Geoghegan J, Brugge W, Fernandez-Del Castillo C, Mino-Kenudson M, Schulick R, Edil BH, Adsay V, Paulino J, van Hooft J, Yachida S, Nara S, Hiraoka N, Yamao K, Hijioka S, van der Merwe S, Goggins M, Canto MI, Ahuja N, Hirose K, Makary M, Weiss MJ, Cameron J, Pittman M, Eshleman JR, Diaz LA Jr, Papadopoulos N, Kinzler KW, Karchin R, Hruban RH, Vogelstein B, Lennon AM. A combination of molecular markers and clinical features improve the classification of pancreatic cysts. Gastroenterology. 2015;149:1501–10.
105. Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, Tan SH, Wu J, Lee MH, Ooi CH, Rha SY, Wong WK, Boussioutas A, Yeoh KG, So J, Yong WP, Tsuburaya A, Grabsch H, Toh HC, Rozen S, Cheong JH, Noh SH, Wan WK, Ajani JA, Lee JS, Tellez MS, Tan P. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. Gastroenterology. 2011;141:476–85, 485.e471-411.
106. TCGA. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202–9.
107. Thompson ED, Zahurak M, Murphy A, Cornish T, Cuka N, Abdelfatah E, Yang S, Duncan M, Ahuja N, Taube JM, Anders RA, Kelly RJ. Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. Gut. 2017;66:794–801.
108. Tsukashita S, Kushima R, Bamba M, Sugihara H, Hattori T. MUC gene expression and histogenesis of adenocarcinoma of the stomach. Int J Cancer. 2001;94:166–70.
109. Ushiku T, Arnason T, Ban S, Hishima T, Shimizu M, Fukayama M, Lauwers GY. Very well-differentiated gastric carcinoma of intestinal type: analysis of diagnostic criteria. Modern Pathol. 2013;26:1620–31.
110. Valente P, Garrido M, Gullo I, Baldaia H, Marques M, Baldaque-Silva F, Lopes J, Carneiro F. Epithelial dysplasia of the stomach with gastric immunophenotype shows features of biological aggressiveness. Gastric Cancer. 2015;18:720–8.
111. van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, Caldas C, Schreiber KE, Hardwick RH, Ausems MG, Bardram L, Benusiglio PR, Bisseling TM, Blair V, Bleiker E, Boussioutas A, Cats A, Coit D, DeGregorio L, Figueiredo J, Ford JM, Heijkoop E, Hermens R, Humar B, Kaurah P, Keller G, Lai J, Ligtenberg MJ, O'Donovan M, Oliveira C, Pinheiro H, Ragunath K, Rasenberg E, Richardson S, Roviello F, Schackert H, Seruca R, Taylor A, Ter Huurne A, Tischkowitz M, Joe ST, van Dijck B, van Grieken NC, van Hillegersberg R, van Sandick JW, Vehof R,

van Krieken JH, Fitzgerald RC. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J Med Genet. 2015;52:361–74.
112. Vogelstein B, Kinzler KW. The path to cancer – three strikes and You're out. N Engl J Med. 2015;373:1895–8.
113. Von Hoff DD, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, Seay T, Tjulandin SA, Ma WW, Saleh MN, Harris M, Reni M, Dowden S, Laheru D, Bahary N, Ramanathan RK, Tabernero J, Hidalgo M, Goldstein D, Van Cutsem E, Wei X, Iglesias J, Renschler MF. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med. 2013;369:1691–703.
114. Voong KR, Davison J, Pawlik TM, Uy MO, Hsu CC, Winter J, Hruban RH, Laheru D, Rudra S, Swartz MJ, Nathan H, Edil BH, Schulick R, Cameron JL, Wolfgang CL, Herman JM. Resected pancreatic adenosquamous carcinoma: clinicopathologic review and evaluation of adjuvant chemotherapy and radiation in 38 patients. Hum Pathol. 2010;41:113–22.
115. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, Quinn MC, Robertson AJ, Fadlullah MZ, Bruxner TJ, Christ AN, Harliwong I, Idrisoglu S, Manning S, Nourse C, Nourbakhsh E, Wani S, Wilson PJ, Markham E, Cloonan N, Anderson MJ, Fink JL, Holmes O, Kazakoff SH, Leonard C, Newell F, Poudel B, Song S, Taylor D, Wood S, Xu Q, Wu J, Pinese M, Cowley MJ, Lee HC, Jones MD, Nagrial AM, Humphris J, Chantrill LA, Chin V, Steinmann AM, Mawson A, Humphrey ES, Colvin EK, Chou A, Scarlett CJ, Pinho AV, Giry-Laterriere M, Rooman I, Samra JS, Kench JG, Pettitt JA, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N, Jamieson NB, Graham JS, Niclou SP, Bjerkvig R, Grutzmann R, Aust D, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Wolfgang CL, Morgan RA, Lawlor RT, Corbo V, Bassi C, Falconi M, Zamboni G, Tortora G, Tempero MA, Gill AJ, Eshleman JR, Pilarsky C, Scarpa A, Musgrove EA, Pearson JV, Biankin AV, Grimmond SM. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015;518:495–501.
116. Wang HH, Wu MS, Shun CT, Wang HP, Lin CC, Lin JT. Lymphoepithelioma-like carcinoma of the stomach: a subset of gastric carcinoma with distinct clinicopathological features and high prevalence of Epstein-Barr virus infection. Hepato-Gastroenterology. 1999;46:1214–9.
117. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, Siu HC, Deng S, Chu KM, Law S, Chan KH, Chan AS, Tsui WY, Ho SL, Chan AK, Man JL, Foglizzo V, Ng MK, Chan AS, Ching YP, Cheng GH, Xie T, Fernandez J, Li VS, Clevers H, Rejto PA, Mao M, Leung SY. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014;46:573–82.
118. Westra WH, Sturm PJ, Drillenburg P, Choti MA, Klimstra DS, Abores-Saavedra J, Montag A, Offerhaus GJ, Hruban RH. K-ras oncogene mutations in osteoclast-like giant cell tumors of the pancreas and liver: genetic evidence to support origin from the duct epithelium. Am J Surg Pathol. 1998;22:1247–54.
119. Wilentz RE, Goggins M, Redston M, Marcus VA, Adsay NV, Sohn TA, Kadkol SS, Yeo CJ, Choti MA, Zahurak M, Johnson KA, Tascilar M, Offerhaus GJ, Hruban RH, Kern SE. Genetic, immunohistochemical, and clinical features of medullary carcinoma of the pancreas: a newly described and characterized entity. Am J Pathol. 2000;156:1641–51.
120. Winter JM, Ting AH, Vilardell F, Gallmeier E, Baylin SB, Hruban RH, Kern SE, Iacobuzio-Donahue CA. Absence of E-cadherin expression distinguishes noncohesive from cohesive pancreatic cancer. Clin Cancer Res. 2008;14:412–8.
121. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, Mollaee M, Wagner KU, Koduru P, Yopp A, Choti MA, Yeo CJ, McCue P, White MA, Knudsen ES. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun. 2015;6:6744.
122. Wong H, Yau T. Molecular targeted therapies in advanced gastric cancer: does tumor histology matter? Ther Adv Gastroenterol. 2013;6:15–31.
123. Wood LD, Hruban RH. Pathology and molecular genetics of pancreatic neoplasms. Cancer J. 2012;18:492–501.
124. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. Science (New York, NY). 2007;318:1108–13.
125. Worthley DL, Phillips KD, Wayte N, Schrader KA, Healey S, Kaurah P, Shulkes A, Grimpen F, Clouston A, Moore D, Cullen D, Ormonde D, Mounkley D, Wen X, Lindor N, Carneiro F, Huntsman DG, Chenevix-Trench G, Suthers GK. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut. 2012;61:774–9.
126. Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, Goggins M, Canto MI, Schulick RD, Edil BH, Wolfgang CL, Klein AP, Diaz LA Jr, Allen PJ, Schmidt CM, Kinzler KW, Papadopoulos N, Hruban RH, Vogelstein B. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med. 2011a;3:92ra66.
127. Wu J, Jiao Y, Dal Molin M, Maitra A, de Wilde RF, Wood LD, Eshleman JR, Goggins MG, Wolfgang CL, Canto MI, Schulick RD, Edil BH, Choti MA, Adsay V, Klimstra DS, Offerhaus GJ, Klein AP, Kopelovich L, Carter H, Karchin R, Allen PJ, Schmidt CM, Naito Y, Diaz LA Jr, Kinzler KW, Papadopoulos N, Hruban RH, Vogelstein B. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. Proc Natl Acad Sci U S A. 2011b;108:21188–93.
128. Yachida S, Vakiani E, White CM, Zhong Y, Saunders T, Morgan R, de Wilde RF, Maitra A, Hicks J, Demarzo AM, Shi C, Sharma R, Laheru D, Edil BH, Wolfgang CL, Schulick RD, Hruban RH, Tang LH, Klimstra DS, Iacobuzio-Donahue CA. Small cell and large cell neuroendocrine carcinomas of the pancreas are genetically similar and distinct from well-differentiated pancreatic neuroendocrine tumors. Am J Surg Pathol. 2012;36:173–84.
129. Yamazawa S, Ushiku T, Shinozaki-Ushiku A, Hayashi A, Iwasaki A, Abe H, Tagashira A, Yamashita H, Seto Y, Aburatani H, Fukayama M. Gastric cancer with primitive enterocyte phenotype: an aggressive subgroup of intestinal-type adenocarcinoma. Am J Surg Pathol. 2017;41(7):989–97.
130. Yanaru-Fujisawa R, Nakamura S, Moriyama T, Esaki M, Tsuchigame T, Gushima M, Hirahashi M, Nagai E, Matsumoto T, Kitazono T. Familial fundic gland polyposis with gastric cancer. Gut. 2012;61:1103–4.
131. Yao JC, Shah MH, Ito T, Bohas CL, Wolin EM, Van Cutsem E, Hobday TJ, Okusaka T, Capdevila J, de Vries EG, Tomassetti P, Pavel ME, Hoosen S, Haas T, Lincy J, Lebwohl D, Oberg K. Everolimus for advanced pancreatic neuroendocrine tumors. N Engl J Med. 2011;364:514–23.
132. Yao T, Utsunomiya T, Oya M, Nishiyama K, Tsuneyoshi M. Extremely well-differentiated adenocarcinoma of the stomach: clinicopathological and immunohistochemical features. World J Gastroenterol. 2006;12:2510–6.
133. Yu J, Sadakari Y, Shindo K, Suenaga M, Brant A, Almario JA, Borges M, Barkley T, Fesharakizadeh S, Ford M, Hruban RH, Shin EJ, Lennon AM, Canto MI, Goggins M. Digital next-

generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms. Gut. 2016;66(9):1677–87.
134. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, Hruban RH, Cote ML, McWilliams RR, Roberts NJ, Cannon-Albright LA, Li D, Moyes K, Wenstrup RJ, Hartman AR, Seminara D, Klein AP, Petersen GM. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med. 2015;17:569–77.
135. Zheng HC, Li XH, Hara T, Masuda S, Yang XH, Guan YF, Takano Y. Mixed-type gastric carcinomas exhibit more aggressive features and indicate the histogenesis of carcinomas. Virchows Arch. 2008;452:525–34.

Molecular Pathology of Genitourinary Cancers: Translating the Cancer Genome to the Clinic

30

Martin J. Magers, Joshua I. Warrick, and Scott A. Tomlins

Introduction

Genitourinary malignancies, including cancers of the prostate, urinary bladder, kidney, testis, and penis, are major causes of cancer morbidity and mortality in the USA. Interrogation of the cancer genome and transcriptome, through single-gene assays (including assessment of gene products by immunohistochemistry [IHC]), multiplexed panels, and targeted or full sequencing, has led to major advances in our understanding of the molecular underpinnings of numerous cancers. High-throughput technologies, such as DNA microarrays and next-generation sequencing (NGS), combined with large international efforts to comprehensively interrogate cancer genomes and transcriptomes such as The Cancer Genome Atlas (TCGA), will likely lead to a complete cataloguing of the aberrations present in genitourinary cancers. The shift toward selecting the right therapy for the molecular alteration(s) driving a patient's particular cancer in the era of precision medicine will increase the clinical demand for routine cancer genome/transcriptome assessment. Importantly, pathologists are ideally suited to be leading the efforts to understand the range and diversity of these aberrations, how they can be assessed in routine specimens, and which assay(s) can be used best to answer important clinical questions (e.g., "which is the best therapy for my patient with bladder cancer?"). In this chapter, we aim to provide an overview of the range of driving genome or transcriptome alterations in common genitourinary cancers. We have focused on important single genes, multigene panels, and findings from exome−/genome-wide interrogation. We have attempted to place these lesions and related assays into a clinical context, particularly regarding current and future translation in relation to areas of clinical need. Lastly, given the explosion in reports and assays for interrogating the cancer genome from the single gene level (through techniques such as IHC, fluorescence in situ hybridization [FISH], capillary sequencing) to full genome-scale sequencing, we have sought to point out salient issues to be considered by the pathologist when thinking about implementing novel biomarkers or assays.

M. J. Magers
Department of Pathology, University of Michigan Medical School, Ann Arbor, MI, USA

Department of Pathology, Indiana University School of Medicine, Indianapolis, IN, USA

J. I. Warrick
Departments of Pathology and Surgery, Penn State College of Medicine, Hershey, PA, USA

S. A. Tomlins (✉)
Departments of Pathology and Urology, University of Michigan Medical School, Ann Arbor, MI, USA

Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, MI, USA

Rogel Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA
e-mail: tomlinss@umich.edu

The Molecular Pathology of Prostate Cancer

Prostate cancer is the leading type of cancer and a leading cause of cancer death affecting American men [1]. Current prognostic models rely heavily on pathologic grade and stage. Tumor grade is determined by the Gleason grading system, which assigns numeric values (range 1–5) to tumor architecture, and in simplified terms sums the two most prevalent patterns to achieve an overall Gleason score (range 2–10). Cancers with Gleason score 2–5 are uncommon. Tumors with Gleason score 3 + 3 = 6 have indolent behavior, with an extremely low chance of causing patient death. In contrast, those with a higher Gleason score have greater potential for metastasis and causing death, which increases with the score [2, 3]. Recently, a system of Grade Groups (range 1–5) for Gleason scores was proposed. Briefly, the Grade Groups are as follows: Gleason score ≤6 is Grade Group 1; Gleason score 3 + 4 = 7 is Grade Group 2; Gleason score 4 + 3 = 7 is Grade Group 3; Gleason score = 8 is Grade Group 4; and Gleason score ≥9 is Grade Group 5. Grade Groupings not only are more intuitive for

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_30

patients than Gleason scores, but Gleason score = 7 is more accurately stratified [4, 5]. Pathologic stage also strongly correlates with prognosis, with higher rates of metastasis and death being associated with cancer extending outside the prostate.

Single Genes in Prostate Cancer

ETS Gene Fusions

Fusion genes resulting from rearrangements involving members of the ETS transcription factor family are the most common known molecular abnormality in prostate cancer, seen in ~50% of cases detected by serum PSA screening [6–8]. The most common rearrangement involves either chromosomal deletion or insertion of chromosome 21, resulting in fusion of the 5′ untranslated region of *TMPRSS2*, an androgen-regulated gene, with the ETS family member *ERG*. This fusion results in androgen-driven expression of full length (or minimally N-terminally truncated) ERG protein product (Fig. 30.1). The vast majority (>90%) of ETS rearrangements involves *ERG*, while the remaining ETS fusions include *ETV1* (chromosome 7), *ETV5* (chromosome 3), or *ETV4* (chromosome 17) as common 3′ partners. Similarly, although *TMPRSS2* is the most common 5′ partner for ERG, other 5′ partners, including *SLC45A3* and *NDRG1*, have been identified. Non-*ERG* ETS gene rearrangements commonly have been identified with a variety of 5′ partners.

ETS rearrangements can readily be detected by FISH [8–10]. In the most commonly used method, fluorescently labeled probes (typically red and green) flanking the regions just 5′ and 3′ to the ETS gene of interest are used. If no ETS rearrangement is present, two fused signals (typically yellow) will be identified per cell, as the probes are close to one another with resulting color addition of red and green signals. Loss of the region 5′ to the ETS gene, as is seen in fusion through deletion, will result in loss of one probe, consequentially showing one yellow signal and one signal the color of the 3′ probe. Similarly, if the material 5′ to the ETS gene is lost to a separate chromosome through insertion, a single yellow signal will be seen in addition to separate red and green signals (Fig. 30.1). Alternative approaches, including three-color FISH, with probes located 5′ to *TMPRSS2*, as well as 5′ and 3′ to *ERG*, have also been described [10].

The *TMPRSS2-ERG* rearrangement results in overexpression of the fusion gene protein product, which is nearly full-length ERG protein with no contribution from *TMPRSS2*. Monoclonal antibodies have therefore been developed against this target, and IHC utilizing these antibodies has been shown to strongly correlate with *TMPRSS2-ERG* fusion by FISH (>95% sensitivity and specificity for detection of translocation) [11–15]. ERG IHC has also been shown to be >99% specific for prostate cancer and high-grade prostatic intraepithelial neoplasia (HGPIN), which is nearly invariably adjacent to ERG^+ carcinoma in prostatectomy specimens [12, 16]. Examples of *ERG* FISH and IHC are shown in Fig. 30.1.

Although FISH for *ERG* rearrangements and IHC for ERG expression are highly concordant in the great majority of cases of prostatic adenocarcinoma, important exceptions are neuroendocrine prostatic carcinoma (typically small-cell carcinoma) and poorly differentiated prostatic adenocarcinomas [e.g., acinar adenocarcinoma, Gleason score 5 + 5 = 10 (Grade Group 5)]. Although this occurs infrequently at presentation, in the setting of prolonged androgen deprivation therapy, prostate cancers develop resistance to hormonal therapy and exhibit loss of androgen receptor (AR) signaling, which may be accompanied by development of a neuroendocrine/small-cell carcinoma phenotype. The *TMPRSS2*:*ERG* transcripts encode a slightly truncated ERG protein product, which is driven by the androgen response elements upstream of *TMPRSS2*. Hence, prostate cancers that have lost AR signaling (e.g., small-cell carcinomas in the setting of androgen deprivation therapy) will not express ERG by IHC, although the *TMPRSS2-ERG* rearrangement is still detectable by FISH (Fig. 30.1). This caveat is important in the setting of determining the site of origin of a cancer of unknown primary. That is, if other AR-regulated products, such as tissue PSA (IHC), are negative in a cancer of unknown primary, there is little value in assessing ERG protein expression. On the other hand, given that *ERG* rearrangements are present at the DNA level in ~50% of all prostate cancers (regardless of AR signaling status), and *ERG* rearrangements are maintained in prostate cancers that dedifferentiate, FISH for *ERG* rearrangement can be helpful in tumors that do not express PSA, but when clinical suspicion for a prostatic origin remains.

The clinical utility of ERG assessment is beginning to emerge, most commonly by IHC, given the ease of incorporating it into existing pathology workflows. Because ERG expression is highly specific for prostate cancer (orders of magnitude more specific than alpha-methylacyl-CoA racemase [AMACR], a protein preferentially expressed by prostate cancer), ERG immunoexpression has shown promise in classifying diagnostically challenging, small acinar foci identiifed on prostate needle core biopsy [17]. The majority of prostate cancers histologically are composed of crowded small acinar glands. Benign processes, including partial atrophy and adenosis, may mimic this histologic appearance, creating a diagnostic challenge. Because ERG is highly spe-

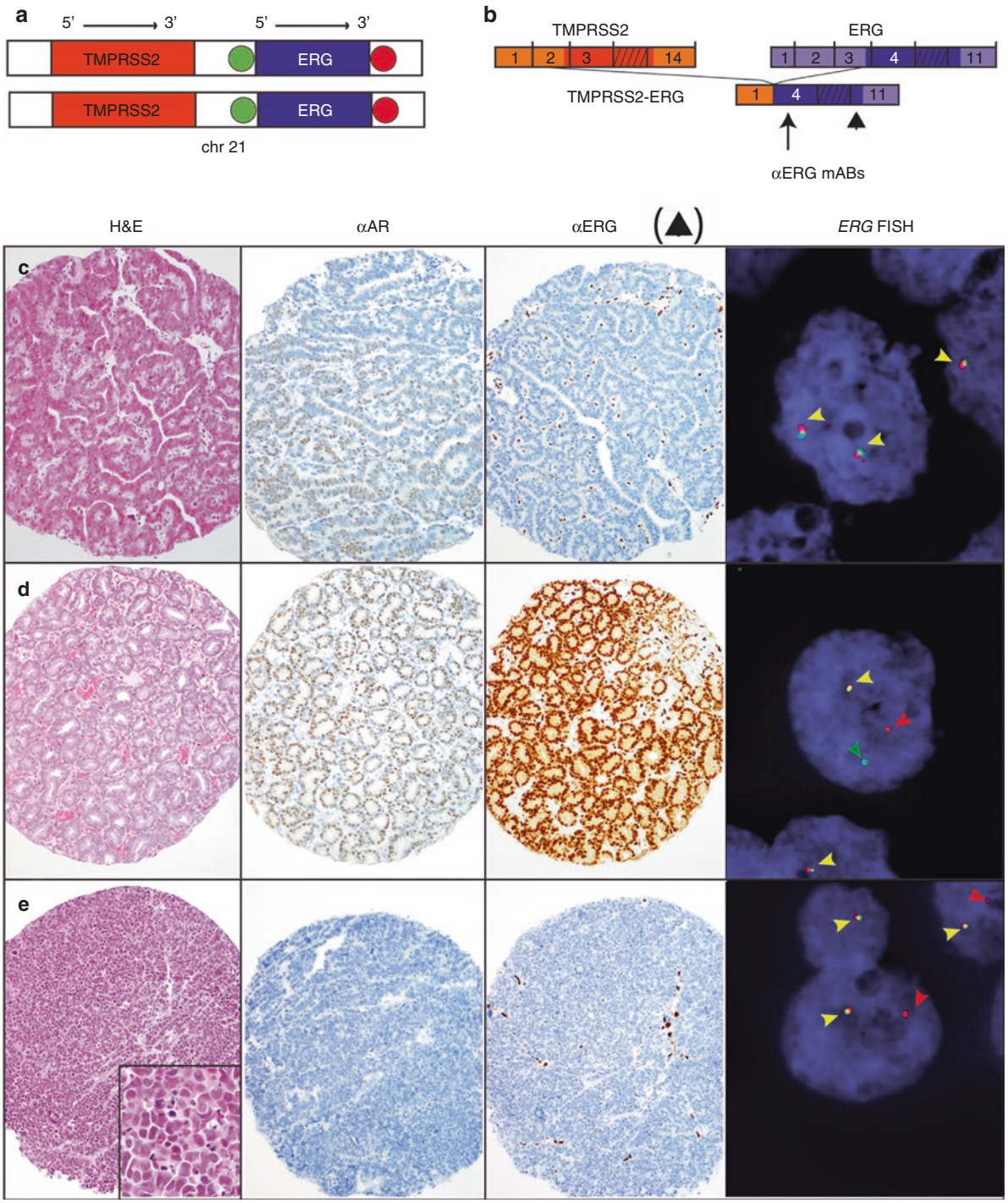


Fig. 30.1 Fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) for detecting *ERG* gene fusions in prostate cancer. (**a**) FISH for *ERG* rearrangements is commonly performed using dual-color split probes flanking *ERG*. (**b**) The transcript structure of *TMPRSS2* and *ERG*, with *boxes* indicating exons, and coding regions in *darker colors*. The structure of the *TMPRSS2-ERG* fusion transcript, which encodes a slightly truncated ERG protein product, is indicated. The advent of monoclonal antibodies (mABs) specific for ERG (clone 9FY CPDR, raised against the N terminus of ERG, *arrow*) and ERG/FLI1 (EPR3864, raised against the *C*-terminus of ERG, *arrowhead*) has enabled IHC-based detection of the *TMPRSS2-ERG* gene fusion product. (**c**–**e**) ERG staining is reflective of ERG rearrangement and androgen receptor (AR) signaling status. Tissue microarray cores from prostate cancer xenografts were characterized by H and E (*left most panel*), IHC for AR and ERG (using EPR3864), and FISH for *ERG* rearrangements (*right most panel*, as in (**a**)). (**c**) Positive AR staining and negative ERG staining in a xenograft with intact *AR* signal and wild-type *ERG* by FISH (*right panel* with all fused [*yellow*] signals). (**d**) Positive AR and ERG staining in a xenograft with intact *AR* signal and *ERG* rearrangement through insertion (*right panel* with separation of one pair of *red* and *green* signals). (**e**) Negative AR and ERG staining in a xenograft with neuroendocrine/small-cell morphology, loss of the *AR* signal and *ERG* rearrangement through deletion (*right panel* with loss of one 5′ [*green*] signal). Original magnification 10× (H and E and IHC), 60× (FISH and inset of (**e**) [*left panel*])

cific for cancer, positive ERG immunostaining in these cases is strong evidence that a focus of crowded glands without basal cells represents cancer and is not a benign mimic (Fig. 30.2). ERG immunohistochemistry has been recommended in best practice guidelines for challenging cases [18], with our recommendation to selectively use it in cases that remain atypical after usual AMACR and basal cell staining.

A small subset of HGPIN lesions on needle biopsy are also ERG$^+$ by IHC (10–30%), which must be considered when using ERG immunostains to support a diagnosis of cancer. Importantly, in prostatectomy sections, ERG$^+$ HGPIN is nearly always located immediately adjacent to ERG$^+$ cancer, while isolated ERG$^-$ HGPIN is frequently observed [12, 16]. Despite this evidence and that provided by Park et al., who showed that ERG-positive status of HGPIN was significantly associated with cancer on rebiopsy in the context of a phase III trial of toremifene (a selective estrogen receptor modulator) vs placebo [19], ERG is not routinely used for risk stratification of HGPIN.

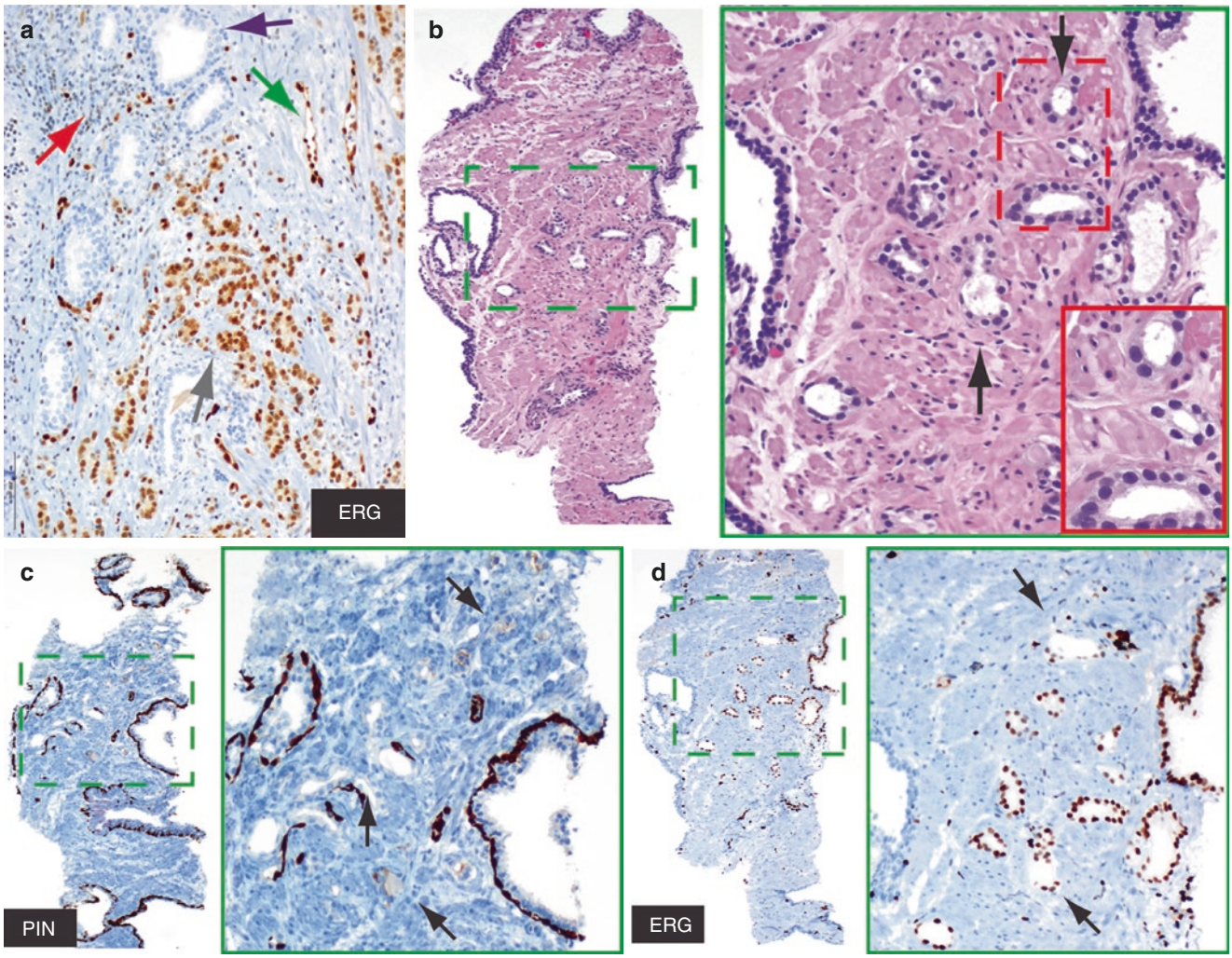


Fig. 30.2 ERG immunohistochemistry (IHC) in prostate cancer. (**a**) Typical ERG staining in a focus of prostate cancer using the EPR3864 antibody. ERG shows strong, diffuse nuclear staining in cancerous glands (*gray arrow*) harboring *ERG* rearrangement. Staining is not present in adjacent benign glands (*purple arrow*). ERG antibodies used for IHC also detect wild-type ERG (and may cross-react with the related ETS protein FLI1), which results in diffuse strong nuclear staining in blood vessels (*green arrow*) and weak staining in tissue lymphocytes (only seen with EPR3864, *red arrow*). Original magnification 20×. (**b**–**d**) Utility of ERG IHC in the diagnostic workup of challenging cases. (**b**) A 12-core needle biopsy had two cores each with a small focus of architecturally and cytologically suspicious glands (*black arrows*, one focus shown). Original magnification 10×, 20× (*green box*) and 60× (*red box*). (**c**–**d**) The core was assessed by IHC for (**c**) basal cell markers (p63 and high molecular weight cytokeratin, brown chromogen) and AMACR (red chromogen) in a cocktail (PIN), and (**d**) ERG (brown chromogen). Original magnification 10×, and 20× (*green boxes*). In our opinion, as high-grade PIN was not in the differential, the presence of ERG staining in the atypical glands (most of which showed only artefactual basal cell staining; see *black arrows* in (**c**)) is consistent with a diagnosis of carcinoma. The other focus on the separate core showed similar staining

Specific Alterations in ETS Gene Fusion-Negative Prostate Cancers

ETS gene fusions are clonal in nearly all cancer foci as demonstrated by FISH and IHC. Combined with their frequency, this provides a basis for basic molecular subtyping of prostate cancer. Importantly, multiple alterations have been found nearly exclusively in ETS fusion-negative (ETS^-) cancers, supporting this molecular subtyping approach. Here, we briefly highlight specific alterations in ETS^- cancers, as well as potentially relevant genes with alterations in both ETS^+ and ETS^- cancers.

SPINK1/SPOP/CHD1

SPINK1 is a trypsin inhibitor first described in the pancreas and has been shown to be overexpressed in ~5–10% of prostate cancers [20, 21]. Across more than 10,000 samples, SPINK1 expression is seen nearly exclusively in ETS^- cancer foci, assessed in part through dual ERG/SPINK1 staining [22, 23]. Importantly, this dual IHC provides a simple assessment for assessing clonality on limited tissue specimens which may be useful in the future, particularly in discontinuously involved cores [24]. Exome sequencing of prostate cancers identified mutations in the cullin ligase *SPOP* in approximately 6–15% of prostate cancers [25, 26], making it one of the most frequently mutated genes in prostate cancer. Missense mutations are clustered in the substrate-binding cleft region, supporting a functional role in prostate cancer. *SPOP* mutations have been identified exclusively in ETS^- prostate cancers, and overlap with $SPINK1^+$ cancers, as well as those harboring alterations in *CHD1*, a chromatin remodeling enzyme, which is deleted or mutated in 5–15% of prostate cancers, is common [25–27]. The clinical significance of the $SPINK1^+/SPOP^{mut}/CHD1^{del}$ subtype of ETS^- prostate cancers is not well defined, and no clear association with differential outcome after prostatectomy has been reported. Importantly, however, recent reports suggest $SPOP^{mut}$ tumors may be resistant to bromodomain (BET) inhibitor-based therapy [28–30].

RAS/RAF/FGFR Family Fusions

RNA-seq-based studies have identified a number of potentially targetable gene fusions involving members of the *RAS*, *RAF*, and *FGFR* family fused typically to androgen-regulated genes, in a total of ~2–5% of prostate cancers, which are exclusively ETS^- [31–34]. Although point mutations in these gene families are very infrequent in Caucasian cohorts (<1%), they may be more frequent in other populations, which correspondingly also have lower ETS^+ prevalence [35, 36]. Given the development of inhibitors that target these alterations in other cancers, we expect that identification of cancers with these alterations may be important for directing therapy in the future.

PTEN

Alterations in classic tumor suppressors and oncogenes, such as *PTEN*, *TP53*, and *MYC*, are relatively frequent in prostate cancer. *TP53* was identified as the most frequently mutated gene through point mutation or indels in castration-resistant prostate cancer (CRPC) [27, 37], and *MYC* is frequently overexpressed in prostate cancer through broad amplifications of 8q, which occur early in prostate cancer development [38]. At present, the clinical utility of these markers is not clear. *PTEN*, a tumor suppressor located on chromosome 10q23, has received intense study in prostate cancer, in large part due to its association with aggressive disease [39–43]. We highlight it here as it provides an important study on the incorporation of prognostic biomarkers into clinical models. PTEN is the most commonly deleted gene in prostate cancer, being deleted or mutated in 15–50% of cases, with *PTEN* alterations being up to three times more common in ETS^+ prostate cancers than ETS^- prostate cancers [25, 26, 37, 40–42].

PTEN deletions have been associated with several clinical and pathologic features of aggressive tumors [39–44]. For example, loss of PTEN by IHC is associated with non-organ-confined disease at prostatectomy and worse recurrence-free survival [41, 45]. Presently, the gold standard for detections of *PTEN* deletions is FISH, which directly detects the abnormality; however, a large multi-institutional study demonstrated that PTEN IHC is likely a cost-effective screening method with reflexive *PTEN* FISH for cases which demonstrate ambiguous or heterogeneous IHC results [40]. Which type of assay best correlates with prostate cancer outcome remains unknown and is an important question for future studies. In contrast to *ETS* status and *SPINK1* expression, which tend to exhibit diffuse expression within a prostate cancer focus, *PTEN* deletions may show considerable heterogeneity, and loss may be seen in only a focal portion of the tumor or nonuniformly in circulating tumor cells [46–48]. Despite numerous studies reporting association of *PTEN* deletion with clinicopathologic parameters of aggressive disease behavior, definite clinical value in assessing *PTEN* has not been shown to date, primarily because studies have not demonstrated that it adds enough prognostic information to change management when combined with standard clinicopathologic data. For example, a recent large study by Krohn et al. with good follow-up data showed that, though statistically significant, *PTEN* deletion status showed a small hazard ratio (1.3) for the prediction of biochemical recurrence in a Cox multivariate regression incorporating *PTEN* deletion status, pathologic tumor stage, serum PSA level, and Gleason

score. In contrast, Gleason score and pathologic stage showed maximum hazard ratios of 6.1 and 5.9, respectively [42].

These findings highlight a crucial consideration in incorporating molecular assays into clinical practice. That is, does the marker add to the best available model in a clinically meaningful manner [49]? For example, in deciding if *PTEN* status should be used to predict outcome (e.g., biochemical recurrence after prostatectomy), evidence of mere independence from other easily assessed parameters at prostatectomy does not necessarily indicate added clinical value, as these parameters can all be easily obtained as a group. Hence, in addition to independence, the new marker must show added predictive value of sufficient magnitude to be clinically actionable when incorporated into the best currently available model (using markers that can be assessed easily). The most relevant method of comparing performance utilizes an area under the curve (AUC) comparison of receiver operator characteristic (ROC) curves or concordance indexes, comparing the current best model to the same model including the new biomarker (*PTEN* in this example). When this analysis was performed on the dataset described by Krohn et al. above, despite the statistically significant independent value of PTEN status, including PTEN (and 6q deletion status) in an optimized multivariate model had minimal impact on the AUC in two scenarios (0.771 vs. 0.767 and 0.782 vs 0.770) [50].

Despite the challenges in identifying prognostic associations, assays to assess the abovementioned single genes are beginning to reveal robust, reproducible molecular subtypes of prostate cancer. The ability to reproducibly identify multiple subtypes should facilitate the stratification of prostate cancer cohorts for biologic behavior and response to therapy (e.g., preclinical evidence suggesting $SPOP^{(mut)}$ tumors are resistant to BET inhibitors) and will help toward the prioritization of subtypes for investigation of novel targeted therapies (even if not directed at these specific biomarkers). The approximate distribution of lesions defining major molecular subtypes, demonstrated by essentially mutual exclusivity (including subtypes defined by comprehensive profiling as described below), is shown in Fig. 30.3.

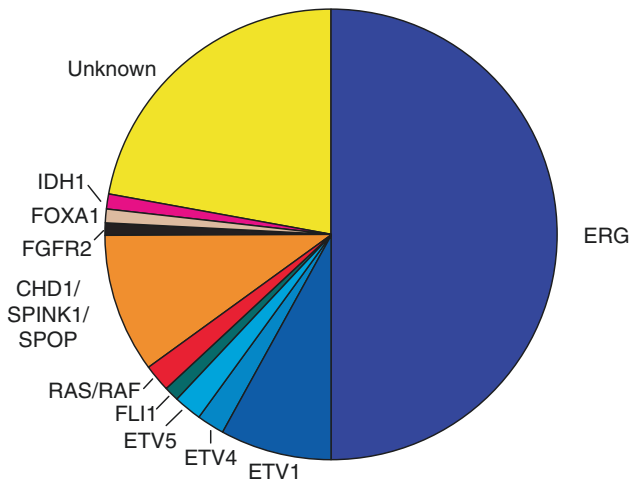


Fig. 30.3 Molecular subtypes of prostate cancer. The approximate distribution of driving molecular lesions in prostate cancers among PSA-screened Caucasians is presented. *ETS* gene fusions (including those involving *ERG*, *ETV1*, *ETV4*, *ETV5*, and *FLI1*) are mutually exclusive with tumors harboring activating gene fusions or mutations in RAS and RAF family members or FGFR2 fusions. A subset of *ETS* and *RAS/RAF* wild-type tumors have *SPINK1* outlier expression, disruption of *CHD1*, and/or *SPOP* mutations. Mutations in *FOXA1* and *IDH1* drive rare molecular subtypes of prostate cancer identified through comprehensive profiling efforts. Approximately 25% of prostate cancers have private or unclear drivers

Applications of the Hereditary Genome in Prostate Cancer: HOXB13 and BRCA

Although in this chapter we have focused on somatic genomic alterations in genitourinary neoplasms (alterations in the cancer genome), hereditary (germ line) mutations are known to predispose to multiple genitourinary neoplasms (e.g., *BRCA2* and prostate cancer or other entities discussed below in the kidney cancer section). As with somatic mutations, the ease of interrogating the genome has led to a large number of studies directed at identifying relatively common hereditary variants that influence the overall risk of developing cancer or the risk of developing aggressive cancer forms. Such hereditary variants, including single nucleotide polymorphisms (SNPs) or copy number variants (CNVs), are commonly interrogated in genome-wide association studies (GWAS), typically through array-based technology. A critical point for the practicing pathologist to consider is that although studies may identify variants that show impressive statistically significant associations (often assessing tens of thousands of cases and controls), such highly significant results do not equate to clinical significance. For example, in an effort to personalize serum PSA cutpoints for prostate cancer detection, Gudmundsson et al. identified six SNPs that were associated with serum PSA levels, each at a significance of $p < 3.0E\text{-}10$ [51]. However, using a model incorporating serum PSA together with these identified SNPs compared to PSA alone only increased the AUC by 0.5% and 1.4%, respectively, in two test cohorts. The limited increase in AUC in part reflects the modest effects of the identified SNPs on PSA levels. Similar findings of limited clinical impact have been frequently identified in several other GWAS studies, suggesting that potential implementation will need to focus on performing necessary clinical utility validation studies [52].

An alternative strategy made feasible by the decreased cost of sequencing, which is allowing the interrogation of thousands of genomes from "case" and "control" patients, is to identify rare variants that may be more strongly associated with cancer development. As an example, through sequencing the 17q21-22 region which had previously been linked to prostate cancer through pedigree analysis of

families with hereditary prostate cancer, Ewing et al. identified the same non-synonymous mutation in *HOXB13* (p.G84E), which co-segregated with prostate cancer in each family [53]. This rare variant (carrier frequency estimate 0.1–1.5%) has consistently shown higher odds ratios (~3–10×) for associations with prostate cancer (including early onset and familial cancer) compared to variants identified by GWAS studies (odds ratios typically 1.1–1.5). Similar findings in other contexts will likely be enabled by WES/WGS efforts [54] and will likely define new "high-risk" criteria that then impact which patients undergo screening and subsequent biopsy. Recently, it was reported that germline mutations in DNA-repair genes are more frequent in men who develop metastatic disease [55]. Because germline mutations may have significant prognostic and/or family planning implications, screening for such mutations at the time of initial diagnosis or even prior to diagnosis may be required in the future, similar to screening for germline mutations in breast cancer patients.

Multigene Panels in Prostate Cancer

Gene expression profiling using DNA microarrays has been instrumental in elucidating gene expression signatures in various cancer types. However, this technique is difficult to perform on routinely processed formalin-fixed paraffin-embedded (FFPE) tissue and is better suited for use with fresh or frozen tissue, which severely limits its clinical utility. Hence, translation of potentially prognostic expression signatures [56] required the development of FFPE compatible assays. For example, Myriad Diagnostics has developed a 40-gene qPCR panel (Polaris™), which may modestly improve on the ability of standard clinicopathologic data to predict death from prostate cancer [57, 58]. This panel, which includes 40 cell cycle genes, has been shown to have potential utility in both radical prostatectomy and needle core biopsy specimens. Genomic Health offers the Oncotype DX® Prostate Cancer Assay, which measures expression of several genes using reverse transcription polymerase chain reaction (RT-PCR) from FFPE needle biopsy tissue [59], similar to the currently available Oncotype DX® Breast Cancer Assay. Likewise, GenomeDx has reported a genomic-clinical classifier model (Decipher™) using expression profiling data from Affymetrix exon microarray analysis of FFPE-isolated RNA. Improved performance compared to a clinical model for predicting clinical recurrence after prostatectomy has been shown with Decipher™. Analogous prognostic utility when performed on biopsy tissue has also been suggested [60–62]. Alternatively, a protein-based assay developed by Metamark Genetics, Inc., was reported to predict for biochemical relapse following prostatectomy when performed on biopsy specimens [63].

For such expression-signature-based classifiers to be routinely used in clinical practice, a number of conditions should be met. First, assays should be able to use standard, clinically relevant samples (e.g., FFPE tissues). Second, they need to be validated in prospective, independent tissue cohorts. If the assay is intended to be used to stratify risk among men under consideration of active surveillance (delaying definitive therapy), it should be applicable to prostate biopsy tissues and must account for potentially unsampled higher grade and/or multifocal prostate cancer, which may not be represented in the diagnostic tissue. Similarly, if assays are to be used to predict "aggressiveness" after prostatectomy (commonly determined by biochemical recurrence), it would be highly desirable that the data generated from the assay would actually influence treatment, such as addition or dose modification of radiation therapy, or addition or modification of the length of antiandrogen therapy. Lastly, it is imperative that the assay adds to the best available clinicopathologic model presently used for predicting the outcome of interest in a clinically meaningful way, rather than only being statistically independent of known clinicopathologic parameters.

Whole-Genome and Whole-Exome Sequencing in Prostate Cancer

Nearly a thousand prostate cancers have now undergone exome sequencing or WGS, including a recent WGS study on 200 non-indolent, localized tumors [27, 37, 64–67]. These studies have confirmed the basic molecular subtypes of prostate cancer defined by nearly mutually exclusive alterations and helped to further elucidate the genomic landscape of prostate cancer. Of note, both targeted and comprehensive sequencing studies have identified ~1% of prostate cancers that lack other known driving alterations as harboring recurrent hotspot mutations in *IDH1* (at residue R132) resulting in global hypermethylation, consistent with a driving oncogenic role [25, 68]. Likewise, a rare subtype of prostate cancer driven by *FOXA1* mutations was also identified through the TCGA study, confirming previous studies identifying this gene as recurrently mutated in prostate cancer [26, 27]. Importantly, WGS has illuminated that prostate cancers harbor large numbers of rearrangements, many involving known cancer-associated genes. A distinctive "closed chain" pattern of rearrangement has also been identified, characterized by complex exchanges within and between chromosomes with no net loss of genetic material, resulting in unique, balanced combinations of chimeric chromosomes [66]. These features were commonly found in *ETS*$^{+}$ prostate cancers, with *ETS* members being involved in the closed chains, further supporting *ETS* gene fusions as defining a distinct molecular subtype. Global Chip-seq also recently supported *ERG* fusion-positive and fusion-negative tumors as distinct molecular subtypes

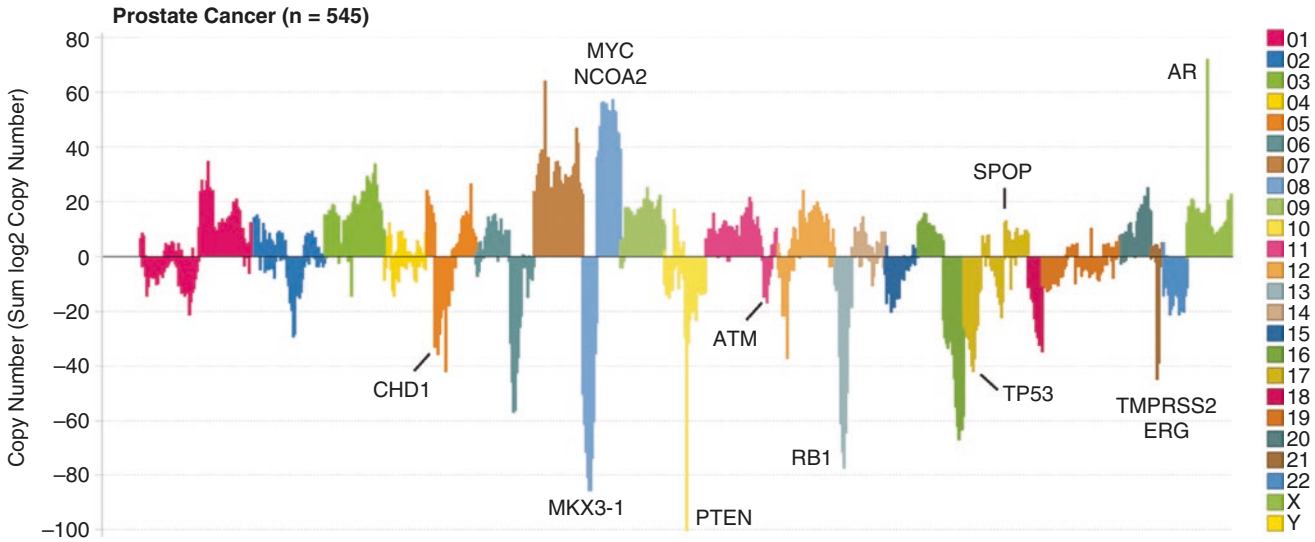


Fig. 30.4 Copy number changes in prostate cancer. Genome-wide copy number profiles from 545 prostate cancers from 4 studies were visualized using the Oncomine Powertools DNA Copy Number Browser (powertools.oncomine.com). The sum of the log2 copy number for each segmented sample is plotted in genomic order (see legend for chromosomes). The locations of genes harboring recurrent copy number gains/losses or mutations are indicated

[69]. Likewise, the Fraser et al. WGS study of non-indolent localized tumors identified recurrent alterations outside of typical coding genes that may have prognostic utility and were not assessed by more common targeted or exome-based approaches. Together, these findings raise several novel potential strategies for prognostic prostate cancer tests [64].

By combining previous array CGH (aCGH) and SNP copy number profiling studies with copy number data from sequencing-based studies, the emerging portrait of the prostate cancer genome suggests that focal gains or losses and point mutations are relatively infrequent (*PTEN* loss and *SPOP* and *TP53* mutations being the most frequent), with rearrangements and broad gains/losses playing important roles (Fig. 30.4). Likewise, whole-transcriptome-based approaches have identified previously uncharacterized transcripts, including noncoding RNAs, as having both functional and potentially prognostic roles in prostate cancer. For example, multiple studies have identified the long noncoding RNA (lncRNA) *SChLAP-1* as one of the most prognostic single-gene markers in prostate cancer [70–76].

Sequencing of Advanced Prostate Cancer for Precision Medicine

Patients with high-grade or advanced prostate cancer are frequently treated with androgen deprivation therapy (medical castration), using drugs such as leuprolide and flutamide. These treatments frequently lead to long-term response, with many patients achieving more than 5 years of disease control. However, all patients will eventually become resistant to standard androgen deprivation therapies, a state known as CRPC. This aggressive state of the disease is associated with high mortality and was previously thought to be androgen independent. However, studies evaluating the mechanisms of castrate resistance have demonstrated intense selection for maintenance of androgen signaling, using varying mechanisms including increased *AR* expression, upregulation of androgen-synthesizing enzymes within cancer cells, development of splice variants of *AR*, and the development of truncated, formerly androgen-dependent proteins that act independently of AR binding [77]. Such observations subsequently led to the hypothesis that many CRPCs are still dependent on androgen signaling. The latter has been borne correct as demonstrated through the efficacy of next-generation antiandrogen signaling drugs, including abiraterone and enzalutamide, in post-chemotherapy CRPC patients.

In other solid tumors, such as lung cancer, molecular testing is often utilized for selection of optimal medical therapy in patients with advanced disease. Given that CRPC is an aggressive disease state amenable to treatment with agents targeting novel aspects of prostate cancer biology, it is likely that molecular testing will play an increasing role in the management of advanced prostate cancer in the near future. The exomes of more than 250 CRPC tumors have now been profiled, with studies consistently showing amplification and mutation of *AR*, loss and deletion of *TP53* and *PTEN*, and mutations in chromatin/histone remodeling genes, including several members of the *MLL* complex and *CHD1*. Of note, a substantial number of patients, despite having clinical resistance to androgen deprivation, showed strong expression of the *TMPRSS2-ERG* gene fusion, indicating active *AR* signaling [27, 37, 67].

Such a profile of androgen signaling assessment including expression of *AR* and target genes, *ETS* gene fusion genomic and transcript status, and presence of *AR* copy number alterations and mutations may be used in the future to select and predict responses to novel antiandrogen drugs.

CRPC profiling has also identified numerous examples of potentially targetable alterations, several of which appear to be enriched in CRPC compared to clinically localized disease. For example, both somatic and germline alterations (mutations and deletions) have been identified in *BRCA1*/*BRCA2*/*ATM*, consistent with disruption of DNA damage repair mechanisms, suggesting potential response to poly(ADP-ribose) polymerase family (PARP) inhibitors [27, 37, 67, 78]. Likewise, rare but potentially targetable alterations in DNA mismatch repair genes (e.g., *MSH2*, *MSH6*), *PIK3CA*/*B*, *R-spondin* (*RSPO2*), *BRAF/RAF1*, *APC*, β-catenin (*CTNNB1*), and *ZBTB16* (*PLZF*) have all been consistently reported across studies. Although assessing "actionability" is challenging, Robinson et al. estimated that 89% of patients in their study harbored a clinically actionable aberration, 65% of whom harbored a targetable mutation in a gene other than *AR* [79]. Given the expanding landscape of targetable mutations in CRPC and improving repertoire of molecular inhibitors, therapy may need to be individualized for small subsets of patients with CRPC. An example of such an effort is the MI-ONCOSEQ study launched at the University of Michigan [79]. In this ongoing study, comprehensive exome and transcriptome sequencing is performed in patients with advanced cancers (including CRPC patients), for whom available treatment options have been exhausted. Patients meet with genetic counselors, and sequencing results are discussed at a multidisciplinary tumor board in an effort to discuss potentially actionable alterations that could be targeted with established or investigational agents. As an example, this program has been used to identify an androgen-driven *FGFR2* fusion in a patient with CRPC, which could be targeted by both FGFR family inhibitors and antiandrogens [33]. Perhaps the most near-term clinically relevant example of the potential translatability of these studies was published by Mateo et al. who showed men with metastatic CRPC harboring DNA-repair defects may respond to PARP inhibition with olaparib (leading to breakthrough designation for this therapy) [80], which may occur in up to 25% of men with CRPC through both germline and somatic alterations [37, 55, 78, 80].

The majority of genomic studies performed on prostate cancer have used fresh or frozen tissue, which may not be always available for widespread clinical use. Sequencing-based assays that are compatible with FFPE tissues are more applicable for routine use. For example, Foundation Medicine has developed a multiplexed sequencing/copy number assay that can be performed from FFPE [81]. Beltran et al. have used this assay to sequence 3320 exons from 182 cancer-associated genes, as well as 37 introns in 14 commonly rearranged genes in a series of 45 prostate cancers including a subset of CRPCs [34]. Their results were consistent with previous studies, showing *PTEN* deletions, *TMPRSS2-ERG* gene fusions, *TP53* mutations, and *MYC* and *AR* amplifications in CRPC in similar frequencies as previously described. Such gene panel-based assay is desirable since it provides useful ancillary genomic data on FFPE specimen in a single assay. More recently, numerous FFPE compatible assays designed for or applicable to prostate cancer have been reported, including a large series from Memorial Sloan Kettering Cancer Center as part of their institutional effort to systematize molecular profiling for patients with advanced cancer [68, 78]. An important caveat is that optimal therapy selection may require assessment of both the genome and the transcriptome, in an effort to understand current driving aberrations. For example, perhaps the most important information to assess in CRPC is *AR* signaling status. Genomic identification of *AR* amplification or mutation, or evidence of *TMPRSS2-ER*G rearrangements, does not necessarily inform on *AR* signaling status [37, 67, 68]. That is, CRPCs with evidence of active androgen signaling (retained expression of *AR* and of *AR*-regulated genes, including *ERG*) may be candidates for more aggressive androgen deprivation therapy, irrespective of genomic status of *AR* and *TMPRSS2-ERG*, while those with androgen signaling loss (and no expression of *AR* or *TMPRSS2-ERG*) may not.

Genomic/Transcriptomic Alterations for Early Detection of Prostate Cancer

Screening for prostate cancer with serum PSA is currently controversial, and several professional organizations in the USA have recommended against general screening. PSA has both sensitivity and specificity limitations and leads to the detection of a significant proportion of small indolent prostate cancers that may not become clinically relevant during a man's lifetime. Hence, intense efforts and resources are being spent on biomarkers for the early detection of prostate cancer in pursuit of a better screening approach.

Prostate Health Index (*phi*), a combination of serum PSA, free PSA, and [−2] pro-PSA, has been FDA approved for risk stratification in the pre-biopsy setting and outperforms serum PSA or free PSA alone [82]. An inherent limitation in this test is that [−2] pro-PSA, the most cancer specific of the three markers, shows at least weak expression in >70% of benign prostate epithelium, consistent with PSA and its derivatives being more prostate specific than prostate cancer specific [83]. An alternative assay including measurement of four kallikrein forms (total PSA, free PSA, intact PSA, and kallikrein-related peptidase 2 [hK2] has been reported to predict cancer risk more accurately than PSA alone and is available as the 4Kscore LDT [84].

Two of the most advanced urine-based biomarkers for prostate cancer are *PCA3* and *TMPRSS2:ERG*. *PCA3* is a noncoding RNA expressed in prostate cancer and HGPIN that is readily detectable in urine [85]. Progensa® is a currently available commercial clinical assay for urine *PCA3* measurement, which calculates a score based on the ratio of *PCA3* RNA to PSA mRNA in urine samples collected following "attentive" digital rectal exam. Progensa® was cleared by the Food and Drug Administration (FDA) for use in determining whether men with a negative prostate biopsy (triggered most commonly by PSA screening) should obtain a repeat biopsy. Using the urine *PCA3* assay score at a cutoff of 25, a negative result has a 90% negative predictive value (NPV) for prostate cancer [86]. Repeat biopsy in men with a negative urine *PCA3* study can thus be avoided. *PCA3* has also been shown to outperform serum PSA in predicting the presence of cancer in first-time prostate biopsies, and incorporation into multivariate models such as the Prostate Cancer Prevention Trial Risk Calculator (PCPTRC) improves performance [87–90].

TMPRSS2-ERG gene fusion transcripts are detectable in the urine of patients with *ERG*$^{+}$ prostate cancer. Like *PCA3*, models including urine *TMPRSS2-ERG* measurement have shown improvement over traditional predictors in multiple cohorts [91–93]. The combined test can be performed on the same urine specimen sent for *PCA3* testing [88, 89]. It includes analysis of serum PSA, urine PCA3, and urine TMPRSS2-ERG. It is termed miProstate Score (MiPS) and is clinically available from the Michigan Center for Translational Pathology lab [89]. Importantly, analysis of ERG by IHC in prostatectomy samples paired with the testing of urine samples demonstrates that the urine *TMPRSS2-ERG* score (calculated similarly to the Progensa *PCA3* score) is strongly correlated to the total *ERG*+ tumor burden (by summing the largest dimension of all tumor foci) [16]. At present there is no known biologic explanation for continued elevation of urine *TMPRSS2-ERG* following prostatectomy except for residual prostate cancer.

Other groups have identified alternative urine biomarkers for potential use in detection of high-grade prostate carcinoma. Van Neste et al. reported that detection of *HOXC6* and *DLX1* mRNA in urine samples identified men at risk of harboring high-grade prostatic carcinoma when combined with traditional clinical risk factors [94]. Likewise, Leyten et al. demonstrated that a three-gene panel including *HOXC6*, *DLX1*, and *TDRD1* may be used to identify men at risk for high-grade prostatic carcinoma, including those with low serum PSA [95]. These markers are clinically available as the SelectMDx urine test. Alternatively, a non-DRE urine-based exosome gene expression assay, using expression of *ERG* and *PCA3* (normalized to *SPDEF*), was reported to be highly sensitive (92%) for high-grade prostatic carcinoma; however, approximately 25% of samples were non-informative in both the training and validation cohorts [96]. Though this exosome gene expression assay demonstrated relatively low specificity (34%), the high negative predictive value (91%) could render the assay of high clinical utility. This assay is clinically available as the ExoDx®Prostate(IntelliScore) test.

Liquid Biopsy for Precision Medicine in Advanced Prostate Cancer

The use of urine samples for *TMPRSS2-ERG* and *PCA3* detection highlights the utility of other sources besides tissues and serum for interrogating the cancer genome. For example, the quantity of circulating tumor cells (CTCs) detectable in the blood of patients with metastatic prostate cancer, and their change in number in response to therapy, can be prognostic [97, 98]. Perhaps more importantly, detection of important genetic driving events, such as *AR* amplification and signaling status, *PTEN* loss, and *TMPRSS2-ERG* rearrangement status, have all been demonstrated in CTCs [99–103], suggesting that such specimens may be utilized for genetic interrogation when tissue is not readily available. Of particular relevance to CRPC due to therapeutics based largely on the AR pathway, discovery of an *androgen receptor* splice variant that lacks the ligand-binding domain but remains constitutively active (*AR*v7) may be particularly relevant for determining therapy in patients with CRPC. Of particular note, detection of *ARv7* in CTCs was reported to be associated with resistance to enzalutamide and abiraterone and decreased progression-free survival [99, 101]. Likewise, patients with detectable pretherapy *ARv7* have superior survival on a taxane-based therapy compared to an AR signaling inhibitor [101].

Similarly, the analysis of circulating free DNA in plasma or urine also allows for interrogation of the cancer genome given the depth of sequencing coverage enabled by current NGS platforms, and preliminary results suggest potential utility in prostate cancer [94–96]. Most notably, multiple groups have shown that *AR* copy number/mutation from circulating cell-free DNA (cfDNA) predicts resistance to second-generation antiandrogen therapies [102–106]. Likewise, more comprehensive cfDNA has been reported, and utility for tracking "reversion" and resistance mutations seems particularly promising, particularly in the context of clinical trials of PARP inhibitors (where *BRCA1/2* reversion mutations are observed after therapy that restore BRCA1/2 function [107–109]). Hence, taken together, the advances made in liquid biopsy approaches and sequencing technologies support an increasing role for non-tissue-based biospecimens in the management of advanced prostate cancer.

The Molecular Pathology of Bladder Cancer

The majority of bladder cancers are urothelial carcinomas, which have widely diverse histopathologic features and clinical behavior. Noninvasive low-grade papillary urothelial carcinoma progresses to invasive cancer in only a minority of cases, while flat urothelial carcinoma in situ has a much higher risk of progression to invasive cancer. Patients with noninvasive carcinoma and those with invasive urothelial carcinoma confined to the lamina propria (pathologic stage pT1) have a relatively good prognosis and are candidates for conservative therapy (referred to as non-muscle invasive bladder cancer [NMIBC]). Patients with invasive disease involving the muscularis propria or beyond (pathologic stage T2 or greater, referred to as MIBC) have high mortality rates and require aggressive therapy.

Current oncogenetic models in urothelial carcinomas suggest two separate molecular routes for NMIBC and MIBC [110, 111]. The first, for NMIBC, is defined by early, activating point mutations in *FGFR3*, RAS family genes, *PIK3CA*, or other oncogenes. A small fraction (~15%) in this group subsequently develop loss of function of *TP53*, *RB1*, *PTEN*, or other tumor suppressor genes, resulting in progression to MIBC which may ultimately lead to metastasis and death. In the second route, loss of function of tumor suppressor genes occurs early, without preceding mutations in *FGFR3*, RAS family genes, or *PIK3CA*. This route is associated with "flat" urothelial carcinoma in situ as a precursor lesion and a higher risk of progression to invasive carcinoma. Hence, these associations between initiating events and subsequent behavior are not absolute.

Single Genes in Bladder Cancer

FGFR3 and TP53

As indicated above, although $FGFR3^{(mut)}$ and $TP53^{(mut)}$ tumors appear to define largely different pathways of tumor progression, many $FGFR3^{(mut)}$ tumors develop mutations in *TP53*. In fact, *TP53* mutations are thought to be a major molecular route by which many $FGFR3^{(mut)}$ tumors progress to muscle-invasive disease. As a single biomarker, *FGFR3* mutational status has good performance for predicting which patients will develop locally advanced disease. Burger et al. have described that *FGFR3* mutation is protective against progression in patients with high-grade pTa/pT1 urothelial carcinoma. Progression-free survival at 5 years was 100% for patients with high-grade disease and *FGFR3* mutation, but approximately 45% for those with high-grade disease without it [112]. Mutational status was not as useful in evaluating urothelial carcinomas as a whole, however, as multivariate analysis including tumor grade showed that *FGFR3* mutational status was not a statistically significant predictor of progression. Importantly, assessment of only *FGFR3* mutations does not incorporate more recently described recurrent *FGFR3* fusions that comprise up to 10% of all *FGFR3* alterations in some series [33, 113–115].

Although the association of *TP53* mutation with aggressive urothelial cancer supports an important functional role, studies of *TP53* as a biomarker have been somewhat disappointing. *TP53* mutational status only modestly improves on standard clinicopathologic data (pathological stage, presence of angiolymphatic invasion, concomitant flat urothelial carcinoma in situ (CIS), patient age, patient gender, and chemotherapy status) for the prediction of disease recurrence and cancer-specific mortality in a cohort of patients with invasive bladder urothelial carcinoma [116]. Similarly, *TP53* mutational status, as a single biomarker, only modestly improves the ability to predict time to progression from high-grade pT1 to muscle-invasive disease (e.g., HR 1.47), and no such predictive value seems to be present in pTa tumors [117]. Despite this lack of demonstrated clinical utility, mutations in *TP53* are diverse in urothelial carcinoma, and more sophisticated *TP53* analysis may yield better results in the future. For example, George et al. have shown both clinical outcome and p53 status by IHC are associated with specific (i.e., location of exon, and single vs. multiple mutations) *TP53* gene mutations. Specifically, single mutations in exon 5 were shown to be associated with wild-type p53 expression by IHC and less aggressive behavior. In contrast, tumors with multiple *TP53* mutations showed p53 overexpression by IHC in most cases and aggressive tumor behavior. Tumors with single mutations in exon 8 of *TP53* also showed p53 overexpression in most cases but intermediately aggressive tumor behavior [118].

Ki-67

Immunoexpression of proliferation marker Ki-67 has shown modest prognostic value. A large trial indicated that Ki-67 independently improved prediction of disease recurrence and cancer-specific survival in multivariate model including pathologic stage, grade, presence of angiolymphatic invasion, concurrent flat CIS, age, and gender [119]. This study also reported that addition of Ki-67 status to a standard multivariable model enhanced its predictive accuracy by 2.9% for disease recurrence and 2.4% for cancer-specific survival. Nevertheless, Ki-67 staining for prognosis is not routinely used for prognosis in clinical practice.

Molecular Subtyping of Bladder Cancer

Lindgren et al. were among the first to highlight the considerable molecular diversity in urothelial carcinomas in two consecutive studies evaluating gene expression profiles by DNA microarray, chromosomal analysis by aCGH, and mutational status of targeted genes in a large cohort of urothelial carcinoma. The first study unveiled the presence of three groups of urothelial carcinomas (Clusters I–III) based on unsupervised gene expression profile hierarchical clustering [120]. Cluster I was characterized by low expression of cell cycle genes, frequent *FGFR3* mutations, lack of *TP53* mutations, and no losses of chromosome 9. The majority (73%) of cases in this group showed low-grade histology. Cluster III showed overexpression of cell cycle genes, no *FGFR3* mutations, frequent *TP53* mutations, and frequent losses of chromosome 9. No cases in this cluster showed low-grade histology. Cluster II was the largest cluster and exhibited features intermediate between Clusters I and III. In their subsequent study involving a subset of the original cohort, the authors were able to group tumors into two groups (MS1 and MS2), based on mutational status of genes in several pathways, including cell cycle-related genes [121]. As is expected from reducing a three-group classifier to one with only two groups, MS1 tumors carried frequent *FGFR3* mutations, but *TP53* mutations were present in several cases. MS2 tumors were likewise enriched in *TP53* mutations, although some had *FGFR3* mutations. MS2 showed much greater chromosomal instability, seen by increased focal genomic amplifications. *HRAS* and *KRAS* mutations were seen equally in MS1 and MS2. More importantly, this study used genes in its clustering to create multigene prognostic models, which were useful in predicting behavior in subgroups of patients. Specifically, a gene signature based on these clusters was able to predict time to metastasis in a group of patients with high-grade disease undergoing cystectomy (cumulative proportion metastasis free 0.90 vs. 0.45 at 60 months).

Subsequently, data from the TCGA project involving comprehensive analysis of 131 muscle-invasive, high-grade urothelial carcinomas demonstrated recurrent mutations in 32 genes, including previously reported mutations in cell cycle regulation, chromatin regulation, and kinase signaling pathways [122]. After profiling a total of 412 MIBC, the TCGA reported updated results including additional significantly mutated genes, a subset of tumors with high mutation load (and improved outcome) driven by *APOBEC*-mediated mutagenesis, and recurrent *PPARG* fusions [123]. Additionally, the initially reported RNA sequencing of TCGA project tumors revealed four expression clusters. Clusters I and II were enriched for luminal breast-like and urothelial differentiation factors (e.g., luminal-like subtype), and cluster III was similar to basal-like breast cancers (e.g., basal-like subtype) [122]. Unbiased NMG consensus clustering of the expanded TCGA MIBC dataset identified five clusters: luminal papillary, luminal-infiltrated, luminal, basal-squamous, and neuronal [123]. Importantly, although only a small subset of the neuronal cluster showed overt neuroendocrine differentiation, in addition to expressing canonical neural and neuroendocrine markers, the neuronal subtype showed frequent mutations in *RB1* and *TP53* (like small-cell neuroendocrine carcinoma of any organ) and had very poor outcome. This subtype will likely be increasingly important to recognize given the prognostic implications despite the lack of usual small-cell morphology. Lastly, comprehensive transcriptional analysis of NMIBC revealed three major subtypes, with types I and II demonstrating luminal-like characteristics and class III demonstrating basal-like characteristics [124].

Importantly, despite the different number and names of clusters and between studies, the vast majority of bladder cancer classification schemes are largely based on robust basal vs. luminal gene expression profiles. For example, several studies have identified a distinct subgroup of urothelial carcinoma with wild-type p53 expression signatures (e.g., p53-like), which in the TCGA classification correspond to the luminal-infiltrated subtype [123, 125, 126]. Of most important clinical interest, RNA-based classifiers have been shown to improve upon the performance of clinical models for predicting post-cystectomy urothelial carcinoma recurrence [127], and the basal-like, luminal-like, and p53-like/luminal-infiltrated subtypes of urothelial carcinoma may have therapeutic and prognostic importance with consistent subtype associations observed across studies. For example, luminal tumors consistently show better outcome than basal-like tumors, although luminal tumors may be targetable by *FGFR3* inhibitors [123, 125, 126, 128, 129].

Multigene Panels in Bladder Cancer

Markers predicting response to neoadjuvant chemotherapy in urothelial bladder carcinoma are also actively sought. Unfortunately, the vast majority of reported predictors suffer from the same limitations as those described for prostate, including lack of validation in independent sample sets, and they have not been shown to add to the best available clinical models. As large efforts like the TCGA identify additional molecular subtypes and the ability to incorporate DNA-based lesions, we anticipate that the predictive ability of such assays will likely improve and will be tested in prospective trials. For example, an active clinical trial using the CO-eXpression ExtrapolatioN (COXEN) score is underway to assess predictability of the COXEN score [130, 131] to direct whether patients with localized, muscle-invasive urothelial carcinoma should receive gemcitabine and cisplatin or dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin (NCT02177695). Most recently, a multigene

expression signature applicable to FFPE specimens was recently reported to be able to perform luminal vs. basal-based subtyping and demonstrate that luminal tumors had the best overall survival, with and without neoadjuvant chemotherapy, while the claudin-low subtype showed poor overall survival regardless of therapy; basal tumors demonstrated the most improvement in overall survival with neoadjuvant therapy [129]. If prospectively validated, such approaches may have both prognostic and predictive utility.

Whole-Genome and Whole-Exome Sequencing in Bladder Cancer

Fewer whole-exome and WGS studies have been performed on urothelial carcinoma compared to prostate cancer. Prior to publication of TCGA data, the largest study of WES/WGS in urothelial carcinoma included 99 tumors [114, 132]. Consistent with previous studies, frequent mutations were seen in *TP53*, *RB1*, *HRAS*, *FGFR3*, and *KRAS*. A high (59%) mutation rate was identified in chromatin remodeling genes, a finding not previously described, and recurrent alterations in *STAG2* were specifically described. Mutations of chromatin and histone modifiers is likely to be a near-uniform finding in cancer, as such genes are commonly mutated in cancers of diverse histologic origin, including the prostate and kidney. The above described TCGA studies confirmed previous single-gene studies, including demonstration of recurrent translocations involving *FGFR3* and *TACC3* on chromosome 4 as well as those involving *PPARG*, frequent activation of the MAP kinase pathway, as well as delineation/confirmation of the above listed expression-based subtypes [122].

Of note, WGS has been performed on a patient with metastatic urothelial carcinoma demonstrating a durable response to mTOR inhibitor agent everolimus [133]. Somatic loss-of-function mutations were identified in *TSC1*, a known regulator of *mTOR* pathway activation. Subsequent targeted mutational analysis in a larger cohort (109 bladder cancers) revealed an 8% *TSC1* mutation rate that also correlated with response to everolimus. This study demonstrated a potential method (and rationale) for identifying genes associated with outlier therapy response, as well as identified a potential therapeutic approach for a subset of patients with urothelial carcinoma.

Copy number alterations in bladder cancer points to a more focal copy number gain and loss compared to other genitourologic malignancies (Fig. 30.5). *CDKN2A* is the most frequently deleted gene in bladder cancer. Loss of this gene locus on chromosome 9p21 has been exploited in urine multitarget FISH assays such as UroVysion. Unlike prostate and kidney cancers, bladder cancer shows focal high-level amplifications of receptor tyrosine kinases such as *ERBB2* and *EGFR* in a mutually exclusive manner (Fig. 30.6) [122]. The amplification levels of the latter are similar to those encountered in breast and lung cancers, where such driver genetic aberrations are targetable. However, despite promising preclinical data [134], initial trials targeting these alterations in bladder cancer have been disappointing [135–137]. Results from recruiting or recently completed trials using more stringent patient selection and novel agents targeting ERBB2 and/or EGFR are eagerly awaited.

Multigene Panels in Bladder Cancer for Early Detection/Recurrence Monitoring

Urothelial carcinomas harbor frequent gains of chromosome 3q, 7p, and 17q and frequent deletions of the 9p21 locus harboring *CDKN2A*. As indicated above, multicolor FISH-based urine assay, UroVysion, has been developed based on these abnormalities. Several studies have demonstrated the utility of multitarget FISH assays in urine samples with the goal of predicting recurrence of urothelial carcinoma in the surveillance setting with an overall sensitivity of 81–87% and a specificity of 92–97% [138, 139]. These sensitivities are superior to standard urine cytology, and the specificity of multitarget FISH assays has also been consistently shown to be comparable to standard urine cytology [140, 141]. The UroVysion assay has also been approved by the US FDA for primary screening for urothelial carcinoma in patients with hematuria [142]. Urine microsatellite, methylation, targeted mutational analyses, NGS, and digital droplet PCR have also shown potential in diagnosis/monitoring of urothelial carcinoma [142–144]. Widespread use of such approaches will require rigorous clinical utility studies like those for UroVysion.

Immunotherapy Markers in Bladder Cancer

Recently, atezolizumab, which inhibits PD-L1 in tumor cells and T-cells, received FDA approval in May 2016 for treatment of patients with chemotherapy-resistant advanced or metastatic urothelial carcinoma [145], and five anti-PD-L1/PD-1 therapies are now approved in chemotherapy refractory/ineligible metastatic urothelial carcinoma [146]. In addition to IHC-based assessment of PD-L1 expression in tumor microenvironment, additional biomarkers to predict response to checkpoint inhibitors are being explored including tumor mutational burden and molecular subtypes [145].

Molecular Pathology of Kidney Cancer

Renal cell carcinomas (RCCs) account for the majority of kidney malignancies. Clear-cell RCC is the most common histologic type followed by papillary and chromophobe RCC. We will focus primarily on clear-cell RCC in this

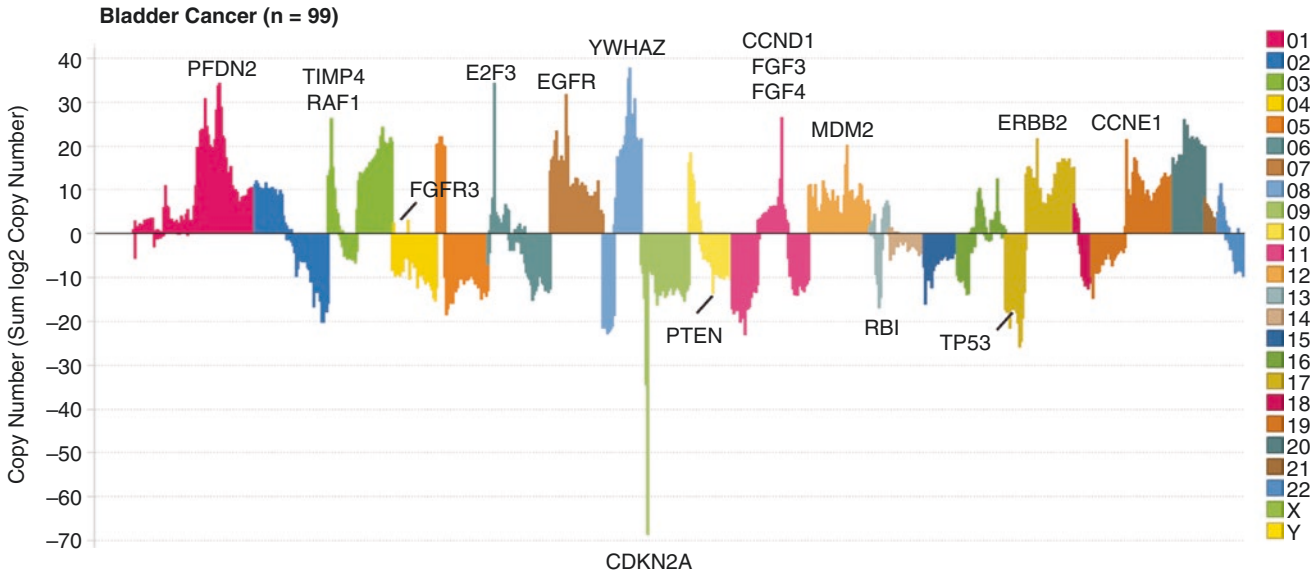


Fig. 30.5 Copy number changes in bladder cancer. Genome-wide copy number profiles from 99 bladder cancers from a single study (TCGA) were visualized using the Oncomine Powertools DNA Copy Number Browser. The sum of the log2 copy number for each segmented sample is plotted in genomic order (see legend for chromosomes). The location of genes harboring recurrent copy number gains/losses or mutations are indicated

chapter, and readers interested in molecular profiling of papillary and chromophobe RCC are referred to published TCGA data on those entitites [147, 148].

Single Genes in Kidney Cancer

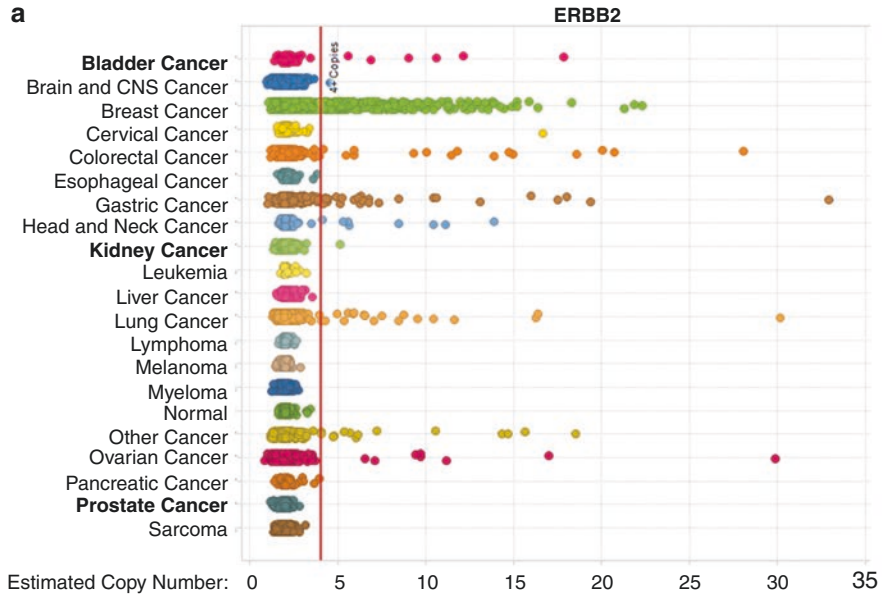
VHL

Nearly all clear-cell RCCs harbor abnormalities in the *VHL* gene located on chromosome 3p25, and inactivation of *VHL* is the single most common abnormality in clear-cell RCCs [149]. Somatic mutations, deletions, insertions, copy loss, and epigenetic silencing have all been observed [149, 150], with lack of any VHL alteration arguing against a strict diagnosis of clear-cell renal cell carcinoma. Loss of 3p is also seen in 80–90% of clear-cell RCCs and was one of the first alterations identified in clear-cell RCC [149, 151] (see Fig. 30.7). *VHL* encodes a member of the ubiquitin ligase complex, which participates in the controlled degradation of numerous cellular proteins, including HIF1α and HIF2α. HIF1 and HIF2 are a transcription factors (dimers composed of α and β subunits) that mediate the cellular response to tissue hypoxia. When activated, HIF1 and HIF2 stimulate angiogenesis and cellular proliferation via increased production of VEGF, erythropoietin, PDGF, TGF-α, and other mediators including the diagnostically useful marker carbonic anhydrase IX (CAIX) [152].

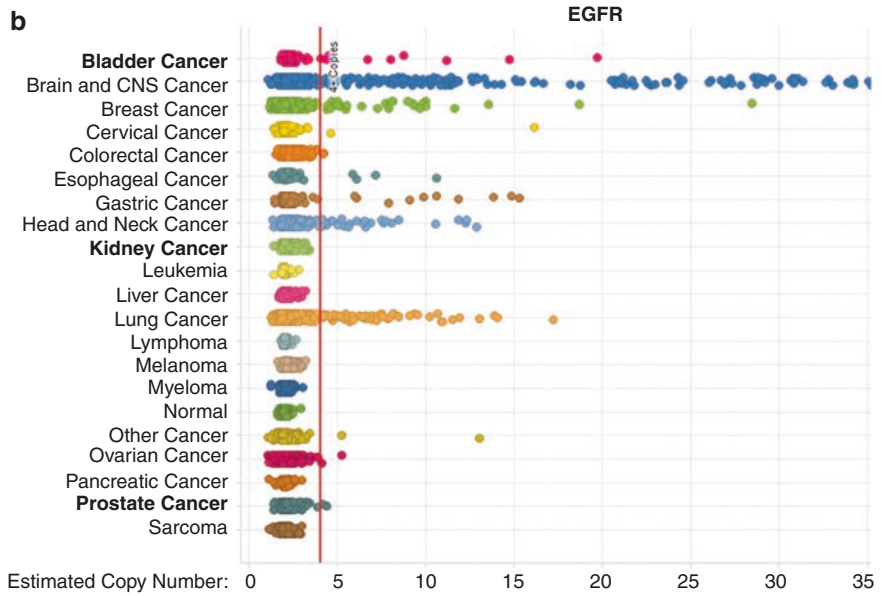
Deletions of 3p can be detected reliably with FISH. Importantly, 3p deletions are not seen in other variants of RCC, such as papillary RCC and chromophobe RCC. Therefore, 3p loss detectable by FISH is useful as a specific diagnostic biomarker for clear-cell RCC in cases that pose diagnostic difficulty [153]; however, as alterations besides 3p deletion can inactivate VHL, it is not entirely sensitive. Of note, tumors that lack VHL alterations may represent distinct diagnostic entities. For example, a morphologically distinct subset of tumors that have clear-cell morphology and thick fibrous bands coursing through the tumor were identified due to their lack of 3p loss (and *VHL* mutations), and instead these tumors have mutations in *TCEB1* [154].

PBRM1

Exome sequencing study by Varela et al. identified *PBRM1*, which encodes a chromatin remodeling enzyme, as another commonly mutated gene in clear-cell RCCs, with truncating mutations seen in 41% (92/227) of cases [155]. The majority of mutations were either indels or nonsense mutations. *PBRM1* is a member of the *SWI/SNF* chromatin remodeling complex involved in the maintenance of DNA stability and the regulation of gene transcription. *PBRM1* resides on chromosome 3p, centromeric to *VHL*. Hence, in addition to *VHL*, losses of 3p usually result in loss of one copy of *PBRM1*. Similar to *VHL*, homozygous inactivation of *PBRM1* via concomitant 3p loss and *PBRM1* mutation in the second allele is a common occurrence in clear-cell RCC. Also similar to *VHL*, nearly all $PBRM1^{(mut)}$ cases express the "hypoxia" phenotype [155].



Cancer Type	> 4 Copies (%)	Total Cases
Breast	10.1%	3470
Gastric	7.2%	417
Bladder	6.1%	99
Other	2.9%	488
Colorectal	1.8%	849
Head & Neck	1.1%	760
Ovarian	1.1%	857
Lung	1.0%	1837
Cervical	0.5%	186
Kidney	0.2%	635
Brain & CNS	0.1%	1454
Esophageal	0.0%	61
Leukemia	0.0%	240
Liver	0.0%	301
Lymphoma	0.0%	332
Melanoma	0.0%	124
Myeloma	0.0%	575
Normal	0.0%	5671
Pancreatic	0.0%	58
Prostate	0.0%	566
Sarcoma	0.0%	227



Cancer Type	> 4 Copies (%)	Total Cases
Brain & CNS	22.1%	1455
Bladder	9.1%	99
Esophageal	6.6%	61
Head & Neck	3.6%	760
Lung	3.2%	1840
Gastric	2.4%	417
Cervical	1.1%	186
Breast	0.8%	3478
Other	0.4%	488
Prostate	0.4%	566
Colorectal	0.2%	849
Ovarian	0.2%	859
Kidney	0.0%	635
Leukemia	0.0%	240
Liver	0.0%	301
Lymphoma	0.0%	332
Melanoma	0.0%	122
Myeloma	0.0%	575
Normal	0.0%	5671
Pancreatic	0.0%	58
Sarcoma	0.0%	227

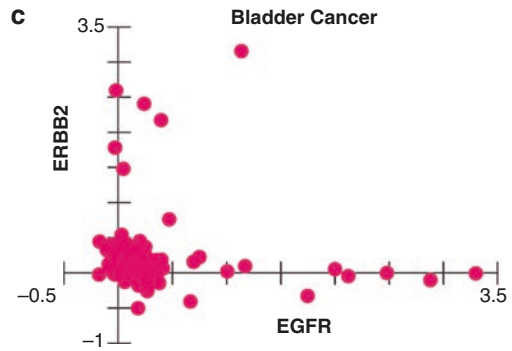


Fig. 30.6 High-level copy gains in *EGFR* and *ERBB2* in bladder cancer. (**a**) Estimated copy number for *ERBB2* across profiled cancer types, visualized using the Oncomine Powertools Integrated Gene Viewer (powertools.oncomine.com). The *red line* indicates four copies. The table shows a frequency of >4 copies across all profiled cases present in the Oncomine database. Genitourinary cancers are bolded. (**b**) As in (**a**), except for *EGFR*. (**c**) *ERBB2* and *EGFR* show mutually exclusive high-level copy number gains in bladder cancer. *ERBB2* and *EGFR* copy number data (log2) for all 99 profiled TCGA bladder cancers downloaded from the Oncomine database (www.oncomine.com) are plotted

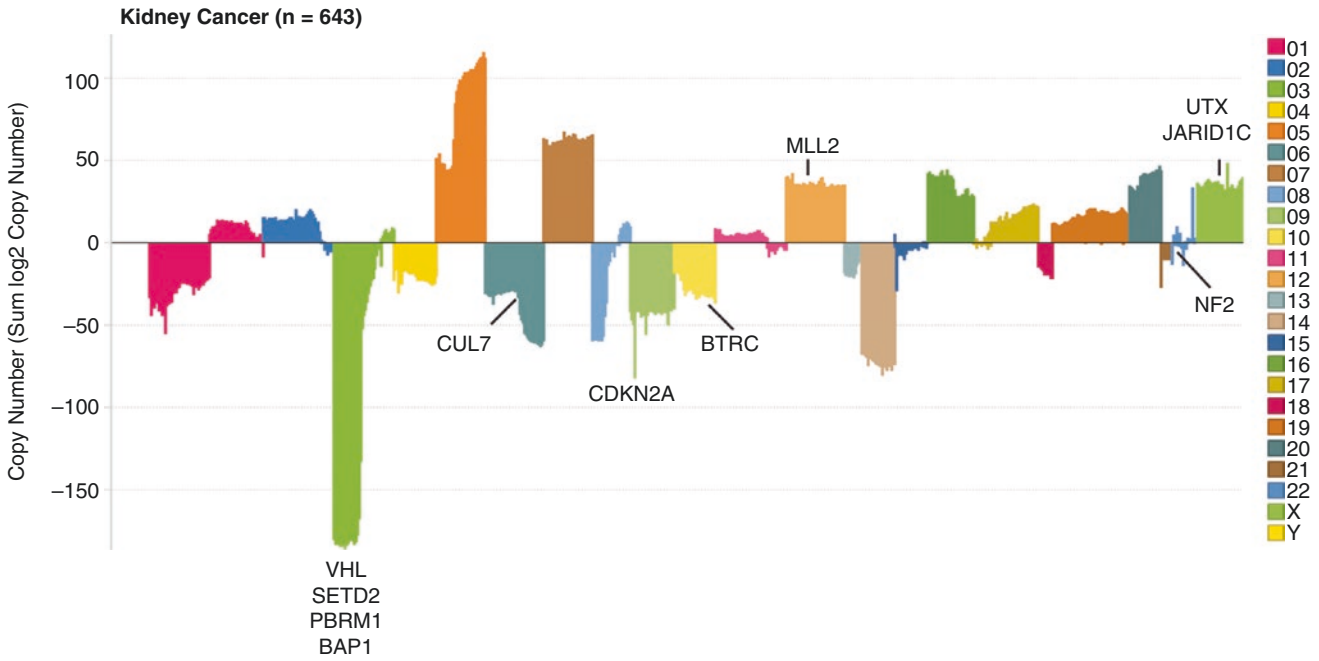


Fig. 30.7 Copy number changes in kidney cancer. Genome-wide copy number profiles from 643 kidney cancers (predominantly clear-cell carcinoma) from 2 studies were visualized using the Oncomine Powertools DNA Copy Number Browser. The sum of the log2 copy number for each segmented sample is plotted in genomic order (see legend for chromosomes). The location of genes harboring recurrent copy number gains/losses or mutations are indicated

Inherited Familial RCC Syndromes and Single-Gene Mutations in Non-Clear-Cell RCC

Although we have focused predominantly on somatic alterations in this chapter, assessment of the germline for variants predisposing or causally associated with specific kidney cancer types will likely become more commonplace with the increasing recognition of hereditary kidney cancer types and the increased use of routine comprehensive genetic/genomic testing [156, 157]. As just one example, hereditary leiomyomatosis and renal cell carcinoma (HLRCC) syndrome is caused by a germline mutation in the fumarate dehydrogenase gene (*FH*) and is associated with an increased incidence of RCCs and cutaneous leiomyomata [158]. Although the syndrome predisposes to several RCC histologic types, a peculiar type of RCC (FH-deficient RCC) with papillary architecture, prominent "cherry-like" nucleoli, and perinucleolar clearing has been associated with this germline defect. Antibodies against FH are available and are useful diagnostically in revealing its loss in FH-deficient RCC.

Single-gene mutations have also been described in sporadic non-clear-cell RCCs. For example, sporadic papillary RCCs have been shown to harbor activating *MET* mutations in a small fraction of cases, similar to their hereditary counterparts [159]. Recurrent gains of chromosomes 7 and 17 are also frequently identified in sporadic papillary RCC, as is overexpression of the protein product of *MET*, with ongoing clinical trials specifically assessing MET inhibitors in this tumor type [160].

Finally, a special type of RCC is defined by recurrent chromosomal translocations involving members of the *MiTF-TFE* genes family [161, 162]. The *MiTF-TFE* family of renal carcinomas contain balanced translocations of one member of the basic helix-loop-helix zipper transcription factors, most commonly *TFE3* on chromosome Xp11 and *TFEB* on chromosome 6p21. The most common translocation partners are *ASPL* (chromosome 17q25) and *PRCC* (chromosome 1q21). Tumors of this type frequently have a unique papillary architecture with abundant clear cytoplasm, are high grade, disproportionately affect younger patients, and are associated with high rates of metastasis. Diagnosis is usually confirmed using well-established *TFE3* and *TFEB* break-apart FISH assays; immunohistochemical assessment of TFE3, TFEB, and cathepsin K proteins has also been used, albeit with less accuracy [163–167]. Of note, potentially distinct tumor types with amplification of *TFEB* have also been recently described [168, 169].

Whole-Genome and Whole-Exome Sequencing in Renal Cell Carcinoma

In addition to identification of mutation in *PBRM1*, exome sequencing studies of clear-cell RCC have shown frequent mutations in other chromatin–/histone-modifying genes. Among the most frequently identified are the histone methylases *MLL2* and *SETD2*, and the histone demethylases *UTX*

and *JARID1C* (*KDM5C*) [155, 170–172]. Collectively, mutations in these genes are seen in nearly 15% of clear-cell RCC. Mutations are commonly missense, splice site, or indels, consistent with a tumor-suppressive function. Histone modification is a primary method of gene expression regulation, and as with urothelial carcinoma and prostate cancer, aberrant modification of histone complexes appears to contribute to abnormal gene expression in a sizable fraction of clear-cell RCCs. A recently published molecular characterization of clear-cell RCC by TCGA Research Network has corroborated the importance of chromatin remodeling genes in clear-cell RCC and has further demonstrated mutations in the SWI/SNF chromatin remodeling complex (which includes *PBRM1*) may have profound effects on numerous other pathways [149]. The TCGA study also found recurrent mutations in the PI(3)K/AKT pathway, as well as frequent mutations in genes involved in cellular metabolism.

A smaller group of clear-cell RCCs harboring truncating mutations in the tumor suppressor *NF2* has also been identified by exome sequencing [171]. *NF2* germline mutations are associated with neurofibromatosis type 2, a syndrome characterized by predisposition to benign and malignant peripheral nerve sheath tumors, meningiomas, and gliomas. While the majority of tumors with *SETD2* or *JARID1C* mutations in this study showed either *VHL* mutations or the hypoxia phenotype, none of the clear-cell RCCs harboring *NF2* mutations appear to contain *VHL* mutations or show the hypoxia phenotype [171]. These findings suggest that *NF2*-mutated clear-cell RCC may represent a distinct molecular subtype (analogous to *TCEB1*-mutated RCC described above) [173].

As mentioned above, the ubiquitin-mediated proteolysis pathway (UMPP) includes *VHL* and functions in the controlled degradation of many cellular proteins, including HIFs. Recurrent mutations in several members of the UMPP have been identified in clear-cell RCCs. Guo et al. sequenced all 135 genes in the UMPP in a set of 98 clear-cell RCCs and found mutations in at least 1 member of that pathway in 50% of tumors [174]. Comprehensive analysis of the clear-cell RCC genome demonstrates few focal aberrations, with nearly universal broad loss of chromosome 3p, as well as broad gains and losses of other chromosomes (Fig. 30.7).

Molecular Prediction of Treatment Response in Kidney Cancer

Traditionally, chemotherapy has not proven to be effective in clear-cell RCC. Immunotherapy with interleukin-2 (IL-2) and interferon-α (INF-α) is effective in a small subset (~10%) of metastatic clear-cell RCCs but is limited by its high toxicity profile. More recently, IL-2-based immunotherapy has been largely replaced by newer treatment strategies targeting VEGF/VEGFR, including small-molecule inhibitors of VEGFR and other tyrosine kinases (e.g., sorafenib and sunitinib) and monoclonal antibodies directed against circulating VEGF (e.g., bevacizumab). Targeted therapy response rates in clear-cell RCC are higher than seen with traditional immunotherapy (~40% vs. 10%, respectively), although toxicity is still an issue [175].

Predictors of response to VEGF-targeted therapy are beginning to emerge [176]. Choueiri et al. have recently showed that loss-of-function mutations in *VHL* were associated with improved response to VEGF-targeted therapy (51% vs. 31% response rate) [177]. In contrast, *VHL* inactivation by other means was not associated with response, as all pooled patients with *VHL* inactivation did not show a statistically significant difference in response rate compared to those with wild-type *VHL*. CAIX expression in clear-cell RCCs has been shown to variably correlate with response to several agents; however, like most other predictive biomarkers in renal cell carcinoma, prospective validation is lacking [176].

A subset of clear-cell RCCs demonstrate overexpression of members of the mTOR pathway which induces cellular proliferation and represses apoptosis. Phosphorylated Akt and S6 (phos-Akt and phos-S6) may be used as markers of pathway activation. *PTEN* is an upstream suppressor of the mTOR pathway. Small molecular inhibitors of the mTOR pathway have been developed (e.g., temsirolimus, everolimus) and have showed improved progression-free survival in a group of patients who progressed on VEGF inhibitor therapy [178]. Despite efforts to identify potential predictive biomarkers of response, none have been prospectively validated.

Newer treatment modalities, including cabozantinib (a tyrosine kinase inhibitor) and nivolumab (a PD-1 checkpoint inhibitor), have been shown to improve survival as compared to everolimus in patients with advanced RCC [179–181]. In a statistical comparison of the cabozantinib and nivolumab trials, cabozantinib initially outperformed nivolumab with cabozantinib having longer overall survival in the first few months of treatment. Nivolumab, however, appeared more effective in the long term, and the authors postulated that patients with a poor prognosis may benefit more from cabozantinib, while nivolumab would be more optimal for patients with a better prognosis, though more evidence is needed [182]. Recently, it was shown that cabozantinib may be better even as a first-line therapy over sunitinib in patients with intermediate- or high-risk metastatic RCC. In the future, it is likely that RCC treatment will be guided by molecular and immunohistochemical markers such as PD-L1 expression and expression of tyrosine kinases. These agents may move earlier in the treatment course (including potentially adjuvant therapy) [175, 183]. Of interest, although clear-cell RCC has a low overall mutation rate (unlike almost all other PD-L1 responsive tumor types), a recent study proposed that efficacy in clear-cell RCC is due to the high number of neoantigens induced by the high number (and proportion) of frameshifting insertion/deletion mutations in this tumor type [184].

Molecular Pathology of Testis and Penile Cancers

Tumors of the testis and penis have received considerably less attention regarding genomic analysis compared to those of the prostate, bladder, and kidney. However, molecular perturbations described in both offer unique insights into tumor behavior and treatment response. The following is a brief discussion of some salient molecular alterations in these two organs.

Germ cell malignancies account for the great majority of testicular tumors, with seminoma being the predominant histologic type. As expected in germ cell lineage, seminomas express markers of totipotency, including *OCT 3/4* and *NANOG* [185, 186]. They also frequently show 12p gains, most commonly as isochromosome 12p [187]. In contrast to other solid tumors, *TP53* mutations are rare [188–190]. Seminomas, as well as the majority of other testicular germ cell tumors, are extraordinarily responsive to chemotherapy with DNA-damaging agents such as cisplatin, etoposide, and bleomycin and radiation with cure rates well exceeding 90%. This is thought to be related to a strong propensity for germ cells to undergo apoptosis as a result of DNA damage [188]. Although relatively uncommon, chemotherapy resistance is well-known and has been associated with microsatellite instability (MSI) and *BRAF* mutations [191]. As the mechanism of chemotherapy resistance is largely unknown, Taylor-Weiner et al. recently performed whole-exome and whole-transcriptome sequencing of precursor, primary, and chemoresistant metastatic germ cell malignancies [190]. They found that activating *KRAS* mutations were acquired in the transition of precursor to primary disease, and chemoresistant tumors possessed a greater number of copy number events and lost the pluripotency and apoptosis regulators *NANOG* and *OCT3/4*. These findings support the hypothesis that chemoresistance is driven by loss of ability to undergo apoptosis in this normally mitochondrially "primed" tumor [188, 190].

Squamous cell carcinoma (SCC) represents the majority of malignant penile neoplasms. Like their cervicovaginal and oropharyngeal counterparts, a subset of penile SCC is associated with high-risk human papillomavirus (HPV) infection. As in other sites, high-risk HPV DNA becomes integrated into the host cell genome, leading to production of the viral proteins E6 and E7. These induce cellular proliferation by inactivation of *TP53* and *RB1*, which are associated with increased expression of p16 (*CDKN2A*). The immunophenotype of many HPV-associated tumors is, therefore, $TP53^{-}$/RB^{-}/$p16^{+}$, and p16 may be used as a surrogate marker for HPV infection. As in the oropharynx, $p16^{+}$ penile SCC has been shown to have better prognosis than $p16^{-}$ SCC, demonstrating improved cancer-specific survival [192, 193]. Interestingly, penile SCCs with basaloid morphology are $p16^{+}$, but have a worse prognosis than conventional penile SCC [193, 194], which stands in contrast to the improved prognosis seen in SCC with this morphology in the oropharynx.

Two recent studies analyzed over 100 FFPE tumor samples using comprehensive targeted DNA sequencing to understand the molecular drivers of penile SCC. Frequent alterations were identified in *TP53*, *CDKN2A*, *PIK3CA*, *MYC*, *NOTCH1*, *HRAS*, and *SOX2* [193, 195], similar to SCCs from various other sites [196]. In the penile SCC study from the University of Michigan, there were no significant associations between an individual gene's mutation status and tumor grade, stage, or histology, and both studies found high-risk HPV in only a minority of cases [193, 195]. Importantly, in the University of Michigan cohort, four cases demonstrated *EGFR* amplifications and one case had *CDK4* amplification, genes for which targeted therapies are available. *EGFR* amplification status, however, did not correlate with EGFR protein expression status as all patients were positive for EGFR protein IHC. Furthermore, *EGFR* amplification status was not uniformly conserved between primary and metastatic tumors, and activating *HRAS* mutations were frequent. These findings suggest treatment with an EGFR inhibitor may have utility in treatment of advanced penile SCC, but further studies elucidating the role of *EGFR* amplifications in pathogenesis and the effect of *HRAS* mutations are necessary. Lastly, the PD-1/PD-L1 immune checkpoint has also been investigated as a potential target in penile SCC, and it was demonstrated that the majority of non-HPV-related penile SCCs express PD-L1 as assessed with IHC, providing rationale for anti-PD-1- and anti-PD-L1-based therapies in penile SCC [197].

Conclusion

Our increased understanding of the cancer genome is beginning to impact patients with all types of malignancies of the genitourinary tract. This includes identification of novel groups of "high-risk" patients who will then undergo biopsy after abnormal screening or imaging, to patients with advanced cancer who may undergo comprehensive whole-genome/whole-transcriptome sequencing to optimize therapy. Immunohistochemistry will remain an important methodology to rapidly interrogate expression of markers that may identify specific molecular subtypes for diagnosis, prognosis, or prediction. However, we anticipate that genitourinary pathologists of the future will utilize additional diverse assays, including FISH, qRT-PCR, and NGS, in order to meet the demands for personalized patient care. An overview of the biospecimens, assays, and clinical indications discussed above is illustrated in Fig. 30.8. In this chapter, we provide an overview of the biomarkers and techniques that will likely have the most immediate clinical impact in genitourinary pathology. Nevertheless, the pace of genomic discovery almost guarantees that the most important biomarkers and techniques to impact clinical practice may be yet to be discovered.

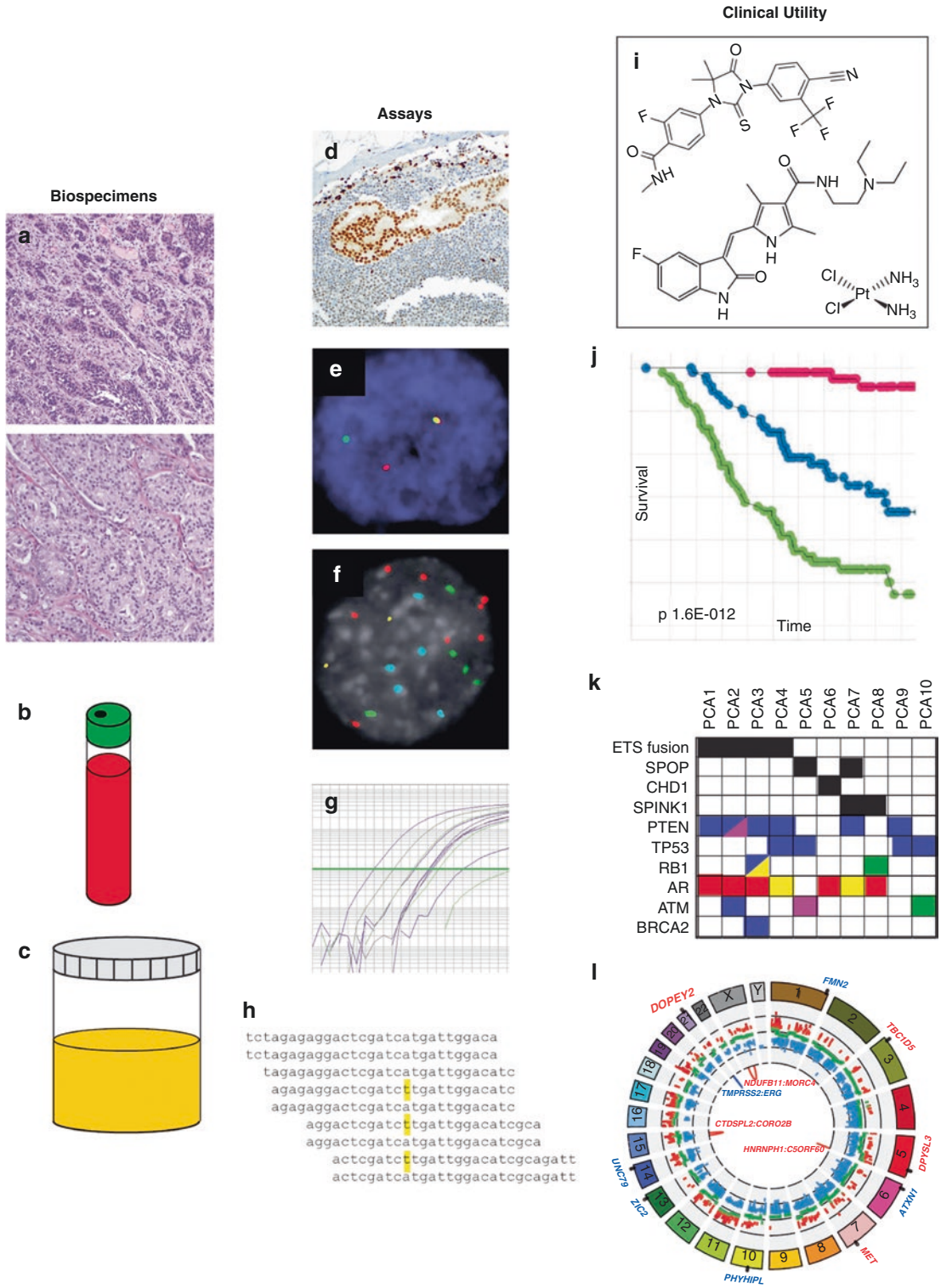


Fig. 30.8 Current and future genomic applications in genitourinary pathology. Examples of biospecimens, assays, and the clinical utility of genomic applications in genitourinary pathology are shown. A variety of specimen types can be utilized for interrogating the cancer genome, including (**a**) routine tissue specimens, (**b**) blood (as a source of protein, circulating tumor cells, or free nucleic acids), and (**c**) urine (as a source of protein, tumor cells, or free nucleic acids). These samples can be used for a variety of analyses, including (**d**) IHC (an ERG-positive prostate cancer lymph node metastasis is shown), (**e**) FISH for rearrangements (split probes showing a *BRAF* rearranged prostate cancer cell), or (**f**) copy number (UroVysion FISH in a bladder cancer cell), gene expression or copy number profiling by microarrays or (**g**) qRT-PCR, or (**h**) NGS of the cancer genome/transcriptome. Such assays are applicable in numerous clinical scenarios, including (**i**) predicting response to therapy, (**j**) outcome, (**k**) basic molecular subtyping (driving alterations in prostate cancer are shown), or (**l**) comprehensive interrogation of genitourinary cancer genomes and transcriptome (circus plot for visualizing the cancer genome, including point mutations/indels, copy number alterations, and gene fusion)

Acknowledgments S.A.T. is a co-inventor on a patent issued to the University of Michigan on ETS fusions in prostate cancer. The diagnostic field of use has been licensed to Gen-Probe, Inc., who has sublicensed certain rights to Ventana Medical Systems. S.A.T. is a co-inventor on a patent filed by the University of Michigan on SPINK1 in prostate cancer. The diagnostic field of use has been licensed to Gen-Probe, Inc., who has sublicensed certain rights to Ventana Medical Systems. S.A.T. is a consultant for and has received honoraria from Ventana Medical Systems, Almac Diagnostics, Astellas/Medivation, Janssen and Sanofi, and has had research sponsored by Astellas/Medivation and GenomeDX. S.A.T. is a co-founder of, consultant for, and the Laboratory Director of Strata Oncology.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017;67(1):7–30.
2. Gleason DF. Classification of prostatic carcinomas. Cancer Chemother Rep. 1966;50(3):125–8.
3. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. Am J Surg Pathol. 2016;40(2):244–52.
4. Epstein JI, Amin MB, Reuter VE, Humphrey PA. Contemporary Gleason grading of prostatic carcinoma: an update with discussion on practical issues to implement the 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. Am J Surg Pathol. 2017;41(4):e1–7.
5. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A contemporary prostate cancer grading system: a validated alternative to the Gleason score. Eur Urol. 2016;69(3):428–35.
6. Pettersson A, Graff RE, Bauer SR, et al. The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2012;21(9):1497–509.
7. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007;448(7153):595–9.
8. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science (New York, NY). 2005;310(5748):644–8.
9. Tomlins SA, Bjartell A, Chinnaiyan AM, et al. ETS gene fusions in prostate cancer: from discovery to daily clinical practice. Eur Urol. 2009;56(2):275–86.
10. Yoshimoto M, Joshua AM, Chilton-Macneill S, et al. Three-color FISH analysis of TMPRSS2/ERG fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. Neoplasia. 2006;8(6):465–9.
11. Braun M, Goltz D, Shaikhibrahim Z, et al. ERG protein expression and genomic rearrangement status in primary and metastatic prostate cancer–a comparative study of two monoclonal antibodies. Prostate Cancer Prostatic Dis. 2012;15(2):165–9.
12. Furusato B, Tan SH, Young D, et al. ERG oncoprotein expression in prostate cancer: clonal progression of ERG-positive tumor cells and potential for ERG-based stratification. Prostate Cancer Prostatic Dis. 2010;13(3):228–37.
13. Park K, Tomlins SA, Mudaliar KM, et al. Antibody-based detection of ERG rearrangement-positive prostate cancer. Neoplasia. 2010;12(7):590–8.
14. van Leenders GJ, Boormans JL, Vissers CJ, et al. Antibody EPR3864 is specific for ERG genomic fusions in prostate cancer: implications for pathological practice. Mod Pathol. 2011;24(8):1128–38.
15. Minner S, Enodien M, Sirma H, et al. ERG status is unrelated to PSA recurrence in radically operated prostate cancer in the absence of antihormonal therapy. Clin Cancer Res. 2011;17(18):5878–88.
16. Young A, Palanisamy N, Siddiqui J, et al. Correlation of urine TMPRSS2:ERG and PCA3 to ERG+ and total prostate cancer burden. Am J Clin Pathol. 2012;138(5):685–96.
17. Shah RB, Tadros Y, Brummell B, Zhou M. The diagnostic use of ERG in resolving an "atypical glands suspicious for cancer" diagnosis in prostate biopsies beyond that provided by basal cell and alpha-methylacyl-CoA-racemase markers. Hum Pathol. 2013;44(5):786–94.
18. Epstein JI, Egevad L, Humphrey PA, Montironi R. Best practices recommendations in the application of immunohistochemistry in the prostate: report from the international society of urologic pathology consensus conference. Am J Surg Pathol. 2014;38(8):e6–e19.
19. Park K, Dalton JT, Narayanan R, et al. TMPRSS2:ERG gene fusion predicts subsequent detection of prostate cancer in patients with high-grade prostatic intraepithelial neoplasia. J Clin Oncol. 2014;32(3):206–11.
20. Grupp K, Diebel F, Sirma H, et al. SPINK1 expression is tightly linked to 6q15- and 5q21-deleted ERG-fusion negative prostate cancers but unrelated to PSA recurrence. Prostate. 2013;73(15):1690–8.
21. Tomlins SA, Rhodes DR, Yu J, et al. The role of SPINK1 in ETS rearrangement-negative prostate cancers. Cancer Cell. 2008;13(6):519–28.
22. Smith SC, Tomlins SA. Prostate cancer SubtyPINg biomarKers and outcome: is clarity emERGing? Clin Cancer Res. 2014;20(18):4733–6.
23. Bhalla R, Kunju LP, Tomlins SA, et al. Novel dual-color immunohistochemical methods for detecting ERG-PTEN and ERG-SPINK1 status in prostate carcinoma. Mod Pathol. 2013;26(6):835–48.
24. Fontugne J, Davis K, Palanisamy N, et al. Clonal evaluation of prostate cancer foci in biopsies with discontinuous tumor involvement by dual ERG/SPINK1 immunohistochemistry. Mod Pathol. 2016;29(2):157–65.
25. Cancer Genome Atlas Research N. The molecular taxonomy of primary prostate cancer. Cell. 2015;163(4):1011–25.
26. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012;44(6):685–9.
27. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012;487(7406):239–43.
28. Zhang P, Wang D, Zhao Y, et al. Intrinsic BET inhibitor resistance in SPOP-mutated prostate cancer is mediated by BET protein stabilization and AKT-mTORC1 activation. Nat Med. 2017;23(9):1055–62.
29. Dai X, Gan W, Li X, et al. Prostate cancer-associated SPOP mutations confer resistance to BET inhibitors through stabilization of BRD4. Nat Med. 2017;23(9):1063–71.
30. Janouskova H, El Tekle G, Bellini E, et al. Opposing effects of cancer-type-specific SPOP mutants on BET protein degradation and sensitivity to BET inhibitors. Nat Med. 2017;23(9):1046–54.
31. Palanisamy N, Ateeq B, Kalyana-Sundaram S, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med. 2010;16(7):793–8.
32. Wang XS, Shankar S, Dhanasekaran SM, et al. Characterization of KRAS rearrangements in metastatic prostate cancer. Cancer Discov. 2011;1(1):35–43.
33. Wu YM, Su F, Kalyana-Sundaram S, et al. Identification of targetable FGFR gene fusions in diverse cancers. Cancer Discov. 2013;3(6):636–47.

34. Beltran H, Yelensky R, Frampton GM, et al. Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. Eur Urol. 2013;63(5):920–6.
35. Shen Y, Lu Y, Yin X, Zhu G, Zhu J. KRAS and BRAF mutations in prostate carcinomas of Chinese patients. Cancer Genet Cytogenet. 2010;198(1):35–9.
36. Ren G, Liu X, Mao X, et al. Identification of frequent BRAF copy number gain and alterations of RAF genes in Chinese prostate cancer. Genes Chromosomes Cancer. 2012;51(11):1014–23.
37. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161(5):1215–28.
38. Qian J, Jenkins RB, Bostwick DG. Detection of chromosomal anomalies and c-myc gene amplification in the cribriform pattern of prostatic intraepithelial neoplasia and carcinoma by fluorescence in situ hybridization. Mod Pathol. 1997;10(11):1113–9.
39. Lotan TL, Carvalho FL, Peskoe SB, et al. PTEN loss is associated with upgrading of prostate cancer from biopsy to radical prostatectomy. Mod Pathol. 2015;28(1):128–37.
40. Lotan TL, Wei W, Ludkovski O, et al. Analytic validation of a clinical-grade PTEN immunohistochemistry assay in prostate cancer by comparison with PTEN FISH. Mod Pathol. 2016;29(8):904–14.
41. Lotan TL, Wei W, Morais CL, et al. PTEN loss as determined by clinical-grade immunohistochemistry assay is associated with worse recurrence-free survival in prostate cancer. Eur Urol Focus. 2016;2(2):180–8.
42. Krohn A, Diedler T, Burkhardt L, et al. Genomic deletion of PTEN is associated with tumor progression and early PSA recurrence in ERG fusion-positive and fusion-negative prostate cancer. Am J Pathol. 2012;181(2):401–12.
43. Chaux A, Peskoe SB, Gonzalez-Roibon N, et al. Loss of PTEN expression is associated with increased risk of recurrence after prostatectomy for clinically localized prostate cancer. Mod Pathol. 2012;25(11):1543–9.
44. Tosoian JJ, Almutairi F, Morais CL, et al. Prevalence and prognostic significance of PTEN loss in African-American and European-American men undergoing radical prostatectomy. Eur Urol. 2017;71(5):697–700.
45. Guedes LB, Tosoian JJ, Hicks J, Ross AE, Lotan TL. PTEN loss in Gleason score 3 + 4 = 7 prostate biopsies is associated with nonorgan confined disease at radical prostatectomy. J Urol. 2017;197(4):1054–9.
46. Suzuki H, Freije D, Nusskern DR, et al. Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues. Cancer Res. 1998;58(2):204–9.
47. Attard G, Swennenhuis JF, Olmos D, et al. Characterization of ERG, AR and PTEN gene status in circulating tumor cells from patients with castration-resistant prostate cancer. Cancer Res. 2009;69(7):2912–8.
48. Sowalsky AG, Ye H, Bubley GJ, Balk SP. Clonal progression of prostate cancers from Gleason grade 3 to grade 4. Cancer Res. 2013;73(3):1050–5.
49. Kattan MW. Judging new markers by their ability to improve predictive accuracy. J Natl Cancer Inst. 2003;95(9):634–5.
50. Lennartz M, Minner S, Brasch S, et al. The combination of DNA ploidy status and PTEN/6q15 deletions provides strong and independent prognostic information in prostate cancer. Clin Cancer Res. 2016;22(11):2802–11.
51. Gudmundsson J, Besenbacher S, Sulem P, et al. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med. 2010;2(62):62ra92.
52. Benafif S, Eeles R. Genetic predisposition to prostate cancer. Br Med Bull. 2016;120(1):75–89.
53. Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med. 2012;366(2):141–9.
54. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. Nat Genet. 2012;44(12):1326–9.
55. Pritchard CC, Mateo J, Walsh MF, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N Engl J Med. 2016;375(5):443–53.
56. Penney KL, Sinnott JA, Fall K, et al. mRNA expression signature of Gleason grade predicts lethal prostate cancer. J Clin Oncol. 2011;29(17):2391–6.
57. Cuzick J, Berney DM, Fisher G, et al. Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. Br J Cancer. 2012;106(6):1095–9.
58. Cuzick J, Swanson GP, Fisher G, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol. 2011;12(3):245–55.
59. Klein EA, Cooperberg MR, Magi-Galluzzi C, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. Eur Urol. 2014;66(3):550–60.
60. Karnes RJ, Bergstralh EJ, Davicioni E, et al. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. J Urol. 2013;190(6):2047–53.
61. Klein EA, Haddad Z, Yousefi K, et al. Decipher genomic classifier measured on prostate biopsy predicts metastasis risk. Urology. 2016;90:148–52.
62. Knudsen BS, Kim HL, Erho N, et al. Application of a clinical whole-transcriptome assay for staging and prognosis of prostate cancer diagnosed in needle core biopsy specimens. J Mol Diagn. 2016;18(3):395–406.
63. Saad F, Latour M, Lattouf JB, et al. Biopsy based proteomic assay predicts risk of biochemical recurrence after radical prostatectomy. J Urol. 2017;197(4):1034–40.
64. Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541(7637):359–64.
65. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013;153(3):666–77.
66. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. Nature. 2011;470(7333):214–20.
67. Kumar A, Coleman I, Morrissey C, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med. 2016;22(4):369–78.
68. Hovelson DH, McDaniel AS, Cani AK, et al. Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. Neoplasia. 2015;17(4):385–99.
69. Kron KJ, Murison A, Zhou S, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat Genet. 2017;49(9):1336–45.
70. Chua MLK, Lo W, Pintilie M, et al. A prostate cancer "nimbosus": genomic instability and SChLAP1 dysregulation underpin aggression of intraductal and cribriform subpathologies. Eur Urol. 2017;72(5):665–74.
71. Mehra R, Shi Y, Udager AM, et al. A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. Neoplasia. 2014;16(12):1121–7.
72. Mehra R, Udager AM, Ahearn TU, et al. Overexpression of the long non-coding RNA SChLAP1 independently predicts lethal prostate cancer. Eur Urol. 2016;70(4):549–52.
73. Bottcher R, Hoogland AM, Dits N, et al. Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. Oncotarget. 2015;6(6):4036–50.

74. Prensner JR, Iyer MK, Sahu A, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet. 2013;45(11):1392–8.
75. Prensner JR, Zhao S, Erho N, et al. RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. Lancet Oncol. 2014;15(13):1469–80.
76. Prensner JR, Iyer MK, Balbin OA, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011;29(8):742–9.
77. Nelson PS. Targeting the androgen receptor in prostate cancer–a resilient foe. N Engl J Med. 2014;371(11):1067–9.
78. Abida W, Armenia J, Gopalan A, et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. JCO Precision Oncology. 2017;1:1–16.
79. Robinson DR, Wu YM, Lonigro RJ, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548(7667):297–303.
80. Mateo J, Carreira S, Sandhu S, et al. DNA-repair defects and olaparib in metastatic prostate cancer. N Engl J Med. 2015;373(18):1697–708.
81. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31(11):1023–31.
82. Ito K, Miyakubo M, Sekine Y, et al. Diagnostic significance of [−2]pro-PSA and prostate dimension-adjusted PSA-related indices in men with total PSA in the 2.0–10.0 ng/mL range. World J Urol. 2013;31(2):305–11.
83. Chan TY, Mikolajczyk SD, Lecksell K, et al. Immunohistochemical staining of prostate cancer with monoclonal antibodies to the precursor of prostate-specific antigen. Urology. 2003;62(1):177–81.
84. Vickers A, Cronin A, Roobol M, et al. Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication. J Clin Oncol. 2010;28(15):2493–8.
85. Salagierski M, Schalken JA. Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2:ERG gene fusion. J Urol. 2012;187(3):795–801.
86. Hessels D, Klein Gunnewiek JM, van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003;44(1):8–15; discussion 15–16.
87. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008;180(4):1303–8; discussion 1308.
88. Sanda MG, Feng Z, Howard DH, et al. Association between combined TMPRSS2:ERG and PCA3 RNA urinary testing and detection of aggressive prostate cancer. JAMA Oncol. 2017;3(8):1085–93.
89. Tomlins SA, Day JR, Lonigro RJ, et al. Urine TMPRSS2:ERG plus PCA3 for individualized prostate cancer risk assessment. Eur Urol. 2016;70(1):45–53.
90. Wei JT, Feng Z, Partin AW, et al. Can urinary PCA3 supplement PSA in the early detection of prostate cancer? J Clin Oncol. 2014;32(36):4066–72.
91. Cornu JN, Cancel-Tassin G, Egrot C, Gaffory C, Haab F, Cussenot O. Urine TMPRSS2:ERG fusion transcript integrated with PCA3 score, genotyping, and biological features are correlated to the results of prostatic biopsies in men at risk of prostate cancer. Prostate. 2012;73(3):242–9.
92. Leyten GH, Hessels D, Jannink SA, et al. Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. Eur Urol. 2014;65(3):534–42.
93. Stephan C, Jung K, Semjonow A, et al. Comparative assessment of urinary prostate cancer antigen 3 and TMPRSS2:ERG gene fusion with the serum [−2]proprostate-specific antigen-based prostate health index for detection of prostate cancer. Clin Chem. 2013;59(1):280–8.
94. Van Neste L, Hendriks RJ, Dijkstra S, et al. Detection of high-grade prostate cancer using a urinary molecular biomarker-based risk score. Eur Urol. 2016;70(5):740–8.
95. Leyten GH, Hessels D, Smit FP, et al. Identification of a candidate gene panel for the early diagnosis of prostate cancer. Clin Cancer Res. 2015;21(13):3061–70.
96. McKiernan J, Donovan MJ, O'Neill V, et al. A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. JAMA Oncol. 2016;2(7):882–9.
97. Scher HI, Jia X, de Bono JS, et al. Circulating tumour cells as prognostic markers in progressive, castration-resistant prostate cancer: a reanalysis of IMMC38 trial data. Lancet Oncol. 2009;10(3):233–9.
98. Danila DC. Circulating tumors cells as biomarkers: progress toward biomarker qualification. Cancer J (Sudbury, Mass). 2011;17(6):438–50.
99. Antonarakis ES, Lu C, Wang H, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. N Engl J Med. 2014;371(11):1028–38.
100. Heller G, Fizazi K, McCormack R, et al. The added value of circulating tumor cell enumeration to standard markers in assessing prognosis in a metastatic castration-resistant prostate cancer population. Clin Cancer Res. 2017;23(8):1967–73.
101. Scher HI, Lu D, Schreiber NA, et al. Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. JAMA Oncol. 2016;2(11):1441–9.
102. Wyatt AW, Azad AA, Volik SV, et al. Genomic alterations in cell-free DNA and enzalutamide resistance in castration-resistant prostate cancer. JAMA Oncol. 2016;2(12):1598–606.
103. Salvi S, Casadio V, Conteduca V, et al. Circulating AR copy number and outcome to enzalutamide in docetaxel-treated metastatic castration-resistant prostate cancer. Oncotarget. 2016;7(25):37839–45.
104. Lallous N, Volik SV, Awrey S, et al. Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. Genome Biol. 2016;17:10.
105. Carreira S, Romanel A, Goodall J, et al. Tumor clone dynamics in lethal prostate cancer. Sci Transl Med. 2014;6(254):254ra125.
106. Romanel A, Gasi Tandefelt D, Conteduca V, et al. Plasma AR and abiraterone-resistant prostate cancer. Sci Transl Med. 2015;7(312):312re310.
107. Wyatt AW, Annala M, Aggarwal R, et al. Concordance of circulating tumor DNA and matched metastatic tissue biopsy in prostate cancer. J Natl Cancer Inst (JNCI). 2017;109(12):djx118-djx118.
108. Quigley D, Alumkal JJ, Wyatt AW, et al. Analysis of circulating cell-free DNA identifies multiclonal heterogeneity of BRCA2 reversion mutations associated with resistance to PARP inhibitors. Cancer Discov. 2017;7:999.
109. Goodall J, Mateo J, Yuan W, et al. Circulating free DNA to guide prostate cancer treatment with PARP inhibition. Cancer Discov. 2017;7(9):1006–17.
110. Castillo-Martin M, Domingo-Domenech J, Karni-Schmidt O, Matos T, Cordon-Cardo C. Molecular pathways of urothelial development and bladder tumorigenesis. Urol Oncol. 2010;28(4):401–8.
111. Netto GJ. Molecular biomarkers in urothelial carcinoma of the bladder: are we there yet? Nat Rev Urol. 2012;9(1):41–51.

112. Burger M, van der Aa MN, van Oers JM, et al. Prediction of progression of non-muscle-invasive bladder cancer by WHO 1973 and 2004 grading and by FGFR3 mutation status: a prospective study. Eur Urol. 2008;54(4):835–43.
113. Williams SV, Hurst CD, Knowles MA. Oncogenic FGFR3 gene fusions in bladder cancer. Hum Mol Genet. 2013;22(4):795–803.
114. Guo G, Sun X, Chen C, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat Genet. 2013;45(12):1459–63.
115. Ross JS, Wang K, Khaira D, et al. Comprehensive genomic profiling of 295 cases of clinically advanced urothelial carcinoma of the urinary bladder reveals a high frequency of clinically relevant genomic alterations. Cancer. 2016;122(5):702–11.
116. Goebell PJ, Groshen SG, Schmitz-Drager BJ. International study-initiative on bladder C. p53 immunohistochemistry in bladder cancer–a new approach to an old question. Urol Oncol. 2010;28(4):377–88.
117. van Rhijn BW, Zuiverloon TC, Vis AN, et al. Molecular grade (FGFR3/MIB-1) and EORTC risk scores are predictive in primary non-muscle-invasive bladder cancer. Eur Urol. 2010;58(3):433–41.
118. George B, Datar RH, Wu L, et al. p53 gene and protein status: the role of p53 alterations in predicting outcome in patients with bladder cancer. J Clin Oncol. 2007;25(34):5352–8.
119. Margulis V, Lotan Y, Karakiewicz PI, et al. Multi-institutional validation of the predictive value of Ki-67 labeling index in patients with urinary bladder cancer. J Natl Cancer Inst. 2009;101(2):114–9.
120. Lindgren D, Liedberg F, Andersson A, et al. Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. Oncogene. 2006;25(18):2685–96.
121. Lindgren D, Frigyesi A, Gudjonsson S, et al. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. Cancer Res. 2010;70(9):3463–72.
122. Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507(7492):315–22.
123. Robertson AG, Kim J, Al-Ahmadie H, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell. 2017;171(3):540–556.e25.
124. Hedegaard J, Lamy P, Nordentoft I, et al. Comprehensive transcriptional analysis of early-stage urothelial carcinoma. Cancer Cell. 2016;30(1):27–42.
125. Choi W, Porten S, Kim S, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell. 2014;25(2):152–65.
126. McConkey DJ, Choi W, Shen Y, et al. A prognostic gene expression signature in the molecular classification of chemotherapy-naive urothelial Cancer is predictive of clinical outcomes from neoadjuvant chemotherapy: a phase 2 trial of dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin with bevacizumab in urothelial Cancer. Eur Urol. 2016;69(5):855–62.
127. Mitra AP, Lam LL, Ghadessi M, et al. Discovery and validation of novel expression signature for postcystectomy recurrence in high-risk bladder cancer. J Natl Cancer Inst. 2014;106(11).
128. Damrauer JS, Hoadley KA, Chism DD, et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. Proc Natl Acad Sci U S A. 2014;111(8):3110–5.
129. Seiler R, Ashab HA, Erho N, et al. Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. Eur Urol. 2017;72(4):544–54.
130. Lee JK, Havaleshko DM, Cho H, et al. A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. Proc Natl Acad Sci U S A. 2007;104(32):13086–91.
131. Smith SC, Baras AS, Lee JK, Theodorescu D. The COXEN principle: translating signatures of in vitro chemosensitivity into tools for clinical outcome prediction and drug discovery in cancer. Cancer Res. 2010;70(5):1753–8.
132. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet. 2011;43(9):875–8.
133. Iyer G, Hanrahan AJ, Milowsky MI, et al. Genome sequencing identifies a basis for everolimus sensitivity. Science (New York, NY). 2012;338(6104):221.
134. Rebouissou S, Bernard-Pierrot I, de Reynies A, et al. EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. Sci Transl Med. 2014;6(244):244ra291.
135. Hussain M, Daignault S, Agarwal N, et al. A randomized phase 2 trial of gemcitabine/cisplatin with or without cetuximab in patients with advanced urothelial carcinoma. Cancer. 2014;120(17):2684–93.
136. Hussain MH, MacVicar GR, Petrylak DP, et al. Trastuzumab, paclitaxel, carboplatin, and gemcitabine in advanced human epidermal growth factor receptor-2/neu-positive urothelial carcinoma: results of a multicenter phase II National Cancer Institute trial. J Clin Oncol. 2007;25(16):2218–24.
137. Powles T, Huddart RA, Elliott T, et al. Phase III, double-blind, randomized trial that compared maintenance Lapatinib versus placebo after first-line chemotherapy in patients with human epidermal growth factor receptor 1/2-positive metastatic bladder cancer. J Clin Oncol. 2017;35(1):48–55.
138. Sokolova IA, Halling KC, Jenkins RB, et al. The development of a multitarget, multicolor fluorescence in situ hybridization assay for the detection of urothelial carcinoma in urine. J Mol Diagn. 2000;2(3):116–23.
139. Halling KC, King W, Sokolova IA, et al. A comparison of cytology and fluorescence in situ hybridization for the detection of urothelial carcinoma. J Urol. 2000;164(5):1768–75.
140. Li HX, Wang MR, Zhao H, Cao J, Li CL, Pan QJ. Comparison of fluorescence in situ hybridization, NMP22 bladderchek, and urinary liquid-based cytology in the detection of bladder urothelial carcinoma. Diagn Cytopathol. 2013;41(10):852–7.
141. Bubendorf L, Grilli B, Sauter G, Mihatsch MJ, Gasser TC, Dalquen P. Multiprobe FISH for enhanced detection of bladder cancer in voided urine specimens and bladder washings. Am J Clin Pathol. 2001;116(1):79–86.
142. Sarosdy MF, Kahn PR, Ziffer MD, et al. Use of a multitarget fluorescence in situ hybridization assay to diagnose bladder cancer in patients with hematuria. J Urol. 2006;176(1):44–7.
143. Ward DG, Bryan RT. Liquid biopsies for bladder cancer. Transl Androl Urol. 2017;6(2):331–5.
144. Scott SN, Ostrovnaya I, Lin CM, et al. Next-generation sequencing of urine specimens: a novel platform for genomic analysis in patients with non-muscle-invasive urothelial carcinoma treated with bacille Calmette-Guerin. Cancer. 2017;125(6):416–26.
145. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016;387(10031):1909–20.
146. Gupta S, Gill D, Poole A, Agarwal N. Systemic immunotherapy for urothelial cancer: current trends and future directions. Cancers (Basel). 2017;9(2):15.

147. Cancer Genome Atlas Research N, Linehan WM, Spellman PT, et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. N Engl J Med. 2016;374(2):135–45.
148. Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell. 2014;26(3):319–30.
149. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43–9.
150. Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res. 2008;14(15):4726–34.
151. Zbar B, Brauch H, Talmadge C, Linehan M. Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. Nature. 1987;327(6124):721–4.
152. Stillebroer AB, Mulders PF, Boerman OC, Oyen WJ, Oosterwijk E. Carbonic anhydrase IX in renal cell carcinoma: implications for prognosis, diagnosis, and therapy. Eur Urol. 2010;58(1):75–83.
153. Barocas DA, Mathew S, DelPizzo JJ, et al. Renal cell carcinoma sub-typing by histopathology and fluorescence in situ hybridization on a needle-biopsy specimen. BJU Int. 2007;99(2):290–5.
154. Hakimi AA, Tickoo SK, Jacobsen A, et al. TCEB1-mutated renal cell carcinoma: a distinct genomic and morphological subtype. Mod Pathol. 2015;28(6):845–53.
155. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469(7331):539–42.
156. Nguyen KA, Syed JS, Espenschied CR, et al. Advances in the diagnosis of hereditary kidney cancer: initial results of a multigene panel test. Cancer. 2017;123:4363.
157. Schmidt LS, Linehan WM. Genetic predisposition to kidney cancer. Semin Oncol. 2016;43(5):566–74.
158. Merino MJ, Torres-Cabala C, Pinto P, Linehan WM. The morphologic spectrum of kidney tumors in hereditary leiomyomatosis and renal cell carcinoma (HLRCC) syndrome. Am J Surg Pathol. 2007;31(10):1578–85.
159. Schmidt L, Duh FM, Chen F, et al. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. Nat Genet. 1997;16(1):68–73.
160. Shuch B, Hahn AW, Agarwal N. Current treatment landscape of advanced papillary renal cancer. J Clin Oncol. 2017;35(26):2981–3. JCO2017743328.
161. Argani P, Hawkins A, Griffin CA, et al. A distinctive pediatric renal neoplasm characterized by epithelioid morphology, basement membrane production, focal HMB45 immunoreactivity, and t(6;11)(p21.1;q12) chromosome translocation. Am J Pathol. 2001;158(6):2089–96.
162. Argani P, Antonescu CR, Illei PB, et al. Primary renal neoplasms with the ASPL-TFE3 gene fusion of alveolar soft part sarcoma: a distinctive tumor entity previously included among renal cell carcinomas of children and adolescents. Am J Pathol. 2001;159(1):179–92.
163. Green WM, Yonescu R, Morsberger L, et al. Utilization of a TFE3 break-apart FISH assay in a renal tumor consultation service. Am J Surg Pathol. 2013;37(8):1150–63.
164. Rao Q, Williamson SR, Zhang S, et al. TFE3 break-apart FISH has a higher sensitivity for Xp11.2 translocation-associated renal cell carcinoma compared with TFE3 or cathepsin K immunohistochemical staining alone: expanding the morphologic spectrum. Am J Surg Pathol. 2013;37(6):804–15.
165. Mosquera JM, Dal Cin P, Mertz KD, et al. Validation of a TFE3 break-apart FISH assay for Xp11.2 translocation renal cell carcinomas. Diagn Mol Pathol. 2011;20(3):129–37.
166. Skala SL, Xiao H, Udager AM, et al. Detection of 6 TFEB-amplified renal cell carcinomas and 25 renal cell carcinomas with MITF translocations: systematic morphologic analysis of 85 cases evaluated by clinical TFE3 and TFEB FISH assays. Mod Pathol. 2017;31(1):179–97.
167. Zhong M, De Angelo P, Osborne L, et al. Dual-color, break-apart FISH assay on paraffin-embedded tissues as an adjunct to diagnosis of Xp11 translocation renal cell carcinoma and alveolar soft part sarcoma. Am J Surg Pathol. 2010;34(6):757–66.
168. Williamson SR, Grignon DJ, Cheng L, et al. Renal cell carcinoma with chromosome 6p amplification including the TFEB gene: a novel mechanism of tumor pathogenesis? Am J Surg Pathol. 2017;41(3):287–98.
169. Argani P, Reuter VE, Zhang L, et al. TFEB-amplified renal cell carcinomas: an aggressive molecular subset demonstrating variable melanocytic marker expression and morphologic heterogeneity. Am J Surg Pathol. 2016;40(11):1484–95.
170. Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet. 2013;45(8):860–7.
171. Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463(7279):360–3.
172. van Haaften G, Dalgliesh GL, Davies H, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet. 2009;41(5):521–3.
173. Chen YB, Xu J, Skanderup AJ, et al. Molecular analysis of aggressive renal cell carcinoma with unclassified histology reveals distinct subsets. Nat Commun. 2016;7:13131.
174. Guo G, Gui Y, Gao S, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nat Genet. 2011;44(1):17–9.
175. Posadas EM, Limvorasak S, Figlin RA. Targeted therapies for renal cell carcinoma. Nat Rev Nephrol. 2017;13(8):496–511.
176. Winer AG, Motzer RJ, Hakimi AA. Prognostic biomarkers for response to vascular endothelial growth factor-targeted therapy for renal cell carcinoma. Urol Clin North Am. 2016;43(1):95–104.
177. Choueiri TK, Plantade A, Elson P, et al. Efficacy of sunitinib and sorafenib in metastatic papillary and chromophobe renal cell carcinoma. J Clin Oncol. 2008;26(1):127–31.
178. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. Lancet. 2008;372(9637):449–56.
179. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. Lancet Oncol. 2016;17(7):917–27.
180. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus Everolimus in advanced renal-cell carcinoma. N Engl J Med. 2015;373(19):1814–23.
181. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in advanced renal-cell carcinoma. N Engl J Med. 2015;373(19):1803–13.
182. Wiecek W, Karcher H. Nivolumab versus Cabozantinib: comparing overall survival in metastatic renal cell carcinoma. PLoS One. 2016;11(6):e0155389.
183. Choueiri TK, Halabi S, Sanford BL, et al. Cabozantinib versus Sunitinib as initial targeted therapy for patients with metastatic renal cell carcinoma of poor or intermediate risk: the alliance A031203 CABOSUN trial. J Clin Oncol. 2017;35(6):591–7.
184. Turajlic S, Litchfield K, Xu H, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol. 2017;18(8):1009–21.
185. Looijenga LH, Stoop H, de Leeuw HP, et al. POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. Cancer Res. 2003;63(9):2244–50.
186. Ulbright TM, Tickoo SK, Berney DM, Srigley JR. Members of the IIiDUPG. Best practices recommendations in the application of immunohistochemistry in testicular tumors: report from the

International Society of Urological Pathology consensus conference. Am J Surg Pathol. 2014;38(8):e50–9.
187. Bosl GJ, Dmitrovsky E, Reuter VE, et al. Isochromosome of the short arm of chromosome 12: clinically useful markers for male germ cell tumors. J Natl Cancer Inst. 1989;81(24):1874–8.
188. di Pietro A, Vries EG, Gietema JA, Spierings DC, de Jong S. Testicular germ cell tumours: the paradigm of chemo-sensitive solid tumours. Int J Biochem Cell Biol. 2005;37(12):2437–56.
189. Lutzker SG, Mathew R, Taller DR. A p53 dose-response relationship for sensitivity to DNA damage in isogenic teratocarcinoma cells. Oncogene. 2001;20(23):2982–6.
190. Taylor-Weiner A, Zack T, O'Donnell E, et al. Genomic evolution and chemoresistance in germ-cell tumours. Nature. 2016;540(7631):114–8.
191. Honecker F, Wermann H, Mayer F, et al. Microsatellite instability, mismatch repair deficiency, and BRAF mutation in treatment-resistant germ cell tumors. J Clin Oncol. 2009;27(13):2129–36.
192. Bethune G, Campbell J, Rocker A, Bell D, Rendon R, Merrimen J. Clinical and pathologic factors of prognostic significance in penile squamous cell carcinoma in a North American population. Urology. 2012;79(5):1092–7.
193. McDaniel AS, Hovelson DH, Cani AK, et al. Genomic profiling of penile squamous cell carcinoma reveals new opportunities for targeted therapy. Cancer Res. 2015;75(24):5219–27.
194. Chaux A, Reuter V, Lezcano C, Velazquez EF, Torres J, Cubilla AL. Comparison of morphologic features and outcome of resected recurrent and nonrecurrent squamous cell carcinoma of the penis: a study of 81 cases. Am J Surg Pathol. 2009;33(9):1299–306.
195. Ali SM, Pal SK, Wang K, et al. Comprehensive genomic profiling of advanced penile carcinoma suggests a high frequency of clinically relevant genomic alterations. Oncologist. 2016;21(1):33–9.
196. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014;158(4):929–44.
197. Udager AM, Liu TY, Skala SL, et al. Frequent PD-L1 expression in primary and metastatic penile squamous cell carcinoma: potential opportunities for immunotherapeutic approaches. Ann Oncol. 2016;27(9):1706–12.

Genomic Applications in Gynecologic Malignancies

31

Sarah Chiang, Luciano G. Martelotto, and Britta Weigelt

Introduction

The uterine corpus represents the most common site for gynecologic malignancies in North America and Europe with an estimated 63,230 new cases and 11,350 deaths in 2018 in the USA [1]. While the death rates decreased for the majority of the most common cancer types, since 2000, both the incidence and death rates for cancers of the uterus increased [2]. In fact, despite advances in diagnosis and therapy, when compared to 1975, the 5-year relative survival rate for patients with uterine cancer has decreased [2].

Cancer of the uterine corpus is a heterogeneous disease comprising multiple entities with distinct risk factors, histopathological features, and clinical outcomes. Approximately 75–80% of patients with endometrial carcinoma are diagnosed at an early stage (stage I/II; disease confined to the uterus) and are managed by surgery with or without adjuvant radiotherapy resulting in a 5-year overall survival rate of 74–91% [3]. Of these early-stage carcinomas, approximately 15–20% recur, and despite advances in adjuvant chemotherapy and radiation strategies, the outcome of women diagnosed with advanced or recurrent disease remains poor with a median overall survival of 5–15 months [4–6]. Malignant uterine mesenchymal neoplasms, albeit rare, generally have an aggressive clinical behavior and may pose significant diagnostic and disease management challenges. Thus, there is a need to improve current treatment strategies and incorporate targeted therapies in standard regimens, to recognize those patients with a high risk of recurrence, and to select the optimal systemic therapy for a given patient with uterine cancer. Molecular genetic studies aiming to identify specific molecular markers for classification, risk stratification, and treatment decision-making continue to unveil the repertoire of alterations in uterine cancer. Molecular profiling efforts have focused mainly on the most common histological subtypes of endometrial carcinomas, the endometrioid and serous adenocarcinomas, but more recently also on rare histological subtypes, all of which we discuss in this chapter. In addition, we describe our understanding of common and rare uterine mesenchymal tumors at the molecular level.

S. Chiang · B. Weigelt (✉)
Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
e-mail: weigeltb@mskcc.org

L. G. Martelotto
Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

The University of Melbourne Centre for Cancer Research, University of Melbourne, Victorian Comprehensive Cancer Center, Melbourne, VIC, Australia

Endometrial Carcinoma

Classification and Biomarkers of Endometrial Carcinoma

For optimal disease management of gynecological cancers, the extent of disease is determined on the basis of the International Federation of Gynecology and Obstetrics (FIGO) system. To improve the prognostic information provided by FIGO staging, the guidelines for endometrial carcinoma underwent revision in 2009 [7] based on the knowledge gathered since the first surgical-pathological FIGO staging system described in 1988.

Traditionally, endometrial carcinomas are classified based on clinical, epidemiological, and "endocrine-metabolic" features into two types [8]: type I tumors, which are low-grade and associated with unopposed estrogen stimulation and endometrial hyperplasia, and type II tumors, which are typically high-grade and initially considered unrelated to hormonal factors or hyperplasia. A pooled analysis of endometrial cancer risk factors in 14,069 endometrial cancer cases and 35,312 controls, however, revealed that type I and type II endometrial cancers share many common etiologic factors, and it has been suggested that the etiology of type II tumors may not be completely estrogen-independent [9]. For histologic subtyping of endometrial cancer, modified

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_31

versions of the World Health Organization (WHO) [10] and International Society of Gynecological Pathologists classifications [11, 12] have been used (see Table 31.1). Noteworthy, although carcinosarcomas (also called malignant mixed Mullerian tumors) are designated as mixed epithelial and mesenchymal tumors according to the WHO, recent studies have provided evidence to suggest that these neoplasms derive from a transformed epithelial cell and are now considered by many in the pathology community as high-grade carcinomas undergoing sarcomatous differentiation [13] and, to some extent, the uterine counterpart of metaplastic carcinoma of the breast [14].

Table 31.1 Histological classification of endometrial adenocarcinoma

World Health Organization [10]	Blaustein [12]	Clement and Young [11]
Endometrioid Squamous differentiation Villoglandular Secretory	Endometrioid Squamous differentiation Villoglandular Secretory Ciliated	Endometrioid Typical Secretory With papillae Villoglandular Small non-villous papillae Microglandular Sertoliform Tumors with cords and hyalinization With metaplastic changes Squamous differentiation Clear cell change, not otherwise specified Surface changes resembling syncytial metaplasia or microglandular hyperplasia Ciliated cells Oxyphilic (or oncocytic) cells With spindled epithelial cells (sarcomatoid)
Mucinous	Mucinous	Mucinous
Serous	Serous	Serous
Clear cell	Clear cell Squamous Transitional	Clear cell Squamous Transitional
Neuroendocrine Low-grade neuroendocrine tumor Carcinoid High-grade neuroendocrine carcinoma Small cell Large cell		Neuroendocrine
Undifferentiated	Undifferentiated	Undifferentiated
Mixed	Mixed	Mixed (minor component accounts for at least 10% of the tumor)
Other poorly differentiated carcinomas	Glassy cell Yolk sac tumor Giant cell Choriocarcinoma	Lymphoepithelioma-like Hepatoid Giant cell With trophoblastic differentiation With yolk sac tumor
Secondary carcinomas		Secondary carcinomas

A three-tiered FIGO grading system based on architectural pattern and nuclear features is used to grade some types of endometrial carcinomas, namely, endometrioid and mucinous adenocarcinomas [15]. According to the percentage of solid nonglandular and non-squamous growth, tumors are assigned grade 1 (≤5% of solid growth), grade 2 (6–50% of solid growth), or grade 3 (>50% of solid growth). The overall grade is raised by one if marked nuclear atypia is present [16]. FIGO grading has generally not been recommended for non-endometrioid (i.e., serous and clear cell), mixed epithelial, or morphologically heterogeneous tumors, due to potential lack of correlation with clinical outcome. Alternative grading schemes, such as a binary system irrespective of tumor subtype, have been proposed and may be more reproducible in clinical practice [17, 18].

In addition to histological type and grade, several surgical and pathological parameters, including FIGO stage, depth of myometrial invasion, lymphovascular invasion, cervical involvement, lymph node status, and DNA ploidy, have been shown to be predictors of prognosis in patients with endometrial carcinoma [19] and are used to guide treatment. There is no consensus, however, as to the prognostic and predictive factors to be used, and therefore the definition of risk groups in endometrial cancer is variable [20].

While there are specific, established histologic criteria for various subtypes of endometrial carcinoma (Fig. 31.1), tumors that demonstrate overlapping morphologic features or are poorly differentiated continue to pose significant diagnostic challenges. Inter-observer reproducibility in the diagnosis of high-grade endometrial carcinoma is limited, even among expert gynecologic pathologists [21, 22]. The differential diagnoses of serous versus clear cell or serous versus FIGO grade 3 endometrioid carcinomas represent the most frequent areas of disagreement [21, 23]. In such cases, ancillary immunohistochemical studies in conjunction with morphologic interpretation have proven helpful in tumor classification [23–25] (Table 31.2). Endometrioid, serous, and clear cell carcinomas all express pan-cytokeratins (CKs), epithelial membrane antigen (EMA), and glycoprotein-associated markers CA125, Ber-EP4, and B72.3. All three subtypes are also usually CK7-positive and CK20-negative. Endometrioid adenocarcinomas that are typically estrogen receptor (ER)- and progesterone

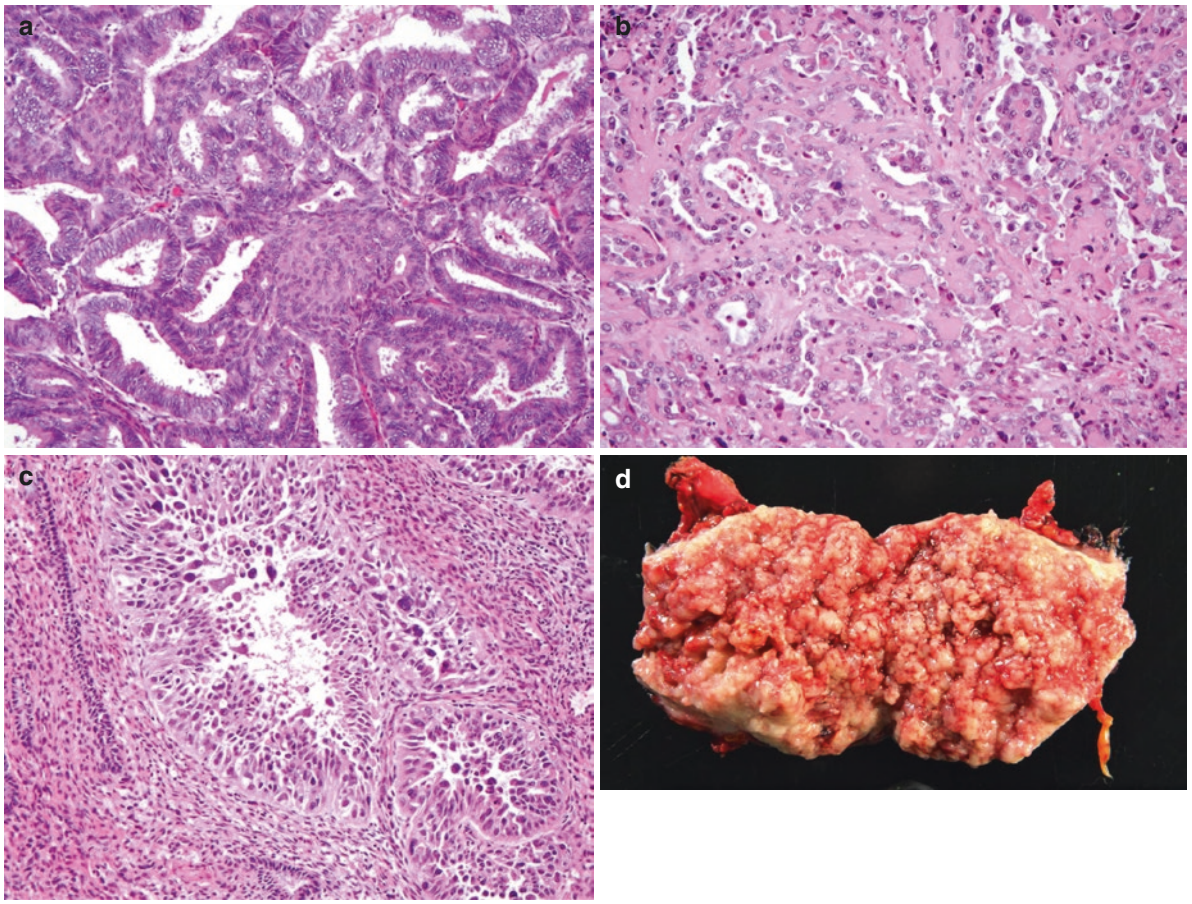


Fig. 31.1 Major subtypes of endometrial adenocarcinoma. (**a**) FIGO grade I endometrioid adenocarcinoma typically consists of well-differentiated endometrioid glands often with foci of squamous metaplasia. (**b**) Cystic glands are lined by a single layer of oxyphilic cells and embedded in hyalinized matrix in a clear cell carcinoma. (**c**) Marked nuclear atypia, tufting, and hobnail architecture are diagnostic of serous carcinoma. (**d**) Exophytic, friable tumor occupies the entire endometrial cavity in an endometrioid adenocarcinoma

Table 31.2 Immunophenotype of endometrioid, serous, and clear cell carcinomas of the uterus

Antibody	Endometrioid	Serous	Clear cell
ER	+	+	Focal + or −
PR	+	+	Focal + or −
MIB1 (Ki-67)	Low	High	Moderate
p53	Heterogeneous or aberrant (strong, diffuse, or completely absent)	Aberrant (strong, diffuse, or completely absent)	Heterogeneous or aberrant (strong, diffuse, or completely absent)
p16	Patchy, variable +	Strong, diffuse +	Moderate +
Keratins	CK7+, CK20−	CK7+, CK20−	CK7+, CK20−
EMA	+	+	+
CA125, Ber-EP4, B72.3	+	+	+
PTEN	−	+	+
HNF-1b	−	−	+
Napsin	−	−	+

CK cytokeratin, *EMA* epithelial membrane antigen, *ER* estrogen receptor, *PR* progesterone receptor, − negative, + positive

receptor (PR)-positive display heterogeneous p53 and p16 expression, a low Ki-67 proliferation index, and loss of PTEN immunoreactivity. It should be noted that a subset of FIGO grade 3 endometrioid carcinomas, however, shows aberrant p53 immunoexpression (either strong and diffuse or complete absence of staining). Serous carcinomas typically exhibit at least some ER and PR expression, high Ki-67 proliferation index, diffuse and strong p16 staining, and aberrant p53 immunoexpression. Clear cell carcinoma is often negative or only focally positive for ER and PR and may show either wild-type or aberrant p53 expression (Table 31.2). Napsin A and HNF-1b expression may be seen in, but is not limited to clear cell carcinoma [26, 27]. Undifferentiated or dedifferentiated carcinomas, the latter associated with a low-grade endometrioid component, show only weak or focal keratin expression, but most examples demonstrate intense staining of rare cells with EMA and CK18 [28]. Tumors may exhibit aberrant p53 staining and loss of PAX8 expression [29].

Molecular Genetic Classification of Endometrial Cancer

The limited reproducibility of the histologic subtyping of some high-grade endometrial carcinomas [21, 22, 30] coupled with burgeoning molecular data on the genetic alterations found in subsets of endometrial cancer have fueled efforts in redefining the histologic classification of these tumors and to define an ancillary molecular-based classification scheme [31]. Endometrioid endometrial cancers are characterized by a high mutational frequency, including mutations affecting *PTEN*, *PIK3CA*, *KRAS*, *FGFR2*, and *CTNNB1*, and by microsatellite instability (MSI) often due to *MLH1* promoter hypermethylation [32] in up to 45% cases (reviewed in [33]). More recently, recurrent mutations in *PIK3R1* and *ARID1A* have been reported [34–36]. In contrast, serous carcinomas harbor a high frequency of *TP53* and *PPP2R1A* mutations [36, 37], as well as overexpression and amplification of *HER2* in a subset of cases [38, 39]. Given that (i) frequencies of some mutations vary by histological grade [35, 40], (ii) mutational profiles of endometrioid and serous endometrial carcinomas partially overlap, and (iii) there is genetic heterogeneity within each of these entities, single gene mutations or small gene panels, although not sufficient to allow for a purely mutation-based classification, may serve as an aid for morphological classification. In particular, given the binary output of mutational analyses, it has been suggested that these molecular markers may be easier to interpret than immunohistochemical results [40]. As some molecular alterations are preferentially but not exclusively found in serous versus endometrioid cancers, such as *HER2* amplification or *TP53* mutations, it may not be entirely surprising that these markers are associated with prognosis when all types of endometrial carcinoma are considered together [41].

In breast cancer, gene expression profiling has led to a molecular classification of the disease [42] and to commercially available prognostic gene signatures as predictors of outcome and guides for treatment (reviewed in [43]). In endometrial cancer, the evidence from microarray-based expression profiling studies suggests that different histological subtypes harbor distinct transcriptomic [44–46] and distinct microRNA profiles [47] and that several genes are overexpressed and amplified in specific subtypes, such as *STK15* in serous and clear cell carcinomas [48]. Furthermore, it has been observed that at the transcriptomic level, stage I serous cancers are similar to stage I grade 3 endometrioid cancers [49] and that high-grade endometrial cancers can be classified into two subgroups with distinct molecular alterations using a panel of 22 genes involved in the PI3K-AKT pathway [50]. Gene expression sets associated with prognosis have also been reported, including a risk score stratifying clinically/pathologically intermediate-risk endometrial cancer patients into high- and low-risk recurrence groups with significant differences in time to recurrence [51]. Gene expression profiling has also yielded gene expression clusters that are predictive of recurrence in endometrioid and serous cancers [52]. A meta-analysis of 12 gene expression microarray studies followed by validation using RNA-sequencing data identified a 9-gene signature associated with outcome in patients with endometrioid endometrial cancers [53]. Analysis of endometrial cancers at the genomic level using array-based methods has revealed that, in contrast to serous carcinomas and carcinosarcomas, endometrioid tumors harbor only few gene copy number alterations [54]. These studies have also shown that within the group of endometrioid cancers, high levels of chromosomal instability are associated with poor prognosis [55]. Unlike breast cancer, however, due to the small number of cases analyzed so far and the lack of robust validation of gene sets or signatures in independent datasets, microarray-based studies are yet to yield clinically utilized assays for patients with endometrial cancer.

Advances in high-throughput sequencing technologies have allowed for the characterization of complete genomes at base-pair resolution in a time- and cost-effective manner. Using massively parallel sequencing (also called "next-generation sequencing (NGS)")-based technologies, whole-exome sequencing of serous endometrial cancers has revealed frequent occurrence of somatic mutations in chromatin-remodeling genes (e.g., *CHD4*) and ubiquitin ligase complex genes such as *FBXW7* [56–58] and amplification of *CCNE1* [57, 58], a target of FBXW7-mediated ubiquitination, in addition to previously recognized mutations in *TP53*, *PIK3CA*, and *PPP2R1A*. In 2013, The Cancer Genome Atlas (TCGA) [59] group performed a comprehensive integrative genomic analysis of endometrioid and serous endometrial cancers, which identified four groups (Table 31.3):

1. A subset of endometrioid endometrial carcinomas (7%) characterized by very high mutation rates ("ultramutated") and hotspot mutations in the exonuclease domain of *POLE* [60], the catalytic subunit of the DNA polymerase epsilon, and a favorable outcome (*POLE* (ultramutated) subtype).
2. Cancers characterized by microsatellite instability (MSI) and associated hyper-mutation with a background mutation rate approximately tenfold greater than non-MSI tumors. These tumors are exclusively of endometrioid type, have few DNA copy number alterations, and harbor recurrent frameshift deletions in *RPL22* as well as *KRAS* mutations (MSI (hypermutated) subtype).
3. Microsatellite-stable endometrioid cancers with lower mutation frequency than MSI tumors, low copy number changes, and high frequency of *CTNNB1* mutations (52%) (copy number low (endometrioid) subtype).
4. A group comprising all serous endometrial and a subset (25%) of FIGO grade 3 endometrioid carcinomas, char-

Table 31.3 Genomic subtypes of endometrioid and serous endometrial carcinoma

	POLE (ultramutated)	MSI (hypermutated)	Copy number low (endometrioid)	Copy number high (serous-like)
Mutation rate	Very high	High	Intermediate	2.3×10^{-6}
Copy number alterations	Few	Few	Few	High
Microsatellite instability	Mixed (high, low, stable)	High; *MLH1* promoter methylation	Stable	Stable
Characteristic genes mutated	*POLE*	*PTEN, RPL22, KRAS, ARID5B*	*PTEN, CTNNB1*	*TP53, FBXW7, PPP2R1A*
Histology	Endometrioid	Endometrioid	Endometrioid	Serous
				Endometrioid
Tumor FIGO grade	Mixed (1, 2, and 3)	Mixed (1, 2, and 3)	1 and 2	3
Progression-free survival	Best	Intermediate	Intermediate	Worst

Based on whole-exome sequencing, gene copy number, and microsatellite instability (MSI) analysis of 248 endometrioid and serous endometrial cancers by The Cancer Genome Atlas Network [59]

acterized by extensive DNA copy number alterations and genomic instability, similar to those seen in high-grade serous ovarian carcinoma (see separate chapter). These tumors harbor frequent *TP53* (90%), *FBXW7* (22%), and *PP2R1A* (22%) mutations (copy number high (serous-like) subtype).

This integrative genomic analysis allowed for the identification of mutations preferentially associated with specific subgroups of endometrial cancers, such as *ARID5B* mutations in 23.1% of MSI endometrioid cancers versus 5.6% and 0% in microsatellite-stable endometrioid and serous carcinomas, respectively [59]. However, an overlap in the mutational repertoire between these genomic groups was observed as was the genetic heterogeneity within a given group (i.e., not all cases in a genomically defined group harbored a specific mutation). Furthermore, a subset of high-grade endometrioid tumors was shown to harbor copy number and mutational profiles similar to those of serous carcinomas, providing evidence to suggest that these may share a similar biology and be driven by similar genetic alterations, providing molecular evidence to warrant studies investigating whether these tumors can be treated similarly.

Importantly, the molecular classification of endometrioid and serous endometrial carcinomas is not a mere academic exercise, but has prognostic and predictive implications. In the TCGA study, the association of the molecular subtypes with outcome was reported, with endometrial cancers of *POLE* (ultramutated) subtype having the best and copy number (serous-like) subtype having the worst outcomes (Table 31.3) [59]. The favorable outcome of *POLE* (ultramutated) tumors, in particular those of high grade, has been subsequently confirmed in independent studies [61–65]. For the translation of the molecular classification into the clinic, a pragmatic surrogate classifier, which identifies four molecular subtypes that are analogous, but not identical to those described by TCGA, has been developed [66, 67]. Using a combination of immunohistochemistry for DNA mismatch repair (MMR) proteins and p53 as well as mutation analysis of the *POLE* exonuclease domain (exons 9–14), molecular subtypes can be identified, which are significantly associated with outcome [66–68]: endometrioid and serous carcinomas (i) showing loss of DNA MMR protein(s) are classified as MMR deficient (surrogate for MSI (hypermutated)); (ii) harboring *POLE* exonuclease domain mutations are classified as of *POLE* subtype; (iii) demonstrating aberrant p53 immunohistochemical staining are classified as p53 abnormal, corresponding to the copy number high (serous-like) subtype; and (iv) showing normal DNA MMR protein expression, lacking *POLE* exonuclease domain mutations, and displaying p53 wild-type immunohistochemical expression are classified as p53 wild-type, corresponding to the copy number low (endometrioid) subtype [66, 67].

This association with outcome is not only observed when applied to unselected endometrioid and serous endometrial cancers, but the molecular classification also identifies distinct prognostic subgroups when applied to FIGO grade 3 endometrioid endometrial cancers only and in a multivariate analysis was found to be an independent prognostic factor for recurrence-free survival in these tumors [69]. The refinement of prognosis in FIGO grade 3 endometrioid tumors is important, as this group is heterogeneous in terms of recurrence rates (14–20%) [70, 71], molecular profile (approximately 25% are of copy number high (serous-like) subtype in the TCGA study [59]), and need for therapeutic intervention.

The presence of *POLE* exonuclease domain mutations and the association with favorable prognosis is not restricted to endometrial cancers of endometrioid histology, but has also been reported for undifferentiated/dedifferentiated [72] and for clear cell carcinomas of the uterus [73]. Using a surrogate, all four molecular subtypes described by TCGA were identified in a series of 17 undifferentiated/dedifferentiated carci-

nomas, with the majority being either DNA MMR deficient (MSI; 45%) or *POLE* ultramutated (11%) [74], highlighting the molecular heterogeneity of this rare aggressive subtype of uterine cancer. Whether the subtype classification also refines prognosis in undifferentiated/dedifferentiated carcinomas has yet to be studied. Massively parallel sequencing analysis of endometrial clear cell carcinomas revealed that subsets molecularly resembled either serous and/or endometrioid endometrial cancer, with *TP53*, *PIK3CA*, *PPP2R1A*, *ARID1A*, *PIK3R1*, and *SPOP* being most frequently affected by somatic mutations [73, 75]. Akin to endometrioid cancers, clear cell carcinomas could be stratified into the four molecular subtypes, including DNA MMR deficient and *POLE* ultramutated, and were associated with disease-free survival [73]. The molecular characterization of endometrial clear cell carcinomas highlighted the genetic heterogeneity of these tumors and suggests that their classification of being generally "type II" tumors may not be warranted [73].

Uterine carcinosarcomas have not yet been formally classified into molecular subtypes; however, several massively parallel sequencing studies revealed the genomic landscape of these lesions. Uterine carcinosarcomas are characterized by frequent mutations effecting *TP53* and chromatin-remodeling genes but also recurrent *PTEN*, *PIK3CA*, *PPP2R1A*, *FBXW7*, and *KRAS* mutations, similar to both endometrioid and serous endometrial cancers [76–78]. Furthermore, an epithelial-to-mesenchymal transition (EMT) gene signature was found in a subset of cases [77, 78]. It should be noted, however, that a small subset of uterine carcinosarcomas were found to be MSI-high or even to harbor *POLE* exonuclease domain mutations [76–78]. A formal classification of uterine carcinosarcomas into the molecular subtypes described for endometrioid and serous carcinomas, and their association with outcome has yet to be performed; however, genetic heterogeneity is also expected in this histologic subtype.

The molecular classification of endometrial cancers has also therapeutic implications. Both *POLE* (ultramutated) and MSI (hypermutated) endometrial cancers have been shown to have high neoantigen loads and high numbers of tumor infiltrating lymphocytes (TILs) and overexpress programmed cell death protein 1 (PD-1) and programmed death-ligand 1 (PD-L1) [79, 80], making them excellent candidates for immune checkpoint inhibitors. In clinical trials, immune checkpoint inhibitor therapy demonstrated durable antitumor activity in PD-L1-positive and MSI-high/DNA MMR-deficient advanced endometrial cancers [81, 82]. In May 2017, the PD-1 inhibitor pembrolizumab was approved by the Food and Drug Administration (FDA) for the treatment of adult and pediatric patients with unresectable or metastatic, MSI-high, or DNA MMR-deficient solid tumors, regardless of tumor site or histology [83]. With approximately 30% of endometrial cancers being MSI-high or DNA MMR deficient, this approval agnostic of organ site or histology will have an impact on the treatment of endometrial cancer.

With the advancements of sequencing technologies, we not only improved our understanding of cancer genomics, but also the implementation of genomic profiling in the clinical setting is becoming readily available. This shift toward genome-driven oncology care [84, 85] will likely affect the management of patients with endometrial cancer in the near future. In fact, the results of recent clinical trials targeting either specific oncogenic drivers in patients with endometrial cancer (e.g., HER2-positive serous endometrial cancers [86]) or genomically selected patients agnostic of histology/tumor type (so-called basket trials [84, 85]; e.g., AKT1 *E17K* mutated cancers [87]) may result in novel therapies being available to patients with endometrial carcinomas harboring specific genetic alterations.

The PI3K Pathway in Endometrial Cancer

The PI3K/AKT/mTOR pathway is altered in the vast majority of endometrioid endometrial cancers. The most frequently mutated member genes are *PTEN* (78%), the main negative regulator of the PI3K pathway; *PIK3CA* (53%) and *PIK3R1* (37%), the catalytic and regulatory subunits of PI3K, respectively; and *KRAS* (25%) [59], which interacts with p110alpha (i.e., the catalytic subunit of PI3K) [88]. Furthermore, 42% of serous carcinomas were reported to harbor a *PIK3CA* alteration [34, 59]. Unlike in some other cancer types, mutations in the PI3K pathway are not mutually exclusive, and coexistence of *PTEN* and *PIK3CA* mutations and/or *KRAS* or *PTEN* and *PIK3R1* and/or *KRAS* mutations in endometrioid endometrial cancers is common [34, 35, 59, 89] (Fig. 31.2).

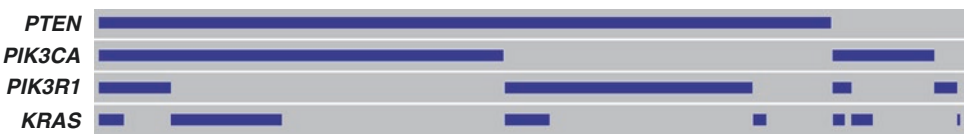


Fig. 31.2 Coexisting PI3K pathway mutations in endometrioid endometrial cancer. Endometrial carcinomas of endometrioid histology with mutations in *PTEN*, *PIK3CA*, *PIK3R1*, and/or *KRAS* (blue bars) were selected from The Cancer Genome Atlas study through cBioPortal [59, 267] (n = 187), and the patterns of co-occurring mutations affecting these genes are shown. Blue bars indicate presence of mutation

The functional consequences of harboring multiple co-occurring mutations in a tumor and their epistatic interactions are not yet entirely understood. It has been suggested that DNA MMR (see below) deficiency may partly contribute to the high frequency of mutations affecting different components of the PI3K pathway in endometrioid cancers [35]; however, this phenomenon seems to be equally frequent in MMR-proficient cases [59]. A significant proportion of *PIK3R1* alterations (p85; regulatory subunit of PI3K) has been found to be located within the iSH2 domain [34, 35], which mediates the binding of p85 to p110α and may constitute a mutational hotspot [90]. Given that not all *PIK3R1* mutations demonstrate gain of function in *in vitro* models [34, 35] and that neomorphic *PIK3R1* mutations activating the MAPK pathway have been identified [91], it is unclear whether they are functionally equivalent to activating mutations in *PIK3CA*. In addition, the spectrum of somatic *PIK3CA* mutations within endometrial carcinomas has been found to be more varied than that of colorectal and breast cancers [92]. In contrast to colorectal, brain, gastric, or breast cancers, where >75% of alterations have been shown to occur in two hotspots in the helical and kinase domains [93], in endometrial cancer, *PIK3CA* mutations are distributed throughout the gene, and somatic, activating mutations in the adapter-binding domain (ABD), ABD-linker region, and C2 domains of p110a are also frequent [35, 92].

Given the high prevalence of PI3K pathway alterations in endometrial cancer, clinical trials assessing the efficacy of inhibitors of this pathway in patients with endometrial cancer have been/are being performed (reviewed in [33, 41, 94]). First results from early clinical trials revealed that a subset of chemotherapy-naïve patients with advanced endometrial cancer are responsive to single-agent rapamycin analogs (i.e., allosteric mTOR inhibitors) and that therapeutic responses or stabilization of disease can be seen across histological types [6, 95–97]. These trials also suggest that not only endometrioid, but also a subset of serous carcinomas may be dependent on the PI3K pathway. Predictive markers associated with response to mTORC1 inhibitors have yet to be identified. The latter could be due to the small number and heterogeneous groups of endometrial cancer patients included in the trials to date. *In vitro* analyses of endometrioid endometrial cancer cell lines have suggested that inhibitors targeting different components of the PI3K pathway may be associated with distinct genomic predictors [98]. However, results from ongoing clinical trials are eagerly awaited.

Lynch Syndrome

Although the majority of endometrial cancers occur sporadically, approximately 2% of cases arise in the setting of the hereditary Lynch syndrome (also referred to as hereditary nonpolyposis colorectal carcinoma (HNPCC) syndrome) [99, 100]. This autosomal dominant disease is associated with germline mutations in DNA MMR genes, *MLH1*, *MSH2*, *MSH6*, or *PMS2,* as well as *EPCAM* deletions resulting in *MSH2* promoter methylation. The syndrome is associated with increased risk of endometrial, colorectal, gastric, and other cancers. The risk for endometrial cancer is estimated to be up to 60% [101–103]. Genetic alterations in DNA MMR genes lead to the accumulation of unstable microsatellite sequences throughout the genome (i.e., MSI). Women with Lynch syndrome who are diagnosed with endometrial carcinoma also have an increased risk of developing colorectal cancer [104] and several other tumors including breast cancer [105]; thus their recognition is important for family screening, early detection, or interventions to reduce the risk of additional cancers.

Identification of individuals with Lynch syndrome is based on the Amsterdam and Bethesda criteria [103, 106], which, in contrast to colorectal cancer, have proven ineffective for endometrial cancer patients [99]. In fact, guidelines for the identification of endometrial cancer patients who have Lynch syndrome have yet to be developed, and different screening algorithms have been suggested [107, 108].

A definite diagnosis of Lynch syndrome is established by germline mutational analysis of DNA MMR genes *MLH1*, *MSH2*, *MSH6*, and/or *PMS2*. Also, MSI analysis using polymerase chain reaction of two mononucleotide (BAT25, BAT26) and three dinucleotide (D2S123, D5S346, D17S250) microsatellite markers has been recommended by the National Cancer Institute (Bethesda guidelines) for the identification of Lynch patients [109]. It has been shown, however, that a panel of five mononucleotide repeat markers rather than mono- and dinucleotide markers as used in the Bethesda guidelines may provide a more accurate evaluation of tumor MSI in colorectal [106, 110] and endometrial cancers (BAT-25, BAT-26, NR-21, NR-24, and NR-27) [111]. When using the five Bethesda markers, high-frequency MSI (MSI-H) is present if two or more markers show instability, and low-frequency MSI (MSI-L) if one marker shows instability. In contrast, when using five mononucleotide repeat markers, three or more mutant alleles are typically required to indicate MSI-H [110]. It should be noted that endometrial cancer represents the most common clinical manifestation in female *MSH6* germline mutation carriers [112, 113]. However, carcinomas with *MSH6* mutations may be microsatellite stable (MSS) or MSI-L [107, 114] and, therefore, potentially may be missed by MSI analysis. In addition, as mentioned above, a subset of sporadic endometrial carcinomas show MSI due to *MLH1* promoter methylation, which cannot be differentiated by MSI analysis.

The role of the pathologist in the identification of patients with Lynch syndrome-associated endometrial cancer should not be underestimated. Immunohistochemical analysis of the

four MMR proteins, which has been shown to be sensitive and specific for the detection of germline MMR abnormalities and of tumors with MSI [99, 115], has been advocated in the context of endometrial cancers arising not only in patients with personal or family history of Lynch syndrome or Lynch syndrome-associated tumors, but generally in endometrial cancer patients of any age (both ages <60 and ≥60) [107, 108, 116, 117]. Loss of MLH1/PMS2 expression in the absence of *MLH1* promoter methylation, loss of MSH2 and/or MSH6 expression, and potentially loss of PMS2 alone indicate the presence of an MMR germline mutation and identify high-risk Lynch syndrome patients to be considered for germline testing.

With the introduction of cost-effective multiplex NGS-based test panels evaluating cancer susceptibility genes, including MMR genes [118], however, MMR mutational screening may be implemented in the near future as a primary screen rather than a confirmatory test in individuals with high risk of Lynch syndrome.

Uterine Mesenchymal Tumors

Introduction

Uterine mesenchymal tumors comprise a heterogeneous group of neoplasms, the overwhelming majority of which are leiomyomas. Malignant mesenchymal tumors are rare, accounting for 2–5% of uterine cancers. Differentiation of benign mesenchymal lesions from their malignant counterparts is crucial due to distinct clinical outcomes. Some mesenchymal tumors pose diagnostic challenges in surgical pathology practice, in particular endometrial stromal tumors (ESTs) and the biological spectrum of smooth muscle tumors (SMTs), as these may display variable and overlapping morphologic features. There have been several important developments in our understanding of the molecular genetics of these tumors over the recent years that have improved classification, diagnosis, and prediction of outcome. Massively parallel sequencing methods have also recently identified novel genetic alterations leading to the discovery of new uterine sarcoma subtypes.

Smooth Muscle Tumors

SMTs represent the most prevalent mesenchymal neoplasms in the uterus and are classified as leiomyoma, SMT of uncertain malignant potential (STUMP), and leiomyosarcoma of usual/spindle cell, epithelioid, or myxoid types [10]. Uterine leiomyoma represents the vast majority of uterine SMTs, which frequently affects women of reproductive age and is the leading indication for hysterectomy in the USA [119, 120]. In contrast, leiomyosarcoma represents only a small percentage of these neoplasms (40% of all uterine sarcomas; 1–3% of all uterine malignancies [121]). Leiomyosarcomas are thought to arise independently from leiomyomas [122]. They are highly aggressive tumors with a 5-year survival of 15–60% [123]. Diagnostic criteria for leiomyosarcomas vary by histologic subtype, but are generally based on a combination of histologic features including the presence of nuclear atypia, mitotic rate, and tumor cell necrosis (Fig. 31.3) [124]. Neoplasms that exhibit some, but not sufficient histologic criteria for the diagnosis of leiomyosarcoma are typically classified as STUMP. There is, however, significant interobserver variability in the assessment of these morphologic features [125, 126].

Currently available immunohistochemical markers are not reliable in distinguishing benign and malignant SMTs. ER and PR are positive in the vast majority of leiomyomas and up to 60% of leiomyosarcomas [127–133]. Diffuse p16 and aberrant p53 expression is usually present in leiomyosarcomas; however, these markers can also be detected in leiomyomas with bizarre nuclei (also referred as atypi-

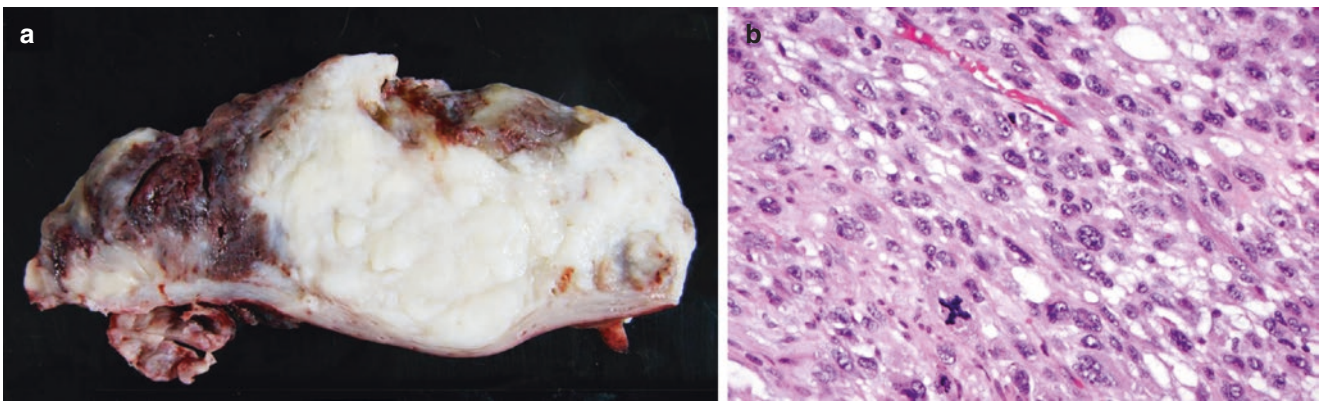


Fig. 31.3 Leiomyosarcoma. (**a**) Hemorrhage and necrosis are grossly evident in an otherwise white, whorled mass. (**b**) Marked nuclear pleomorphism and brisk mitotic activity with atypical mitotic figures are seen in a spindle cell leiomyosarcoma

cal or symplastic leiomyoma) and less frequently in STUMPs [134–144]. Elevated MIB1 (Ki-67) proliferation index can be seen in leiomyosarcoma, leiomyoma with bizarre nuclei, and mitotically active leiomyoma [137, 141–143]. Over the past years, various genetic alterations have been discovered in uterine leiomyomas. These include recurrent mutations in mediator complex subunit 12 (*MED12*), which have been reported in up to 80% of conventional uterine leiomyomas [121, 145–151]. *MED12* is a component of the mediator complex involved in the transcription of RNA polymerase II-dependent genes [152] and also plays a role in the Wnt/β-catenin and hedgehog signaling pathways [153, 154] and in the regulation of *NANOG* and *NANOG* target genes [155]. *MED12* mutations have also been found in a subset of leiomyoma variants, including cellular leiomyomas [156], leiomyomas with bizarre nuclei [150], and mitotically active leiomyomas [156]. In contrast, mutations in *MED12* are less frequently found in STUMPs [121, 156] and leiomyosarcomas [121, 147, 150, 151]. However, the number of such samples analyzed to date remains relatively small. While it is generally thought that leiomyosarcomas arise independently from leiomyomas [122], the presence of *MED12* mutations in a subset of leiomyosarcomas suggests that a small subgroup of leiomyosarcomas may, in fact, originate from leiomyomas. In addition, leiomyomas have been shown to harbor the recurrent t(12;14)(q15;q23~24) translocation in approximately 10% of cases or other 12q14~15 chromosomal rearrangements involving the *HMGA2* and *RAD51B* loci, rearrangements involving 6p21 and Xq22 (affecting *COL4A5* and *COL4A6*), deletions of 7q, and trisomy 12, among other genetic aberrations [157, 158] (Table 31.4). Interestingly, complex chromosomal rearrangements resembling chromothripsis [159] have been documented in these benign tumors, and a subset of physically distinct uterine leiomyomas from the same patient have been shown to be clonally related [158]. These observations suggest that even hallmarks of catastrophic genetic events, such as chromothripsis, cannot be employed to differentiate benign from malignant SMTs. Germline and somatic mutations in fumarate hydratase (*FH*), which encodes an enzyme that converts fumarate to malate in the tricarboxylic acid (Krebs) cycle, have been found in sporadic (i.e., non-hereditary) and hereditary leiomyomas with distinctive morphologic features. The latter is seen in hereditary leiomyomatosis renal cell carcinoma (HLRCC) syndrome, an autosomal dominant disorder characterized by uterine and cutaneous leiomyomatosis as well as renal cell carcinoma [160]. Both somatic and germline *FH* mutations lead to an accumulation of fumarate and formation of S-(2-succino)-cysteine (2SC). FH and 2SC antibodies have also been developed to detect this biochemical modification [161]. Leiomyomas with FH deficiency often exhibit prominent nucleoli, perinucleolar halos, and eosinophilic globules [162] (Fig. 31.4). These histologic findings are seen in both sporadic leiomyomas with bizarre nuclei as well as leiomyomas in patients with germline *FH* mutations [162–164]. More than 50% of leiomyomas with bizarre nuclei display aberrant FH/2SC expression defined by absent FH and diffuse 2SC staining [162]. And among leiomyomas with bizarre nuclei and aberrant FH/2SC expression, somatic *FH* genetic alterations, including homozygous deletions, missense mutations coupled with loss of heterozygosity, and splice site mutations, are detected in more than 90% of tumors [162]. A significant subset of leiomyomas with bizarre nuclei display normal FH/2SC expression, however, defined by diffuse FH and absent 2SC staining. These tumors frequently harbor *TP53* and/or *RB1* alterations compared to tumors with aberrant

Table 31.4 The most recurrent genomic aberrations in sporadic uterine mesenchymal tumors

	Chromosomal rearrangement	Gene copy number aberration	Mutated gene
Leiomyoma	t(12;14)(q15;q23~24) and other 12q14~15 (involving *HMGA2* and *RAD51B*) 6p21 Chromothripsis	del(7)(q22q32) Trisomy chromosome 12	*MED12* *FH*
Leiomyosarcoma	Complex and frequent 1p	Complex and frequent 1p deletions	*TP53* *ATRX* *RB1* *PTEN*
Endometrial stromal nodule	t(7;17)(p15;q21) (*JAZF1-SUZ12*)		
Low-grade endometrial stromal sarcoma	t(7;17)(p15;q21) (*JAZF1-SUZ12*) t(6;7)(p21;p15) (*JAZF1-PHF1*) t(6;10;10) (p21;q22;p11) (*PHF1-EPC1*) t(1;6)(p34;p21) (*MEAF6-PHF1*)		
High-grade endometrial stromal sarcoma	der(22)t(X;22) (p11;q13) *(ZC3H7B-BCOR)* t(10;17)(q22;p13) (*YWHAE–NUTM2*)		
Undifferentiated uterine sarcoma	t(7;17)(p15;q21) (*JAZF1-SUZ12*)* t(10;17)(q22;p13) (*YWHAE–NUTM2*)		*TP53*
Fibrosarcoma	*NTRK1* and *NTRK3* rearrangements		
Inflammatory myofibroblastic tumor	*ALK1* rearrangements		

*Only two cases have been described [186, 202]

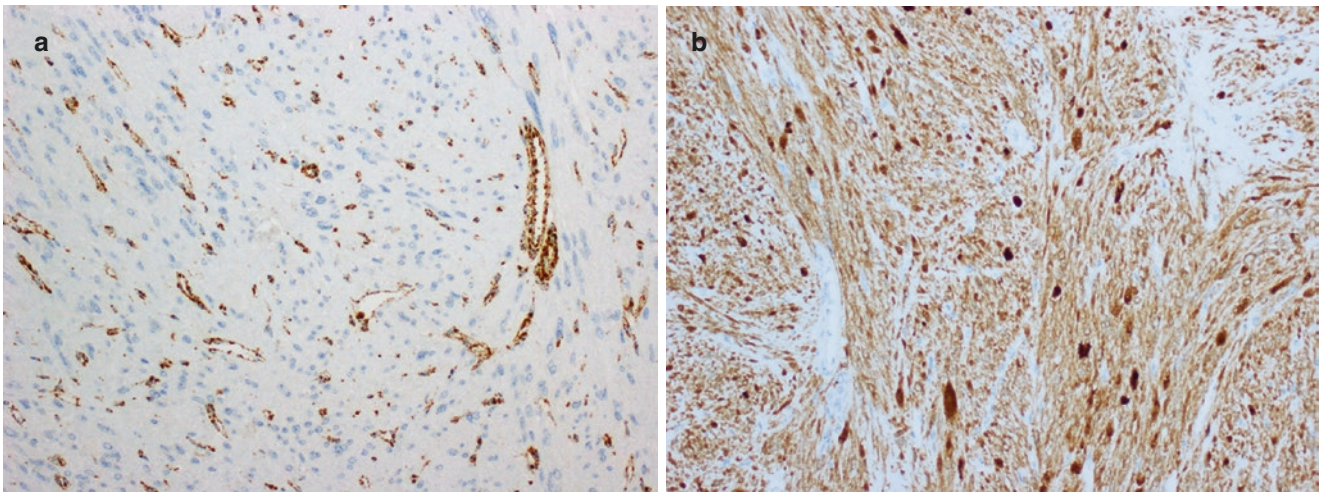


Fig. 31.4 Leiomyoma with fumarate hydratase deficiency. Aberrant FH/2SC expression is illustrated by (**a**) loss of FH and (**b**) diffuse 2SC expression

FH/2SC expression [162]. In clinical practice, aberrant FH/2SC expression in leiomyomas with bizarre nuclei is helpful in prompting review of personal medical and family history as well as referral to clinical genetics.

Leiomyosarcomas often display high levels of genetic instability, so much so that no identical karyotypes were found in 68 cases published in the literature between 1994 and 2004 (reviewed in [122]). There are widespread DNA copy number alterations including chromothripsis and whole-genome duplication [165]. Chromosomal alterations involving 1p translocations and deletions [166] have been reported in a subset of leiomyosarcomas. Recent studies have shown near-universal deletions and mutations in *TP53*, *RB1*, and *PTEN* [165, 167–170]. Alternative telomere lengthening due to recurrent alterations in telomere maintenance genes such as *ATRX*, *RBL2*, and *SP100* has also been detected in the majority of leiomyosarcomas [165, 169, 171]. Alterations in homologous recombination genes, structural rearrangements, and enrichment of specific mutational signatures have also been identified [165]. Compared to soft tissue leiomyosarcomas, uterine leiomyosarcomas show hypomethylation of the *ESR1* gene and a higher DNA damage response score [170]. It is currently unclear whether the genetic alterations recently described in leiomyomas, such as the chromosomal rearrangements affecting *COL4A5* and *COL4A6* [158], are unique to benign lesions.

In summary, no reliable immunohistochemical or molecular markers have been identified for the discrimination between benign and malignant SMT. The role of these ancillary markers is limited. If used at all, the above markers should be used in conjunction with thorough histologic analysis to establish a definitive diagnosis.

Endometrial Stromal Tumors

Endometrial stromal tumors (ESTs) are rare neoplasms that are classified into endometrial stromal nodule (ESN), low-grade endometrial stromal sarcoma (LGESS), high-grade endometrial stromal sarcoma (HGESS), and undifferentiated uterine sarcoma (UUS) according to the latest WHO classification [10]. Genomic approaches such as RNA sequencing have identified novel fusions among ESTs, particularly LGESS and HGESS, in recent years.

Endometrial Stromal Nodule and Low-Grade Endometrial Stromal Sarcoma

Both ESNs and LGESSs are composed of bland cells resembling proliferative endometrium stroma and can exhibit a number of variant histologic features. ESNs are benign and have well-circumscribed borders without invasion of surrounding tissue in contrast to LGESS that demonstrates infiltrative growth and vascular invasion. LGESSs are relatively indolent with a 10-year overall survival rate ranging from 65% to 76% [172].

Ancillary immunohistochemical studies may be useful in the diagnosis of ESN and LGESS. A panel of markers including CD10 and smooth muscle markers such as desmin, h-caldesmon, and HDAC8 has been recommended to discriminate ESNs and LGESSs from SMTs [173–175]. It should be noted that the expression of these markers should be interpreted in conjunction with tumor morphology given that areas of smooth muscle differentiation may express any of the aforementioned smooth muscle markers. Inhibin, calretinin, CD99, melan A, and WT-1 which are usually expressed in sex cord stromal tumors can also be found in "sex cord-like" foci of ESTs [176–182]. In addition, LGESSs

frequently express ER and PR [183–188], nuclear β-catenin [186, 189–192], EGFR [193–195], and c-kit [194, 196] and less commonly androgen receptor [197], aromatase [198], PDGF-α, PDGF-β, or VEGF [196, 199]. It should be noted, however, that mutations in *KIT*, *PDGFR-α*, *PDGFR-β*, or *EGFR* have not been detected in LGESS [194, 195]; hence, the proteins they encode are unlikely to constitute optimal therapeutic targets for LGESS.

ESNs and LGESSs are characterized by recurrent chromosomal translocations (Table 31.4), including the t(7;17)(p15;q21), t(6;7)(p21;p15), t(6;10;10)(p21;q22;p11), t(1;6)(p34;p21), and 46,XX,t(5;6)(q31;p21), which result in the formation of the chimeric transcripts *JAZF1-SUZ12*, *JAZF1-PHF1*, *EPC1-PHF1*, *MEAF6-PHF1*, *and BRD8-PHF1*, respectively (reviewed in [200, 201]). When present, gene fusions appear to be mutually exclusive and have been observed in both conventional ESNs and LGESSs and less frequently tumors that display variant morphology. The fusion of *JAZF1* and *SUZ12* (also known as *JJAZ1*) is the most common genetic alteration in ESTs and has been reported in 68% of ESNs, 43% of LGESSs, and 9% of UUSs [186, 200, 202–208]. *JAZF1-PHF1* and *EPC1-PHF1* rearrangements have been observed in 10% and 6% of LGESSs, respectively [200, 207, 209]. *MEAF6-PHF1*, *MBTD1-CXorf67*, and *BRD8-PHF1* gene fusions have recently been characterized in rare LGESSs [201, 210–212], and *PHF1* and *JAZF1* rearrangements without known partners have also been observed [207, 208]. The functions of the chimeric proteins resulting from the various gene rearrangements identified in ESTs are not resolved at this time. The most common fusion joins the first three exons of *JAZF1*, the function of which is relatively unknown, to the last 15 exons of *SUZ12*, a component of the polycomb repressive complex 2 [213]. Low levels of *JAZF1-SUZ12* mRNA, which are thought to arise from trans-splicing of precursor mRNAs for *JAZF1* and *SUZ12* genes, and the JAZF1-SUZ12 protein have been detected in normal endometrial stromal cell lines and in normal late-secretory and early-proliferative phase endometrium [213, 214]. Based on these findings, it has been suggested that acquisition of the *JAZF1-SUZ12* fusion via chromosomal translocation may be an early event in the pathogenesis of ESTs. Expression of JAZF1-SUZ12 has been shown to promote cell proliferation in vitro, however, only when accompanied by suppression of endogenous wild-type SUZ12 (from the unarranged allele) [213]. Interestingly, the non-rearranged wild-type *JAZF1* allele is active in ESNs, but silenced in LGESSs harboring the *JAZF1-SUZ12* fusion, providing evidence to suggest that LGESSs arise from ESNs via genetic or epigenetic silencing of the non-rearranged allele [213]. The presence of *JAZF1-SUZ12* fusion in rare UUSs also raises the possibility that a small subset of ESNs and LGESSs may transform into UUSs via dedifferentiation [186, 202]; it should be noted, however, that the vast majority of LGESSs are unlikely to stem from ESNs, given that these tumors appear to be genetically distinct.

High-Grade Endometrial Stromal Sarcoma

In the current WHO classification, HGESS describes tumors specifically harboring the t(10;17)(q22-23;p13) translocation resulting in fusion of *YWHAE* and *NUTM2* (also known as *FAM22*) genes [10]. Tumors with this fusion exhibit a high-grade round cell component consisting of sheets, nests, and, less frequently, rosettes, sex cord-like and pseudoglandular/pseudopapillary patterns of cells with enlarged, intermediate-sized round nuclei associated with increased mitotic activity and tumor cell necrosis (Fig. 31.5) [187, 215]. A subset of these tumors also demonstrates a low-grade fibroblastic or fibromyxoid component reminiscent of LGESS [187, 216,

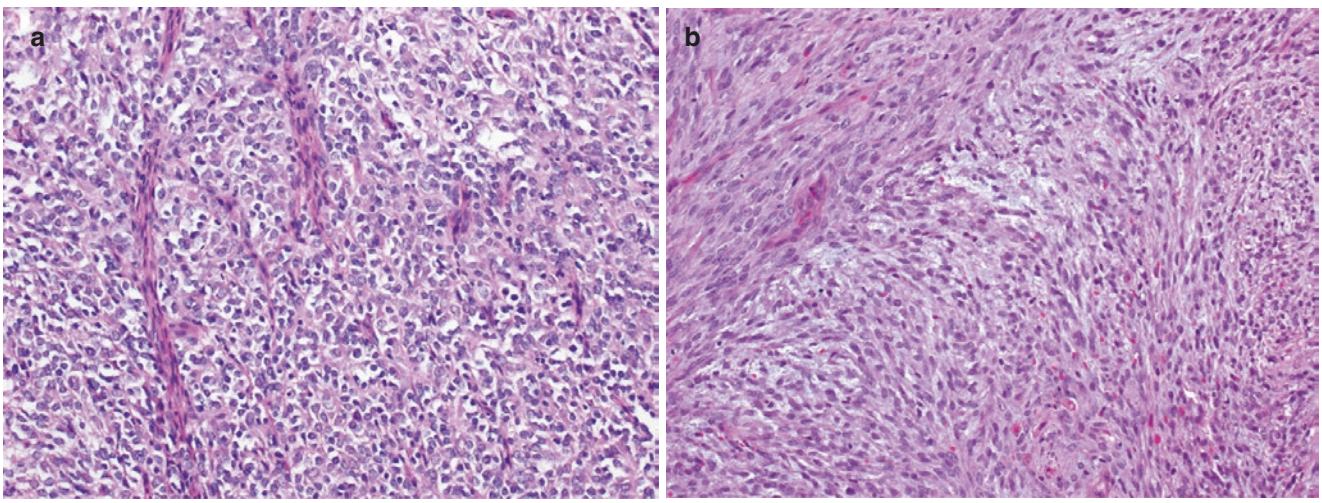


Fig. 31.5 High-grade Endometrial stromal sarcoma. (**a**) *YWHAE-NUTM2* high-grade endometrial stromal sarcoma consists of a high-grade round cell component. (**b**) Haphazard fascicles of spindle cells are embedded in myxoid matrix in a *ZC3H7B-BCOR* high-grade endometrial stromal sarcoma

217]. There have been few reports of *YWHAE-NUTM2* fusion found in tumors with only low-grade morphology [218, 219]. Limited clinical data suggest that the prognosis of these tumors is intermediate between that of LGESS and UUS [187, 217].

While the low-grade spindle cell component of HGESS harboring *YWHAE-NUTM2* fusion shares a similar immunoprofile with conventional LGESS and displays CD10, ER, and PR expression, CD10 is typically negative, and hormone receptors are often absent or only focally weakly-to-moderately expressed in the high-grade round cell component. By gene expression profiling, *CCND1* (cyclin D1) has also been reported to be differentially expressed between *YWHAE-NUTM2* HGESS and conventional LGESS, with cyclin D1 being a sensitive and specific diagnostic marker for the identification of *YWHAE-NUTM2* HGESS [220]. BCOR has also recently emerged as a robust diagnostic marker of *YWHAE-NUTM2* HGESS and is strongly and diffusely expressed in the high-grade round cell component, while absent or only focally expressed in the low-grade spindle cell component [221]. The mechanism leading to cyclin D1 and BCOR overexpression in *YWHAE-NUTM2* rearranged tumors remains to be elucidated.

YWHAE-NUTM2 fusions have been shown to lead to nuclear accumulation of the functionally intact 14-3-3ε (*YWHAE*) protein-interaction domain, thereby likely redirecting known cytoplasmic 14-3-3ε protein interactions with phosphoserine-containing proteins to the nuclear compartment [216]. Of interest, *YWHAE-NUTM2* fusion has also been described in a subset of clear cell sarcomas of the kidney [222, 223], but has not been found in other uterine and non-gynecological mesenchymal tumors, adenosarcomas, or carcinosarcomas [216].

ESSs harboring *ZC3H7B-BCOR* gene fusion as a result of the t(X;22)(p11;q13) translocation have recently emerged as another type of HGESS with a distinctive phenotype [224, 225]. *ZC3H7B-BCOR* HGESS is composed of haphazard fascicles of spindle cells displaying mild to moderate nuclear atypia, often associated with abundant myxoid matrix and brisk mitotic activity (Fig. 31.5). Collagen plaques are seen in approximately 50% of tumors. Tongue-like and/or pushing myometrial invasion is typical. Unlike *YWHAE-NUTM2* HGESS, there is no conventional or variant LGESS component [224, 225]. CD10 expression is present and usually diffuse; however, in contrast to LGESS, ER and PR expression is seen in approximately 30% of tumors. While *ZC3H7B-BCOR* HGESS has extensive morphologic overlap with myxoid leiomyosarcoma, it exhibits only limited or absent desmin, SMA, and/or h-caldesmon staining. Diffuse cyclin D1 staining is seen in >80% of tumors, while diffuse BCOR expression is only seen in half. BCOR expression by immunohistochemistry does not appear to correlate with *BCOR* breakpoints in the fusion transcripts [221]. Limited clinical data suggest that patients present at high stage and have worse prognosis compared with published outcomes in LGESS [224, 225].

BCOR internal tandem duplication (ITD) involving exons 15 and 16 has recently been discovered in rare uterine sarcomas that likely represent yet another type of HGESS with distinctive histologic features. HGESS with *BCOR* ITD exhibit a high-grade round cell component and a low-grade spindle cell component similar to *YWHAE-NUTM2* HGESS as well as a high-grade spindle cell component and myxoid stroma reminiscent of *ZC3H7B-BCOR* HGESS. Mitotic activity is brisk, and necrosis is often present. Tongue-like myometrial infiltration and vascular invasion as seen in most ESS are often evident. Many tumors express at least focal CD10 as well as diffuse and strong cyclin D1 and BCOR immunoreactivity [221, 226]. While these are morphologically high-grade sarcomas, given the rarity and only recent discovery of HGESS with *BCOR* ITD, it is difficult to ascertain its clinical behavior at this time.

The BCL-6 corepressor (*BCOR*) gene encodes the BCOR protein, which interacts with PCGF1 within a variant polycomb repressive complex (PRC1) that suppresses gene expression by histone modification [227, 228]. Germline *BCOR* mutations have been found in syndromic diseases such as Lenz microphthalmia and oculo-facio-cardio-dental syndrome [229]. Somatic *BCOR* mutations have been reported in hematologic malignancies [230–233], rhabdomyosarcoma [234], retinoblastoma [235], medulloblastoma [236], central nervous system primitive neuroectodermal tumor [237, 238], primary renal sarcomas [239–243], and round cell sarcomas of the soft tissues [244–247]. *ZC3H7B-BCOR* gene fusions are not limited to HGESS and have also been described in ossifying myofibroblastic tumors [248] with myxoid change. *BCOR* ITD involves duplication of the *BCOR* region within a PUFD domain that facilitates binding with PCGF1 and may affect the function of PRC1 in epigenetic modification. BCOR ITD has been found clear cell sarcoma of the kidney [239–241, 243] as well as round cell sarcomas of the soft tissues [246, 247] that share morphologic overlap with HGESS harboring the same aberration.

Undifferentiated Uterine Sarcoma

By definition, UUS demonstrates significant cytologic atypia, bears no morphologic similarity to endometrial stroma, often demonstrates destructive myometrial invasion, and reportedly has a poor prognosis (median overall survival 1–3 years) [172, 249–251]. Given the morphologic heterogeneity observed in UUS, some authors have suggested further subtyping of UUS into uniform (UUS-U) and pleomorphic types (UUS-P) to stratify sarcomas into those that demonstrate uniform cytologic atypia with enlarged, hyperchromatic nuclei and prominent nucleoli and those that exhibit marked nuclear pleomorphism, respectively [186]. ER and

PR are expressed in up to 50% of UUS-U, but not in UUS-P [186, 188, 217]. Most UUS-U and approximately one-third of UUS-P express nuclear β-catenin, whereas aberrant p53 expression is more often associated with UUS-P than UUS-U [186, 188, 217]. Cyclin D1 immunoexpression has been found only rarely in UUS-U [220]. However, UUS remains a diagnosis of exclusion with no morphologic or immunohistochemical line of cell differentiation. Genetic aberrations in UUS are essentially unknown with the exception of *YWHAE-NUTM2* [252] and *JAZF1-SUZ12* [216] fusions identified in rare tumors, suggesting that tumors classified as UUS may represent misdiagnosed or dedifferentiated ESS [252]. One recent retrospective study showed that approximately 80% of tumors classified as UUS harbored HGESS-associated mutations, such as *YWHAE* and *BCOR* genetic abnormalities, by fluorescence in situ hybridization and next-generation targeted RNA sequencing [253].

Other Rare Uterine Mesenchymal Tumors

NTRK rearrangements have recently been reported in uterine spindle cell sarcomas morphologically resembling soft tissue fibrosarcoma [254] (Fig. 31.6). This novel uterine sarcoma subtype typically arises in the cervix and less commonly in the corpus and affects women of premenopausal age. Tumors are composed of spindle cells with moderate cytologic atypia that may form a storiform or herringbone pattern. Marked nuclear pleomorphism with bizarre cells resembling UUS-P may be seen, but fascicles of intermediate-grade spindle cells are usually seen in the background. The mitotic index is often brisk, and necrosis may be present. The immunoprofile is non-specific; there may be focal SMA and S100 positivity, but desmin, ER, PR, CD34, and SOX10 are negative, and H3K27me3 expression is retained [254]. TrkA and/or pan-Trk immunoexpression is seen among all *NTRK* fusion-positive uterine sarcomas; however, depending on the fusion partner, the type of staining may vary. These tumors are underpinned by *NTRK1* and *NTRK3* rearrangements with various fusion partners.

Neurotrophic tyrosine kinase receptors *NTRK1*, *NTRK2*, and *NTRK3* genes encode TrkA, TrkB, and TrkC proteins, which bind neurotrophins, thereby inducing receptor dimerization, phosphorylation, and activation of downstream signaling cascades via PI3K, RAS/MAPK/ERK, and PLC-gamma [255–257]. *NTRK* rearrangements are seen in a variety of common cancers but have not been associated with unique phenotypes until the identification of *NTRK* fusion-positive uterine sarcomas. While the clinical behavior of *NTRK* fusion-positive uterine sarcomas is uncertain due to their rarity, limited clinical data suggest an aggressive clinical course. Larotrectinib and entrectinib are two leading Trk inhibitors that have shown promising results in *NTRK* fusion-positive tumors [258, 259]. Patients with this type of uterine sarcoma who develop progressive disease may benefit from enrollment in clinical trials with Trk inhibitors.

Inflammatory myofibroblastic tumors (IMT) are rare in the uterus, but share identical morphologic features with IMTs located elsewhere. They are composed of spindle cells that are associated with myxoid, compact, or hyalinized patterns. Nuclear atypia is often mild to moderate, with rare examples exhibiting marked pleomorphism. Inflammation is usually conspicuous and consists of a lymphoplasmacytic infiltrate. While many patients with uterine IMT have a favorable prognosis, some may develop progressive disease. ALK expression is often present and associated with *ALK* rearrangements with a variety of fusion partners [260–262].

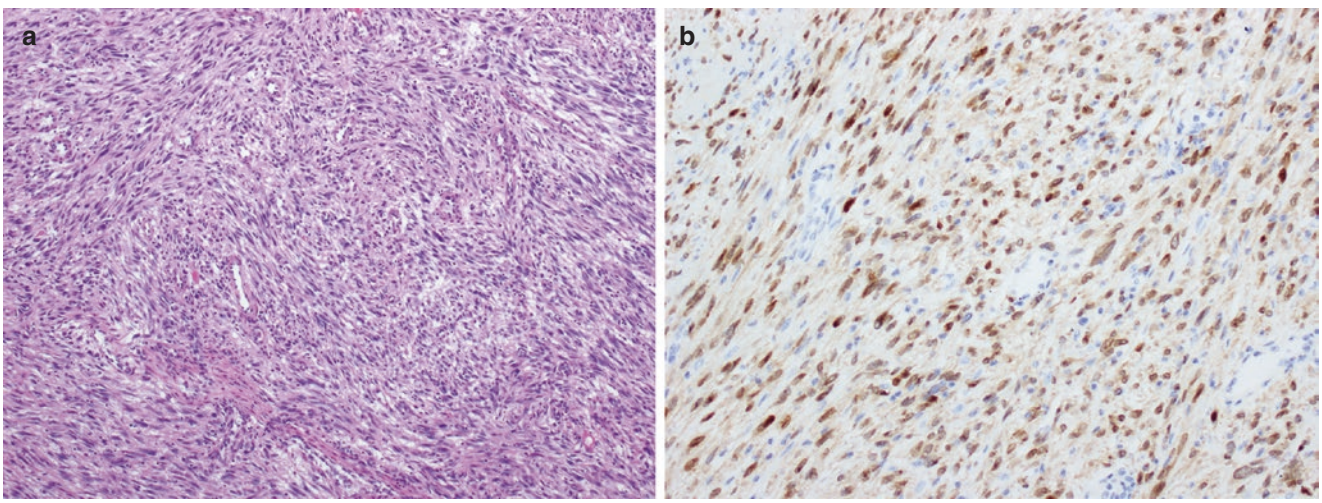


Fig. 31.6 *NTRK* fusion-positive uterine sarcoma with features of fibrosarcoma. (**a**) Spindle cells with moderate nuclear pleomorphism form a storiform pattern and (**b**) exhibit diffuse cytoplasmic and strong perinuclear Trk A expression

The characterization of the genetic underpinning of uterine mesenchymal neoplasms is providing fertile ground for the development of molecular genetic tests for their accurate diagnosis, opportunities for the development of immunohistochemical surrogate markers for the presence of these genetic aberrations, and novel targeted therapies. With the efforts of TCGA, the International Cancer Genome Consortium (ICGC), and individual investigators, a wealth of data on the repertoire of genetic alterations found in rare types of uterine cancers has and will become available. These data will be instrumental to find answers for questions that have puzzled pathologists for decades.

Summary

Endometrial carcinoma is a heterogeneous disease and comprises a spectrum of lesions with distinct histological features, clinical behaviors, and molecular characteristics. The dualistic Bokhman model [8] has been of conceptual use; however, it does not cover the histological and molecular heterogeneity of the disease, and its prognostic value is limited. Mesenchymal neoplasms of the uterus comprise a diverse group of tumors and, albeit uncommon, pose important diagnostic challenges. The recent genomic analyses of uterine malignancies, together with the knowledge acquired from the study of other cancer types, are likely to provide a solid basis for the development of an improved prognostic classification system and of targeted therapies for patients with these diseases [31]. We anticipate that diagnostic and predictive biomarkers based on the genetic features of uterine cancers are likely to be incorporated into the armamentarium of diagnostic gynecological pathologists in the near future.

Outlook

The ongoing efforts to characterize the genetic landscapes of cancers are expected to provide further insights into cancers of the uterine corpus and lead to the introduction of genomic applications for endometrial cancer in the near future. The insights from the integrated genomic characterization of endometrial carcinoma by TCGA [59] have called the distinct nature of type I and type II endometrial cancers into question. In fact, these two subsets of endometrial cancers are highly heterogeneous, and similarities between entities classified as type I or type II have been documented (e.g., a subset of FIGO grade 3 endometrioid cancers shows a serous-like genomic profile).

It is probable that with the burgeoning information on the genetic makeup of different types of endometrial cancer, a combined morphological and molecular classification will emerge, which may result in more accurate diagnosis, in particular of high-grade lesions. Based on the results of the integrative analysis carried out by TCGA, one may envision that *POLE* sequencing analysis will soon be used to identify patients with a favorable expected outcome and to discriminate them from potential histologic mimics such as serous carcinomas. On the other hand, the presence of *TP53* mutations in endometrioid endometrial cancers may define the subset of endometrioid carcinomas with a serous-like aggressive behavior. In addition, it has been demonstrated that endometrial and ovarian cancer cells can be detected through massively parallel sequencing analysis of DNA extracted from routine liquid-based cervical cytological specimens/liquids, which may constitute a first step toward a new generation of cancer screening tests [263, 264].

The recent identification of recurrent mutations and fusion genes in uterine mesenchymal tumors is expected to lead to the development of molecular tests, based on in situ hybridization, reverse transcriptase (RT)-PCR, and RNA sequencing, or even immunohistochemical tests, such as cyclin D1 and BCOR expression for the detection of HGESS harboring *YWHAE-NUTM2* fusions and *BCOR* abnormalities [220, 221] as well as Trk A and pan-Trk expression for identification of *NTRK* fusion-positive uterine sarcomas [254].

The integration of biomarker assessment and/or routine tissue collection in clinical trials testing novel targeted agents should be encouraged, and this source of biological material is absolutely essential for the identification of predictive markers. PI3K pathway, including AKT1 inhibitors, is thought to have great potential as targeted therapies in subsets of patients with endometrial carcinoma (reviewed in [33, 41, 86, 87, 94]). The Trk inhibitor larotrectinib has shown promising results in patients with various *NTRK* fusion-positive tumor types [258] and may be beneficial for patients with *NTRK* fusion-positive uterine sarcomas who develop progressive disease. The immune checkpoint inhibitor pembrolizumab has been approved for the treatment of MSI-high or DNA MMR-deficient solid tumors, including endometrial cancers [83]. It should be noted, however, that from a clinical perspective, the patients most in need for more effective therapies are those with advanced, recurrent, or metastatic disease. Genomic studies, however, have so far preferentially focused on the analysis of primary tumors. Conversely, early clinical trials are usually performed in pretreated patients with advanced disease. Therefore, germane to the realization of the potential of precision medicine is that studies ascertaining whether the genetic landscapes and distribution of molecular subtypes of metastatic endometrial cancer are similar or distinct from those of early lesions [265] are carried out.

In the past, endometrial cancer was often perceived as a "benign disease" given the large proportion of women with low-risk disease and high cure rates [266]. Advances in our understanding of uterine cancer revealed the complexity of

the disease, in particular at the molecular level. This information may result in a refinement of the taxonomy of uterine neoplasms and in the identification of genetic biomarkers required for the introduction of molecular target-based therapy for the treatment of women with this disease.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30. https://doi.org/10.3322/caac.21442.
2. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson B, Mariotto A, Lake AJ, Wilson R, Sherman RL, Anderson RN, Henley SJ, Kohler BA, Penberthy L, Feuer EJ, Weir HK. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. J Natl Cancer Inst. 2017;109(9). https://doi.org/10.1093/jnci/djx030.
3. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, Heintz AP, Ngan HY, Pecorelli S. Carcinoma of the corpus uteri. FIGO 26th annual report on the results of treatment in gynecological cancer. Int J Gynaecol Obstet. 2006;1(95 Suppl):S105–43. https://doi.org/10.1016/S0020-7292(06)60031-3.
4. Makker V, Hensley ML, Zhou Q, Iasonos A, Aghajanian CA. Treatment of advanced or recurrent endometrial carcinoma with doxorubicin in patients progressing after paclitaxel/carboplatin: memorial sloan-kettering cancer center experience from 1995 to 2009. Int J Gynecol Cancer. 2013;23:929–34. https://doi.org/10.1097/IGC.0b013e3182915c20.
5. Carey MS, Gawlik C, Fung-Kee-Fung M, Chambers A, Oliver T. Systematic review of systemic therapy for advanced or recurrent endometrial cancer. Gynecol Oncol. 2006;101(1):158–67. https://doi.org/10.1016/j.ygyno.2005.11.019.
6. Oza AM, Elit L, Tsao MS, Kamel-Reid S, Biagi J, Provencher DM, Gotlieb WH, Hoskins PJ, Ghatage P, Tonkin KS, Mackay HJ, Mazurka J, Sederias J, Ivy P, Dancey JE, Eisenhauer EA. Phase II study of temsirolimus in women with recurrent or metastatic endometrial cancer: a trial of the NCIC clinical trials group. J Clin Oncol. 2011;29(24):3278–85. https://doi.org/10.1200/JCO.2010.34.1578.
7. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. Int J Gynaecol Obstet. 2009;105(2):103–4.
8. Bokhman JV. Two pathogenetic types of endometrial carcinoma. Gynecol Oncol. 1983;15(1):10–7.
9. Setiawan VW, Yang HP, Pike MC, McCann SE, Yu H, Xiang YB, Wolk A, Wentzensen N, Weiss NS, Webb PM, van den Brandt PA, van de Vijver K, Thompson PJ, Strom BL, Spurdle AB, Soslow RA, Shu XO, Schairer C, Sacerdote C, Rohan TE, Robien K, Risch HA, Ricceri F, Rebbeck TR, Rastogi R, Prescott J, Polidoro S, Park Y, Olson SH, Moysich KB, Miller AB, McCullough ML, Matsuno RK, Magliocco AM, Lurie G, Lu L, Lissowska J, Liang X, Lacey JV Jr, Kolonel LN, Henderson BE, Hankinson SE, Hakansson N, Goodman MT, Gaudet MM, Garcia-Closas M, Friedenreich CM, Freudenheim JL, Doherty J, De Vivo I, Courneya KS, Cook LS, Chen C, Cerhan JR, Cai H, Brinton LA, Bernstein L, Anderson KE, Anton-Culver H, Schouten LJ, Horn-Ross PL. Type I and II endometrial cancers: have they different risk factors? J Clin Oncol. 2013;31:2607–18. https://doi.org/10.1200/JCO.2012.48.2596.
10. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO classification of tumours of female reproductive organs, World health organization classification of tumors. Lyon: IARC Press; 2014.
11. Clement PB, Young RH. Atlas of gynecologic surgical pathology. 3rd ed. Philadelphia: Saunders/Elsevier; 2013.
12. Blaustein A, Kurman RJ. Blaustein's pathology of the female genital tract. 6th ed. New York: Springer; 2011.
13. McCluggage WG. Uterine carcinosarcomas (malignant mixed Mullerian tumors) are metaplastic carcinomas. Int J Gynecol Cancer. 2002;12(6):687–90.
14. McCluggage WG. Malignant biphasic uterine tumours: carcinosarcomas or metaplastic carcinomas? J Clin Pathol. 2002;55(5):321–5.
15. Announcement. FIGO stages-1988 revision. Gynecol Oncol. 1989;35:125–6.
16. Zaino RJ, Kurman RJ, Diana KL, Morrow CP. The utility of the revised international federation of gynecology and obstetrics histologic grading of endometrial adenocarcinoma using a defined nuclear grading system. a gynecologic oncology group study. Cancer. 1995;75(1):81–6.
17. Alkushi A, Abdul-Rahman ZH, Lim P, Schulzer M, Coldman A, Kalloger SE, Miller D, Gilks CB. Description of a novel system for grading of endometrial carcinoma and comparison with existing grading systems. Am J Surg Pathol. 2005;29(3):295–304.
18. Lax SF, Kurman RJ, Pizer ES, Wu L, Ronnett BM. A binary architectural grading system for uterine endometrial endometrioid carcinoma has superior reproducibility compared with FIGO grading and identifies subsets of advance-stage tumors with favorable and unfavorable prognosis. Am J Surg Pathol. 2000;24(9):1201–8.
19. Prat J. Prognostic parameters of endometrial carcinoma. Hum Pathol. 2004;35(6):649–62.
20. Sorbe B. Predictive and prognostic factors in definition of risk groups in endometrial carcinoma. ISRN Obstet Gynecol. 2012;2012:325790. https://doi.org/10.5402/2012/325790.
21. Gilks CB, Oliva E, Soslow RA. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. Am J Surg Pathol. 2013;37(6):874–81. https://doi.org/10.1097/PAS.0b013e31827f576a.
22. Hoang LN, Kinloch MA, Leo JM, Grondin K, Lee CH, Ewanowich C, Kobel M, Cheng A, Talhouk A, McConechy M, Huntsman DG, McAlpine JN, Soslow RA, Gilks CB. Interobserver agreement in endometrial carcinoma histotype diagnosis varies depending on the cancer genome atlas (TCGA)-based molecular subgroup. Am J Surg Pathol. 2017;41(2):245–52. https://doi.org/10.1097/PAS.0000000000000764.
23. Soslow RA. High-grade endometrial carcinomas – strategies for typing. Histopathology. 2013;62(1):89–110. https://doi.org/10.1111/his.12029.
24. Clarke BA, Gilks CB. Endometrial carcinoma: controversies in histopathological assessment of grade and tumour cell type. J Clin Pathol. 2010;63(5):410–5. https://doi.org/10.1136/jcp.2009.071225.
25. Mittal K, Soslow R, McCluggage WG. Application of immunohistochemistry to gynecologic pathology. Arch Pathol Lab Med. 2008;132(3):402–23. https://doi.org/10.1043/1543-2165(2008)132[402,AOITGP]2.0.CO;2.
26. Fadare O, Liang SX. Diagnostic utility of hepatocyte nuclear factor 1-Beta immunoreactivity in endometrial carcinomas: lack of specificity for endometrial clear cell carcinoma. Appl Immunohistochem Mol Morphol. 2012;20(6):580–7. https://doi.org/10.1097/PAI.0b013e31824973d1.
27. Lim D, Ip PP, Cheung AN, Kiyokawa T, Oliva E. Immunohistochemical comparison of ovarian and uterine endometrioid carcinoma, endometrioid carcinoma with clear cell change, and clear cell carcinoma. Am J Surg Pathol. 2015;39(8):1061–9. https://doi.org/10.1097/PAS.0000000000000436.
28. Tafe LJ, Garg K, Chew I, Tornos C, Soslow RA. Endometrial and ovarian carcinomas with undifferentiated components:

clinically aggressive and frequently underrecognized neoplasms. Mod Pathol. 2010;23(6):781–9. https://doi.org/10.1038/modpathol.2010.41.
29. Ramalingam P, Masand RP, Euscher ED, Malpica A. Undifferentiated carcinoma of the endometrium: an expanded immunohistochemical analysis including PAX-8 and Basal-Like carcinoma surrogate markers. Int J Gynecol Pathol. 2016;35(5):410–8. https://doi.org/10.1097/PGP.0000000000000248.
30. Hussein YR, Broaddus R, Weigelt B, Levine DA, Soslow RA. The genomic heterogeneity of FIGO grade 3 endometrioid carcinoma impacts diagnostic accuracy and reproducibility. Int J Gynecol Pathol. 2016;35(1):16–24. https://doi.org/10.1097/PGP.0000000000000212.
31. McAlpine J, Leon-Castillo A, Bosse T. The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses. J Pathol. 2018;244(5):538–49. https://doi.org/10.1002/path.5034.
32. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, Kovatich AJ, Palazzo JP, Fishel R, Goodfellow PJ. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. Hum Mol Genet. 1999;8(4):661–6.
33. Dedes KJ, Wetterskog D, Mendes-Pereira AM, Natrajan R, Lambros MB, Geyer FC, Vatcheva R, Savage K, Mackay A, Lord CJ, Ashworth A, Reis-Filho JS. PTEN deficiency in endometrioid endometrial adenocarcinomas predicts sensitivity to PARP inhibitors. Sci Transl Med. 2010;2(53):53ra75. https://doi.org/10.1126/scitranslmed.3001538.
34. Urick ME, Rudd ML, Godwin AK, Sgroi D, Merino M, Bell DW. PIK3R1 (p85alpha) is somatically mutated at high frequency in primary endometrial cancer. Cancer Res. 2011;71(12):4061–7. https://doi.org/10.1158/0008-5472.CAN-11-0549.
35. Cheung LW, Hennessy BT, Li J, Yu S, Myers AP, Djordjevic B, Lu Y, Stemke-Hale K, Dyer MD, Zhang F, Ju Z, Cantley LC, Scherer SE, Liang H, Lu KH, Broaddus RR, Mills GB. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. Cancer Discov. 2011;1(2):170–85. https://doi.org/10.1158/2159-8290.CD-11-0039.
36. Wiegand KC, Lee AF, Al-Agha OM, Chow C, Kalloger SE, Scott DW, Steidl C, Wiseman SM, Gascoyne RD, Gilks B, Huntsman DG. Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. J Pathol. 2011;224(3):328–33. https://doi.org/10.1002/path.2911.
37. McConechy MK, Anglesio MS, Kalloger SE, Yang W, Senz J, Chow C, Heravi-Moussavi A, Morin GB, Mes-Masson AM, Carey MS, McAlpine JN, Kwon JS, Prentice LM, Boyd N, Shah SP, Gilks CB, Huntsman DG. Subtype-specific mutation of PPP2R1A in endometrial and ovarian carcinomas. J Pathol. 2011;223(5):567–73. https://doi.org/10.1002/path.2848.
38. Santin AD, Bellone S, Van Stedum S, Bushen W, Palmieri M, Siegel ER, De Las Casas LE, Roman JJ, Burnett A, Pecorelli S. Amplification of c-erbB2 oncogene: a major prognostic indicator in uterine serous papillary carcinoma. Cancer. 2005;104(7):1391–7. https://doi.org/10.1002/cncr.21308.
39. Hetzel DJ, Wilson TO, Keeney GL, Roche PC, Cha SS, Podratz KC. HER-2/neu expression: a major prognostic factor in endometrial cancer. Gynecol Oncol. 1992;47(2):179–85.
40. McConechy MK, Ding J, Cheang MC, Wiegand KC, Senz J, Tone AA, Yang W, Prentice LM, Tse K, Zeng T, McDonald H, Schmidt AP, Mutch DG, McAlpine JN, Hirst M, Shah SP, Lee CH, Goodfellow PJ, Gilks CB, Huntsman DG. Use of mutation profiles to refine the classification of endometrial carcinomas. J Pathol. 2012;228(1):20–30. https://doi.org/10.1002/path.4056.
41. Salvesen HB, Haldorsen IS, Trovik J. Markers for individualised therapy in endometrial carcinoma. Lancet Oncol. 2012;13(8):e353–61. https://doi.org/10.1016/S1470-2045(12)70213-9.
42. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–52. https://doi.org/10.1038/35021093.
43. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J Pathol. 2010;220(2):263–80. https://doi.org/10.1002/path.2648.
44. Risinger JI, Maxwell GL, Chandramouli GV, Jazaeri A, Aprelikova O, Patterson T, Berchuck A, Barrett JC. Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer. Cancer Res. 2003;63(1):6–11.
45. Maxwell GL, Chandramouli GV, Dainty L, Litzi TJ, Berchuck A, Barrett JC, Risinger JI. Microarray analysis of endometrial carcinomas and mixed mullerian tumors reveals distinct gene expression profiles associated with different histologic types of uterine cancer. Clin Cancer Res. 2005;11(11):4056–66. https://doi.org/10.1158/1078-0432.CCR-04-2001.
46. Chen Y, Yao Y, Zhang L, Li X, Wang Y, Zhao L, Wang J, Wang G, Shen D, Wei L, Zhao J. cDNA microarray analysis and immunohistochemistry reveal a distinct molecular phenotype in serous endometrial cancer compared to endometrioid endometrial cancer. Exp Mol Pathol. 2011;91(1):373–84. https://doi.org/10.1016/j.yexmp.2011.04.005.
47. Ratner ES, Tuck D, Richter C, Nallur S, Patel RM, Schultz V, Hui P, Schwartz PE, Rutherford TJ, Weidhaas JB. MicroRNA signatures differentiate uterine cancer tumor subtypes. Gynecol Oncol. 2010;118(3):251–7. https://doi.org/10.1016/j.ygyno.2010.05.010.
48. Moreno-Bueno G, Sanchez-Estevez C, Cassia R, Rodriguez-Perales S, Diaz-Uriarte R, Dominguez O, Hardisson D, Andujar M, Prat J, Matias-Guiu X, Cigudosa JC, Palacios J. Differential gene expression profile in endometrioid and nonendometrioid endometrial carcinoma: STK15 is frequently overexpressed and amplified in nonendometrioid carcinomas. Cancer Res. 2003;63(18):5697–702.
49. Mhawech-Fauceglia P, Wang D, Kesterson J, Syriac S, Clark K, Frederick PJ, Lele S, Liu S. Gene expression profiles in stage I uterine serous carcinoma in comparison to grade 3 and grade 1 stage I endometrioid adenocarcinoma. PLoS One. 2011;6(3):e18066. https://doi.org/10.1371/journal.pone.0018066.
50. Catasus L, D'Angelo E, Pons C, Espinosa I, Prat J. Expression profiling of 22 genes involved in the PI3K-AKT pathway identifies two subgroups of high-grade endometrial carcinomas with different molecular alterations. Mod Pathol. 2010;23(5):694–702. https://doi.org/10.1038/modpathol.2010.44.
51. Ferguson SE, Olshen AB, Viale A, Barakat RR, Boyd J. Stratification of intermediate-risk endometrial cancer patients into groups at high risk or low risk for recurrence based on tumor gene expression profiles. Clin Cancer Res. 2005;11(6):2252–7. https://doi.org/10.1158/1078-0432.CCR-04-1353.
52. Salvesen HB, Carter SL, Mannelqvist M, Dutt A, Getz G, Stefansson IM, Raeder MB, Sos ML, Engelsen IB, Trovik J, Wik E, Greulich H, Bo TH, Jonassen I, Thomas RK, Zander T, Garraway LA, Oyan AM, Sellers WR, Kalland KH, Meyerson M, Akslen LA, Beroukhim R. Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. Proc Natl Acad Sci U S A. 2009;106(12):4834–9. https://doi.org/10.1073/pnas.0806514106.
53. O'Mara TA, Zhao M, Spurdle AB. Meta-analysis of gene expression studies in endometrial cancer identifies gene expression pro-

files associated with aggressive disease and patient outcome. Sci Rep. 2016;6:36677. https://doi.org/10.1038/srep36677.
54. Fles R, Hoogendoorn WE, Platteel I, Scheerman CE, de Leeuw-Mantel G, Mourits MJ, Hollema H, van Leeuwen FE, van Boven HH, Nederlof PM. Genomic profile of endometrial tumors depends on morphological subtype, not on tamoxifen exposure. Genes Chromosomes Cancer. 2010;49(8):699–710. https://doi.org/10.1002/gcc.20781.
55. Murayama-Hosokawa S, Oda K, Nakagawa S, Ishikawa S, Yamamoto S, Shoji K, Ikeda Y, Uehara Y, Fukayama M, McCormick F, Yano T, Taketani Y, Aburatani H. Genome-wide single-nucleotide polymorphism arrays in endometrial carcinomas associate extensive chromosomal instability with poor prognosis and unveil frequent chromosomal imbalances involved in the PI3-kinase pathway. Oncogene. 2010;29(13):1897–908. https://doi.org/10.1038/onc.2009.474.
56. Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, Price JC, Zhang S, England BM, Godwin AK, Sgroi DC, Hieter P, Mullikin JC, Merino MJ, Bell DW. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nat Genet. 2012;44(12):1310–5. https://doi.org/10.1038/ng.2455.
57. Kuhn E, Wu RC, Guan B, Wu G, Zhang J, Wang Y, Song L, Yuan X, Wei L, Roden RB, Kuo KT, Nakayama K, Clarke B, Shaw P, Olvera N, Kurman RJ, Levine DA, Wang TL, Shih Ie M. Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. J Natl Cancer Inst. 2012;104(19):1503–13. https://doi.org/10.1093/jnci/djs345.
58. Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, Guzzo F, English DP, Varughese J, Gasparrini S, Bortolomai I, Buza N, Hui P, Abu-Khalaf M, Ravaggi A, Bignotti E, Bandiera E, Romani C, Todeschini P, Tassi R, Zanotti L, Carrara L, Pecorelli S, Silasi DA, Ratner E, Azodi M, Schwartz PE, Rutherford TJ, Stiegler AL, Mane S, Boggon TJ, Schlessinger J, Lifton RP, Santin AD. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. Proc Natl Acad Sci U S A. 2013;110(8):2916–21. https://doi.org/10.1073/pnas.1222577110.
59. The Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67–73. https://doi.org/10.1038/nature12113.
60. Church DN, Briggs SE, Palles C, Domingo E, Kearsey SJ, Grimes JM, Gorman M, Martin L, Howarth KM, Hodgson SV, Kaur K, Taylor J, Tomlinson IP. DNA polymerase {varepsilon} and delta exonuclease domain mutations in endometrial cancer. Hum Mol Genet. 2013; https://doi.org/10.1093/hmg/ddt131.
61. Church DN, Stelloo E, Nout RA, Valtcheva N, Depreeuw J, ter Haar N, Noske A, Amant F, Tomlinson IP, Wild PJ, Lambrechts D, Jurgenliemk-Schulz IM, Jobsen JJ, Smit VT, Creutzberg CL, Bosse T. Prognostic significance of POLE proofreading mutations in endometrial cancer. J Natl Cancer Inst. 2015;107(1):402. https://doi.org/10.1093/jnci/dju402.
62. Billingsley CC, Cohn DE, Mutch DG, Hade EM, Goodfellow PJ. Prognostic significance of POLE exonuclease domain mutations in high-grade endometrioid endometrial cancer on survival and recurrence: a subanalysis. Int J Gynecol Cancer. 2016;26(5):933–8. https://doi.org/10.1097/IGC.0000000000000681.
63. Billingsley CC, Cohn DE, Mutch DG, Stephens JA, Suarez AA, Goodfellow PJ. Polymerase varepsilon (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. Cancer. 2015;121(3):386–94. https://doi.org/10.1002/cncr.29046.
64. McConechy MK, Talhouk A, Leung S, Chiu D, Yang W, Senz J, Reha-Krantz LJ, Lee CH, Huntsman DG, Gilks CB, McAlpine JN. Endometrial carcinomas with POLE exonuclease domain mutations have a favorable prognosis. Clin Cancer Res. 2016;22(12):2865–73. https://doi.org/10.1158/1078-0432.CCR-15-2233.
65. Meng B, Hoang LN, McIntyre JB, Duggan MA, Nelson GS, Lee CH, Kobel M. POLE exonuclease domain mutation predicts long progression-free survival in grade 3 endometrioid carcinoma of the endometrium. Gynecol Oncol. 2014;134(1):15–9. https://doi.org/10.1016/j.ygyno.2014.05.006.
66. Talhouk A, McConechy MK, Leung S, Li-Chang HH, Kwon JS, Melnyk N, Yang W, Senz J, Boyd N, Karnezis AN, Huntsman DG, Gilks CB, McAlpine JN. A clinically applicable molecular-based classification for endometrial cancers. Br J Cancer. 2015;113(2):299–310. https://doi.org/10.1038/bjc.2015.190.
67. Kommoss S, McConechy MK, Kommoss F, Leung S, Bunz A, Magrill J, Britton H, Kommoss F, Grevenkamp F, Karnezis A, Yang W, Lum A, Kramer B, Taran F, Staebler A, Lax S, Brucker SY, Huntsman DG, Gilks CB, McAlpine JN, Talhouk A. Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series. Ann Oncol. 2018; https://doi.org/10.1093/annonc/mdy058.
68. Talhouk A, McConechy MK, Leung S, Yang W, Lum A, Senz J, Boyd N, Pike J, Anglesio M, Kwon JS, Karnezis AN, Huntsman DG, Gilks CB, McAlpine JN. Confirmation of ProMisE: a simple, genomics-based clinical classifier for endometrial cancer. Cancer. 2017;123(5):802–13. https://doi.org/10.1002/cncr.30496.
69. Bosse T, Nout RA, McAlpine JN, McConechy MK, Britton H, Hussein YR, Gonzalez C, Ganesan R, Steele JC, Harrison BT, Oliva E, Vidal A, Matias-Guiu X, Abu-Rustum NR, Levine DA, Gilks CB, Soslow RA. Molecular classification of grade 3 endometrioid endometrial cancers identifies distinct prognostic subgroups. Am J Surg Pathol. 2018;42(5):561–8. https://doi.org/10.1097/PAS.0000000000001020.
70. Kim HJ, Kim TJ, Lee YY, Choi CH, Lee JW, Bae DS, Kim BG. A comparison of uterine papillary serous, clear cell carcinomas, and grade 3 endometrioid corpus cancers using 2009 FIGO staging system. J Gynecol Oncol. 2013;24(2):120–7. https://doi.org/10.3802/jgo.2013.24.2.120.
71. Wang J, Jia N, Li Q, Wang C, Tao X, Hua K, Feng W. Analysis of recurrence and survival rates in grade 3 endometrioid endometrial carcinoma. Oncol Lett. 2016;12(4):2860–7. https://doi.org/10.3892/ol.2016.4918.
72. Espinosa I, Lee CH, D'Angelo E, Palacios J, Prat J. Undifferentiated and dedifferentiated endometrial carcinomas with POLE exonuclease domain mutations have a favorable prognosis. Am J Surg Pathol. 2017;41(8):1121–8. https://doi.org/10.1097/PAS.0000000000000873.
73. DeLair DF, Burke KA, Selenica P, Lim RS, Scott SN, Middha S, Mohanty AS, Cheng DT, Berger MF, Soslow RA, Weigelt B. The genetic landscape of endometrial clear cell carcinomas. J Pathol. 2017;243(2):230–41. https://doi.org/10.1002/path.4947.
74. Rosa-Rosa JM, Leskela S, Cristobal-Lana E, Santon A, Lopez-Garcia MA, Munoz G, Perez-Mies B, Biscuola M, Prat J, Esther OE, Soslow RA, Matias-Guiu X, Palacios J. Molecular genetic heterogeneity in undifferentiated endometrial carcinomas. Mod Pathol. 2016;29(11):1390–8. https://doi.org/10.1038/modpathol.2016.132.
75. Le Gallo M, Rudd ML, Urick ME, Hansen NF, Zhang S, Program NCS, Lozy F, Sgroi DC, Vidal Bel A, Matias-Guiu X, Broaddus RR, Lu KH, Levine DA, Mutch DG, Goodfellow PJ, Salvesen HB, Mullikin JC, Bell DW. Somatic mutation profiles of clear cell endometrial tumors revealed by whole exome and targeted

gene sequencing. Cancer. 2017;123(17):3261–8. https://doi.org/10.1002/cncr.30745.
76. Jones S, Stransky N, McCord CL, Cerami E, Lagowski J, Kelly D, Angiuoli SV, Sausen M, Kann L, Shukla M, Makar R, Wood LD, Diaz LA Jr, Lengauer C, Velculescu VE. Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes. Nat Commun. 2014;5:5006. https://doi.org/10.1038/ncomms6006.
77. Cherniack AD, Shen H, Walter V, Stewart C, Murray BA, Bowlby R, Hu X, Ling S, Soslow RA, Broaddus RR, Zuna RE, Robertson G, Laird PW, Kucherlapati R, Mills GB, Cancer Genome Atlas Research Network, Weinstein JN, Zhang J, Akbani R, Levine DA. Integrated molecular characterization of uterine carcinosarcoma. Cancer Cell. 2017;31(3):411–23. https://doi.org/10.1016/j.ccell.2017.02.010.
78. Zhao S, Bellone S, Lopez S, Thakral D, Schwab C, English DP, Black J, Cocco E, Choi J, Zammataro L, Predolini F, Bonazzoli E, Bi M, Buza N, Hui P, Wong S, Abu-Khalaf M, Ravaggi A, Bignotti E, Bandiera E, Romani C, Todeschini P, Tassi R, Zanotti L, Odicino F, Pecorelli S, Donzelli C, Ardighieri L, Facchetti F, Falchetti M, Silasi DA, Ratner E, Azodi M, Schwartz PE, Mane S, Angioli R, Terranova C, Quick CM, Edraki B, Bilguvar K, Lee M, Choi M, Stiegler AL, Boggon TJ, Schlessinger J, Lifton RP, Santin AD. Mutational landscape of uterine and ovarian carcinosarcomas implicates histone genes in epithelial-mesenchymal transition. Proc Natl Acad Sci U S A. 2016;113(43):12238–43. https://doi.org/10.1073/pnas.1614120113.
79. Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, Stover E, Strickland KC, D'Andrea AD, Wu CJ, Matulonis UA, Konstantinopoulos PA. Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. JAMA Oncol. 2015;1(9):1319–23. https://doi.org/10.1001/jamaoncol.2015.2151.
80. Eggink FA, Van Gool IC, Leary A, Pollock PM, Crosbie EJ, Mileshkin L, Jordanova ES, Adam J, Freeman-Mills L, Church DN, Creutzberg CL, De Bruyn M, Nijman HW, Bosse T. Immunological profiling of molecularly classified high-risk endometrial cancers identifies POLE-mutant and microsatellite unstable carcinomas as candidates for checkpoint inhibition. Oncoimmunology. 2017;6(2):e1264565. https://doi.org/10.1080/2162402X.2016.1264565.
81. Ott PA, Bang YJ, Berton-Rigaud D, Elez E, Pishvaian MJ, Rugo HS, Puzanov I, Mehnert JM, Aung KL, Lopez J, Carrigan M, Saraf S, Chen M, Soria JC. Safety and antitumor activity of pembrolizumab in advanced programmed death ligand 1-positive endometrial cancer: results from the KEYNOTE-028 study. J Clin Oncol. 2017;35(22):2535–41. https://doi.org/10.1200/JCO.2017.72.5952.
82. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Greten TF, Duffy AG, Ciombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdhoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bever KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA, Diaz LA Jr. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357(6349):409–13. https://doi.org/10.1126/science.aan6733.
83. Lemery S, Keegan P, Pazdur R. First FDA approval agnostic of cancer site – when a biomarker defines the indication. N Engl J Med. 2017;377(15):1409–12. https://doi.org/10.1056/NEJMp1709968.
84. Tao JJ, Schram AM, Hyman DM. Basket studies: redefining clinical trials in the era of genome-driven oncology. Annu Rev Med. 2018;69:319–31. https://doi.org/10.1146/annurev-med-062016-050343.
85. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. Cell. 2017;168(4):584–99. https://doi.org/10.1016/j.cell.2016.12.015.
86. Fader AN, Roque DM, Siegel E, Buza N, Hui P, Abdelghany O, Chambers SK, Secord AA, Havrilesky L, O'Malley DM, Backes F, Nevadunsky N, Edraki B, Pikaart D, Lowery W, ElSahwi KS, Celano P, Bellone S, Azodi M, Litkouhi B, Ratner E, Silasi DA, Schwartz PE, Santin AD. Randomized phase II trial of carboplatin-paclitaxel versus carboplatin-paclitaxel-trastuzumab in uterine serous carcinomas that overexpress human epidermal growth factor receptor 2/neu. J Clin Oncol. 2018:JCO2017765966. https://doi.org/10.1200/JCO.2017.76.5966.
87. Hyman DM, Smyth LM, Donoghue MTA, Westin SN, Bedard PL, Dean EJ, Bando H, El-Khoueiry AB, Perez-Fidalgo JA, Mita A, Schellens JHM, Chang MT, Reichel JB, Bouvier N, Selcuklu SD, Soumerai TE, Torrisi J, Erinjeri JP, Ambrose H, Barrett JC, Dougherty B, Foxley A, Lindemann JPO, McEwen R, Pass M, Schiavon G, Berger MF, Chandarlapaty S, Solit DB, Banerji U, Baselga J, Taylor BS. AKT inhibition in solid tumors with AKT1 mutations. J Clin Oncol. 2017;35(20):2251–9. https://doi.org/10.1200/JCO.2017.73.0143.
88. Gupta S, Ramjaun AR, Haiko P, Wang Y, Warne PH, Nicke B, Nye E, Stamp G, Alitalo K, Downward J. Binding of ras to phosphoinositide 3-kinase p110alpha is required for ras-driven tumorigenesis in mice. Cell. 2007;129(5):957–68. https://doi.org/10.1016/j.cell.2007.03.051.
89. Oda K, Stokoe D, Taketani Y, McCormick F. High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma. Cancer Res. 2005;65(23):10669–73. https://doi.org/10.1158/0008-5472.CAN-05-2620.
90. Jaiswal BS, Janakiraman V, Kljavin NM, Chaudhuri S, Stern HM, Wang W, Kan Z, Dbouk HA, Peters BA, Waring P, Dela Vega T, Kenski DM, Bowman KK, Lorenzo M, Li H, Wu J, Modrusan Z, Stinson J, Eby M, Yue P, Kaminker JS, de Sauvage FJ, Backer JM, Seshagiri S. Somatic mutations in p85alpha promote tumorigenesis through class IA PI3K activation. Cancer Cell. 2009;16(6):463–74. https://doi.org/10.1016/j.ccr.2009.10.016.
91. Cheung LW, Yu S, Zhang D, Li J, Ng PK, Panupinthu N, Mitra S, Ju Z, Yu Q, Liang H, Hawke DH, Lu Y, Broaddus RR, Mills GB. Naturally occurring neomorphic PIK3R1 mutations activate the MAPK pathway, dictating therapeutic response to MAPK pathway inhibitors. Cancer Cell. 2014;26(4):479–94. https://doi.org/10.1016/j.ccell.2014.08.017.
92. Rudd ML, Price JC, Fogoros S, Godwin AK, Sgroi DC, Merino MJ, Bell DW. A unique spectrum of somatic PIK3CA (p110alpha) mutations within primary endometrial carcinomas. Clin Cancer Res. 2011;17(6):1331–40. https://doi.org/10.1158/1078-0432.CCR-10-0540.
93. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004;304(5670):554. https://doi.org/10.1126/science.1096502.
94. Slomovitz BM, Coleman RL. The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. Clin Cancer Res. 2012;18(21):5856–64. https://doi.org/10.1158/1078-0432.CCR-12-0662.
95. Tredan O, Treilleux I, Wang Q, Gane N, Pissaloux D, Bonnin N, Petit T, Cretin J, Bonichon-Lamichhane N, Priou F, Lavau-Denes S, Mari V, Freyer G, Lebrun D, Alexandre J, Ray-Coquard I. Predicting everolimus treatment efficacy in patients with advanced endometrial carcinoma: a GINECO group study. Target Oncol. 2012. https://doi.org/10.1007/s11523-012-0242-9.
96. Ray-Coquard I, Favier L, Weber B, Roemer-Becuwe C, Bougnoux P, Fabbro M, Floquet A, Joly F, Plantade A, Paraiso D, Pujade-Lauraine E. Everolimus as second- or third-line treatment of advanced endometrial cancer: ENDORAD, a phase II

trial of GINECO. Br J Cancer. 2013;108(9):1771–7. https://doi.org/10.1038/bjc.2013.183.
97. Colombo N, McMeekin DS, Schwartz PE, Sessa C, Gehrig PA, Holloway R, Braly P, Matei D, Morosky A, Dodion PF, Einstein MH, Haluska F. Ridaforolimus as a single agent in advanced endometrial cancer: results of a single-arm, phase 2 trial. Br J Cancer. 2013;108(5):1021–6. https://doi.org/10.1038/bjc.2013.59.
98. Weigelt B, Warne PH, Lambros MB, Reis-Filho JS, Downward J. PI3K pathway dependencies in endometrioid endometrial cancer cell lines. Clin Cancer Res. 2013; https://doi.org/10.1158/1078-0432.CCR-12-3815.
99. Hampel H, Frankel W, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, La Jeunesse J, Nakagawa H, Westman JA, Prior TW, Clendenning M, Penzone P, Lombardi J, Dunn P, Cohn DE, Copeland L, Eaton L, Fowler J, Lewandowski G, Vaccarello L, Bell J, Reid G, de la Chapelle A. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. Cancer Res. 2006;66(15):7810–7. https://doi.org/10.1158/0008-5472.CAN-06-1114.
100. Ollikainen M, Abdel-Rahman WM, Moisio AL, Lindroos A, Kariola R, Jarvela I, Poyhonen M, Butzow R, Peltomaki P. Molecular analysis of familial endometrial carcinoma: a manifestation of hereditary nonpolyposis colorectal cancer or a separate syndrome? J Clin Oncol. 2005;23(21):4609–16. https://doi.org/10.1200/JCO.2005.06.055.
101. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomaki P, Mecklin JP, Jarvinen HJ. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer. 1999;81(2):214–8.
102. Meyer LA, Broaddus RR, Lu KH. Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. Cancer Control. 2009;16(1):14–22.
103. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the international collaborative group on HNPCC. Gastroenterology. 1999;116(6):1453–6.
104. Lu KH, Dinh M, Kohlmann W, Watson P, Green J, Syngal S, Bandipalliam P, Chen LM, Allen B, Conrad P, Terdiman J, Sun C, Daniels M, Burke T, Gershenson DM, Lynch H, Lynch P, Broaddus RR. Gynecologic cancer as a "sentinel cancer" for women with hereditary nonpolyposis colorectal cancer syndrome. Obstet Gynecol. 2005;105(3):569–74. https://doi.org/10.1097/01.AOG.0000154885.44002.ae.
105. Win AK, Lindor NM, Winship I, Tucker KM, Buchanan DD, Young JP, Rosty C, Leggett B, Giles GG, Goldblatt J, Macrae FA, Parry S, Kalady MF, Baron JA, Ahnen DJ, Marchand LL, Gallinger S, Haile RW, Newcomb PA, Hopper JL, Jenkins MA. Risks of colorectal and other cancers after endometrial cancer for women with Lynch syndrome. J Natl Cancer Inst. 2013;105(4):274–9. https://doi.org/10.1093/jnci/djs525.
106. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004;96(4):261–8.
107. Folkins AK, Longacre TA. Hereditary gynaecological malignancies: advances in screening and treatment. Histopathology. 2013;62(1):2–30. https://doi.org/10.1111/his.12028.
108. Garg K, Soslow RA. Lynch syndrome (hereditary non-polyposis colorectal cancer) and endometrial carcinoma. J Clin Pathol. 2009;62(8):679–84. https://doi.org/10.1136/jcp.2009.064949.
109. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A national cancer institute workshop on microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 1998;58(22):5248–57.
110. Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, Seruca R, Iacopetta B, Hamelin R. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology. 2002;123(6):1804–11. https://doi.org/10.1053/gast.2002.37070.
111. Wong YF, Cheung TH, Lo KW, Yim SF, Chan LK, Buhard O, Duval A, Chung TK, Hamelin R. Detection of microsatellite instability in endometrial cancer: advantages of a panel of five mononucleotide repeats over the national cancer institute panel of markers. Carcinogenesis. 2006;27(5):951–5. https://doi.org/10.1093/carcin/bgi333.
112. Wijnen J, de Leeuw W, Vasen H, van der Klift H, Moller P, Stormorken A, Meijers-Heijboer H, Lindhout D, Menko F, Vossen S, Moslein G, Tops C, Brocker-Vriends A, Wu Y, Hofstra R, Sijmons R, Cornelisse C, Morreau H, Fodde R. Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat Genet. 1999;23(2):142–4. https://doi.org/10.1038/13773.
113. Hendriks YM, Wagner A, Morreau H, Menko F, Stormorken A, Quehenberger F, Sandkuijl L, Moller P, Genuardi M, Van Houwelingen H, Tops C, Van Puijenbroek M, Verkuijlen P, Kenter G, Van Mil A, Meijers-Heijboer H, Tan GB, Breuning MH, Fodde R, Wijnen JT, Brocker-Vriends AH, Vasen H. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. Gastroenterology. 2004;127(1):17–25.
114. Wu Y, Berends MJ, Mensink RG, Kempinga C, Sijmons RH, van Der Zee AG, Hollema H, Kleibeuker JH, Buys CH, Hofstra RM. Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. Am J Hum Genet. 1999;65(5):1291–8. https://doi.org/10.1086/302612.
115. Modica I, Soslow RA, Black D, Tornos C, Kauff N, Shia J. Utility of immunohistochemistry in predicting microsatellite instability in endometrial carcinoma. Am J Surg Pathol. 2007;31(5):744–51. https://doi.org/10.1097/01.pas.0000213428.61374.06.
116. Goodfellow PJ, Billingsley CC, Lankes HA, Ali S, Cohn DE, Broaddus RJ, Ramirez N, Pritchard CC, Hampel H, Chassen AS, Simmons LV, Schmidt AP, Gao F, Brinton LA, Backes F, Landrum LM, Geller MA, DiSilvestro PA, Pearl ML, Lele SB, Powell MA, Zaino RJ, Mutch D. Combined microsatellite instability, MLH1 methylation analysis, and immunohistochemistry for Lynch syndrome screening in endometrial cancers from GOG210: an NRG oncology and gynecologic oncology group study. J Clin Oncol. 2015;33(36):4301–8. https://doi.org/10.1200/JCO.2015.63.9518.
117. Buchanan DD, Tan YY, Walsh MD, Clendenning M, Metcalf AM, Ferguson K, Arnold ST, Thompson BA, Lose FA, Parsons MT, Walters RJ, Pearson SA, Cummings M, Oehler MK, Blomfield PB, Quinn MA, Kirk JA, Stewart CJ, Obermair A, Young JP, Webb PM, Spurdle AB. Tumor mismatch repair immunohistochemistry and DNA MLH1 methylation testing of patients with endometrial cancer diagnosed at age younger than 60 years optimizes triage for population-level germline mismatch repair gene mutation testing. J Clin Oncol. 2014;32(2):90–100. https://doi.org/10.1200/JCO.2013.51.2129.
118. Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME. Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? J Clin Oncol. 2013;31(10):1267–70. https://doi.org/10.1200/JCO.2012.46.9403.
119. Zimmermann A, Bernuit D, Gerlinger C, Schaefers M, Geppert K. Prevalence, symptoms and management of uterine fibroids: an international internet-based survey of 21,746 women. BMC Womens Health. 2012;12:6. https://doi.org/10.1186/1472-6874-12-6.

120. Cramer SF, Patel A. The frequency of uterine leiomyomas. Am J Clin Pathol. 1990;94(4):435–8.
121. Perot G, Croce S, Ribeiro A, Lagarde P, Velasco V, Neuville A, Coindre JM, Stoeckle E, Floquet A, MacGrogan G, Chibon F. MED12 alterations in both human benign and malignant uterine soft tissue tumors. PLoS One. 2012;7(6):e40015. https://doi.org/10.1371/journal.pone.0040015.
122. Sandberg AA. Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors: leiomyosarcoma. Cancer Genet Cytogenet. 2005;161(1):1–19.
123. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA, editors. SEER cancer statistics review, 1975–2010. Bethesda: National Cancer Institute; 2013.
124. Bell SW, Kempson RL, Hendrickson MR. Problematic uterine smooth muscle neoplasms. a clinicopathologic study of 213 cases. Am J Surg Pathol. 1994;18(6):535–58.
125. Veras E, Malpica A, Deavers MT, Silva EG. Mitosis-specific marker phospho-histone H3 in the assessment of mitotic index in uterine smooth muscle tumors: a pilot study. Int J Gynecol Pathol. 2009;28(4):316–21. https://doi.org/10.1097/PGP.0b013e318193df97.
126. Lim D, Alvarez T, Nucci MR, Gilks B, Longacre T, Soslow RA, Oliva E. Interobserver variability in the interpretation of tumor cell necrosis in uterine leiomyosarcoma. Am J Surg Pathol. 2013;37(5):650–8. https://doi.org/10.1097/PAS.0b013e3182851162.
127. Soper JT, KS MC Jr, Hinshaw W, Creasman WT, KS MC Sr, Clarke-Pearson DL. Cytoplasmic estrogen and progesterone receptor content of uterine sarcomas. Am J Obstet Gynecol. 1984;150(4):342–8.
128. Sutton GP, Stehman FB, Michael H, Young PC, Ehrlich CE. Estrogen and progesterone receptors in uterine sarcomas. Obstet Gynecol. 1986;68(5):709–14.
129. Wade K, Quinn MA, Hammond I, Williams K, Cauchi M. Uterine sarcoma: steroid receptors and response to hormonal therapy. Gynecol Oncol. 1990;39(3):364–7.
130. Zhai YL, Kobayashi Y, Mori A, Orii A, Nikaido T, Konishi I, Fujii S. Expression of steroid receptors, Ki-67, and p53 in uterine leiomyosarcomas. Int J Gynecol Pathol. 1999;18(1):20–8.
131. Mittal K, Demopoulos RI. MIB-1 (Ki-67), p53, estrogen receptor, and progesterone receptor expression in uterine smooth muscle tumors. Hum Pathol. 2001;32(9):984–7. https://doi.org/10.1053/hupa.2001.27113.
132. Bodner K, Bodner-Adler B, Kimberger O, Czerwenka K, Leodolter S, Mayerhofer K. Estrogen and progesterone receptor expression in patients with uterine leiomyosarcoma and correlation with different clinicopathological parameters. Anticancer Res. 2003;23(1B):729–32.
133. Leitao MM, Soslow RA, Nonaka D, Olshen AB, Aghajanian C, Sabbatini P, Dupont J, Hensley M, Sonoda Y, Barakat RR, Anderson S. Tissue microarray immunohistochemical expression of estrogen, progesterone, and androgen receptors in uterine leiomyomata and leiomyosarcoma. Cancer. 2004;101(6):1455–62. https://doi.org/10.1002/cncr.20521.
134. Jeffers MD, Farquharson MA, Richmond JA, McNicol AM. p53 immunoreactivity and mutation of the p53 gene in smooth muscle tumours of the uterine corpus. J Pathol. 1995;177(1):65–70. https://doi.org/10.1002/path.1711770111.
135. Bodner-Adler B, Bodner K, Czerwenka K, Kimberger O, Leodolter S, Mayerhofer K. Expression of p16 protein in patients with uterine smooth muscle tumors: an immunohistochemical analysis. Gynecol Oncol. 2005;96(1):62–6. https://doi.org/10.1016/j.ygyno.2004.09.026.
136. Akhan SE, Yavuz E, Tecer A, Iyibozkurt CA, Topuz S, Tuzlali S, Bengisu E, Berkman S. The expression of Ki-67, p53, estrogen and progesterone receptors affecting survival in uterine leiomyosarcomas. a clinicopathologic study. Gynecol Oncol. 2005;99(1):36–42. https://doi.org/10.1016/j.ygyno.2005.05.019.
137. O'Neill CJ, McBride HA, Connolly LE, McCluggage WG. Uterine leiomyosarcomas are characterized by high p16, p53 and MIB1 expression in comparison with usual leiomyomas, leiomyoma variants and smooth muscle tumours of uncertain malignant potential. Histopathology. 2007;50(7):851–8. https://doi.org/10.1111/j.1365-2559.2007.02699.x.
138. Atkins KA, Arronte N, Darus CJ, Rice LW. The Use of p16 in enhancing the histologic classification of uterine smooth muscle tumors. Am J Surg Pathol. 2008;32(1):98–102. https://doi.org/10.1097/PAS.0b013e3181574d1e.
139. Gannon BR, Manduch M, Childs TJ. Differential immunoreactivity of p16 in leiomyosarcomas and leiomyoma variants. Int J Gynecol Pathol. 2008;27(1):68–73. https://doi.org/10.1097/pgp.0b013e3180ca954f.
140. Chen L, Yang B. Immunohistochemical analysis of p16, p53, and Ki-67 expression in uterine smooth muscle tumors. Int J Gynecol Pathol. 2008;27(3):326–32. https://doi.org/10.1097/PGP.0b013e31815ea7f5.
141. Lee CH, Turbin DA, Sung YC, Espinosa I, Montgomery K, van de Rijn M, Gilks CB. A panel of antibodies to determine site of origin and malignancy in smooth muscle tumors. Mod Pathol. 2009;22(12):1519–31. https://doi.org/10.1038/modpathol.2009.122.
142. Ip PP, Cheung AN, Clement PB. Uterine smooth muscle tumors of uncertain malignant potential (STUMP): a clinicopathologic analysis of 16 cases. Am J Surg Pathol. 2009;33(7):992–1005. https://doi.org/10.1097/PAS.0b013e3181a02d1c.
143. D'Angelo E, Espinosa I, Ali R, Gilks CB, Rijn M, Lee CH, Prat J. Uterine leiomyosarcomas: tumor size, mitotic index, and biomarkers Ki67, and Bcl-2 identify two groups with different prognosis. Gynecol Oncol. 2011;121(2):328–33. https://doi.org/10.1016/j.ygyno.2011.01.022.
144. Hakverdi S, Gungoren A, Yaldiz M, Hakverdi AU, Toprak S. Immunohistochemical analysis of p16 expression in uterine smooth muscle tumors. Eur J Gynaecol Oncol. 2011;32(5):513–5.
145. Makinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ, Gentile M, Yan J, Enge M, Taipale M, Aavikko M, Katainen R, Virolainen E, Bohling T, Koski TA, Launonen V, Sjoberg J, Taipale J, Vahteristo P, Aaltonen LA. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. Science. 2011;334(6053):252–5. https://doi.org/10.1126/science.1208930.
146. Makinen N, Heinonen HR, Moore S, Tomlinson IP, van der Spuy ZM, Aaltonen LA. MED12 exon 2 mutations are common in uterine leiomyomas from South African patients. Oncotarget. 2011;2(12):966–9.
147. Je EM, Kim MR, Min KO, Yoo NJ, Lee SH. Mutational analysis of MED12 exon 2 in uterine leiomyoma and other common tumors. Int J Cancer. 2012;131(6):E1044–7. https://doi.org/10.1002/ijc.27610.
148. McGuire MM, Yatsenko A, Hoffner L, Jones M, Surti U, Rajkovic A. Whole exome sequencing in a random sample of North American women with leiomyomas identifies MED12 mutations in majority of uterine leiomyomas. PLoS One. 2012;7(3):e33251. https://doi.org/10.1371/journal.pone.0033251.
149. Markowski DN, Huhle S, Nimzyk R, Stenman G, Loning T, Bullerdiek J. MED12 mutations occurring in benign and malignant mammalian smooth muscle tumors. Genes Chromosomes Cancer. 2013;52(3):297–304. https://doi.org/10.1002/gcc.22029.
150. Matsubara A, Sekine S, Yoshida M, Yoshida A, Taniguchi H, Kushima R, Tsuda H, Kanai Y. Prevalence of MED12 mutations in uterine and extrauterine smooth muscle tumours. Histopathology. 2013;62(4):657–61. https://doi.org/10.1111/his.12039.

151. de Graaff MA, Cleton-Jansen AM, Szuhai K, Bovee JV. Mediator complex subunit 12 exon 2 mutation analysis in different subtypes of smooth muscle tumors confirms genetic heterogeneity. Hum Pathol. 2013;44(8):1597–604. https://doi.org/10.1016/j.humpath.2013.01.006.
152. Conaway RC, Conaway JW. Function and regulation of the mediator complex. Curr Opin Genet Dev. 2011;21(2):225–30. https://doi.org/10.1016/j.gde.2011.01.013.
153. Kim S, Xu X, Hecht A, Boyer TG. Mediator is a transducer of Wnt/beta-catenin signaling. J Biol Chem. 2006;281(20):14066–75. https://doi.org/10.1074/jbc.M602696200.
154. Zhou H, Kim S, Ishii S, Boyer TG. Mediator modulates Gli3-dependent Sonic hedgehog signaling. Mol Cell Biol. 2006;26(23):8667–82. https://doi.org/10.1128/MCB.00443-06.
155. Tutter AV, Kowalski MP, Baltus GA, Iourgenko V, Labow M, Li E, Kadam S. Role for Med12 in regulation of Nanog and Nanog target genes. J Biol Chem. 2009;284(6):3709–18. https://doi.org/10.1074/jbc.M805677200.
156. Makinen N, Vahteristo P, Kampjarvi K, Arola J, Butzow R, Aaltonen LA. MED12 exon 2 mutations in histopathological uterine leiomyoma variants. Eur J Hum Genet. 2013; https://doi.org/10.1038/ejhg.2013.33.
157. Sandberg AA. Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors: leiomyoma. Cancer Genet Cytogenet. 2005;158(1):1–26. https://doi.org/10.1016/j.cancergencyto.2004.08.025.
158. Mehine M, Kaasinen E, Makinen N, Katainen R, Kampjarvi K, Pitkanen E, Heinonen HR, Butzow R, Kilpivaara O, Kuosmanen A, Ristolainen H, Gentile M, Sjoberg J, Vahteristo P, Aaltonen LA. Characterization of uterine leiomyomas by whole-genome sequencing. N Engl J Med. 2013; https://doi.org/10.1056/NEJMoa1302736.
159. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144(1):27–40. https://doi.org/10.1016/j.cell.2010.11.055.
160. Lehtonen R, Kiuru M, Vanharanta S, Sjoberg J, Aaltonen LM, Aittomaki K, Arola J, Butzow R, Eng C, Husgafvel-Pursiainen K, Isola J, Jarvinen H, Koivisto P, Mecklin JP, Peltomaki P, Salovaara R, Wasenius VM, Karhu A, Launonen V, Nupponen NN, Aaltonen LA. Biallelic inactivation of fumarate hydratase (FH) occurs in nonsyndromic uterine leiomyomas but is rare in other tumors. Am J Pathol. 2004;164(1):17–22. https://doi.org/10.1016/S0002-9440(10)63091-X.
161. Nagai R, Brock JW, Blatnik M, Baatz JE, Bethard J, Walla MD, Thorpe SR, Baynes JW, Frizzell N. Succination of protein thiols during adipocyte maturation: a biomarker of mitochondrial stress. J Biol Chem. 2007;282(47):34219–28. https://doi.org/10.1074/jbc.M703551200.
162. Bennett JA, Weigelt B, Chiang S, Selenica P, Chen YB, Bialik A, Bi R, Schultheis AM, Lim RS, Ng CKY, Morales-Oyarvide V, Young RH, Reuter VE, Soslow RA, Oliva E. Leiomyoma with bizarre nuclei: a morphological, immunohistochemical and molecular analysis of 31 cases. Mod Pathol. 2017;30(10):1476–88. https://doi.org/10.1038/modpathol.2017.56.
163. Harrison WJ, Andrici J, Maclean F, Madadi-Ghahan R, Farzin M, Sioson L, Toon CW, Clarkson A, Watson N, Pickett J, Field M, Crook A, Tucker K, Goodwin A, Anderson L, Srinivasan B, Grossmann P, Martinek P, Ondic O, Hes O, Trpkov K, Clifton-Bligh RJ, Dwight T, Gill AJ. Fumarate hydratase-deficient uterine leiomyomas occur in both the syndromic and sporadic settings. Am J Surg Pathol. 2016;40(5):599–607. https://doi.org/10.1097/PAS.0000000000000573.
164. Joseph NM, Solomon DA, Frizzell N, Rabban JT, Zaloudek C, Garg K. Morphology and immunohistochemistry for 2SC and FH Aid in detection of fumarate hydratase gene aberrations in uterine leiomyomas from young patients. Am J Surg Pathol. 2015;39(11):1529–39. https://doi.org/10.1097/PAS.0000000000000520.
165. Chudasama P, Mughal SS, Sanders MA, Hubschmann D, Chung I, Deeg KI, Wong SH, Rabe S, Hlevnjak M, Zapatka M, Ernst A, Kleinheinz K, Schlesner M, Sieverling L, Klink B, Schrock E, Hoogenboezem RM, Kasper B, Heilig CE, Egerer G, Wolf S, von Kalle C, Eils R, Stenzinger A, Weichert W, Glimm H, Groschel S, Kopp HG, Omlor G, Lehner B, Bauer S, Schimmack S, Ulrich A, Mechtersheimer G, Rippe K, Brors B, Hutter B, Renner M, Hohenberger P, Scholl C, Frohling S. Integrative genomic and transcriptomic analysis of leiomyosarcoma. Nat Commun. 2018;9(1):144. https://doi.org/10.1038/s41467-017-02602-0.
166. Hodge JC, Morton CC. Genetic heterogeneity among uterine leiomyomata: insights into malignant progression. Hum Mol Genet. 2007;16:R7–13. https://doi.org/10.1093/hmg/ddm043.
167. Kuhn E, Yemelyanova A, Wang TL, Kurman R, Shih Ie M Abstract 5536: TP53 and MED12 mutations in uterine smooth muscle tumors. In: AACR 103rd annual meeting, Chicago, IL, USA. vol 8, Supplement 1. Cancer Research. 2012.
168. de Vos S, Wilczynski SP, Fleischhacker M, Koeffler P. p53 alterations in uterine leiomyosarcomas versus leiomyomas. Gynecol Oncol. 1994;54(2):205–8. https://doi.org/10.1006/gyno.1994.1194.
169. Makinen N, Aavikko M, Heikkinen T, Taipale M, Taipale J, Koivisto-Korander R, Butzow R, Vahteristo P. Exome sequencing of uterine leiomyosarcomas identifies frequent mutations in TP53, ATRX, and MED12. PLoS Genet. 2016;12(2):e1005850. https://doi.org/10.1371/journal.pgen.1005850.
170. Network. CGAR. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. Cell. 2017;171(4):950–965.e928. https://doi.org/10.1016/j.cell.2017.10.014.
171. Lee PJ, Yoo NS, Hagemann IS, Pfeifer JD, Cottrell CE, Abel HJ, Duncavage EJ. Spectrum of mutations in leiomyosarcomas identified by clinical targeted next-generation sequencing. Exp Mol Pathol. 2017;102(1):156–61. https://doi.org/10.1016/j.yexmp.2017.01.012.
172. Chew I, Oliva E. Endometrial stromal sarcomas: a review of potential prognostic factors. Adv Anat Pathol. 2010;17(2):113–21. https://doi.org/10.1097/PAP.0b013e3181cfb7c2.
173. Nucci MR, O'Connell JT, Huettner PC, Cviko A, Sun D, Quade BJ. h-Caldesmon expression effectively distinguishes endometrial stromal tumors from uterine smooth muscle tumors. Am J Surg Pathol. 2001;25(4):455–63.
174. Oliva E, Young RH, Amin MB, Clement PB. An immunohistochemical analysis of endometrial stromal and smooth muscle tumors of the uterus: a study of 54 cases emphasizing the importance of using a panel because of overlap in immunoreactivity for individual antibodies. Am J Surg Pathol. 2002;26(4):403–12.
175. de Leval L, Waltregny D, Boniver J, Young RH, Castronovo V, Oliva E. Use of histone deacetylase 8 (HDAC8), a new marker of smooth muscle differentiation, in the classification of mesenchymal tumors of the uterus. Am J Surg Pathol. 2006;30(3):319–27. https://doi.org/10.1097/01.pas.0000188029.63706.31.
176. McCluggage WG, Date A, Bharucha H, Toner PG. Endometrial stromal sarcoma with sex cord-like areas and focal rhabdoid differentiation. Histopathology. 1996;29(4):369–74.
177. Fukunaga M, Miyazawa Y, Ushigome S. Endometrial low-grade stromal sarcoma with ovarian sex cord-like differentiation: report of two cases with an immunohistochemical and flow cytometric study. Pathol Int. 1997;47(6):412–5.

178. Zamecnik M, Michal M. Endometrial stromal nodule with retiform sex-cord-like differentiation. Pathol Res Pract. 1998;194(6):449–53.
179. Baker RJ, Hildebrandt RH, Rouse RV, Hendrickson MR, Longacre TA. Inhibin and CD99 (MIC2) expression in uterine stromal neoplasms with sex-cord-like elements. Hum Pathol. 1999;30(6):671–9.
180. Ohta Y, Suzuki T, Kojima M, Shiokawa A, Mitsuya T. Low-grade endometrial stromal sarcoma with an extensive epithelial-like element. Pathol Int. 2003;53(4):246–51.
181. Sumathi VP, Al-Hussaini M, Connolly LE, Fullerton L, McCluggage WG. Endometrial stromal neoplasms are immunoreactive with WT-1 antibody. Int J Gynecol Pathol. 2004;23(3):241–7.
182. Irving JA, Carinelli S, Prat J. Uterine tumors resembling ovarian sex cord tumors are polyphenotypic neoplasms with true sex cord differentiation. Mod Pathol. 2006;19(1):17–24. https://doi.org/10.1038/modpathol.3800475.
183. Reich O, Regauer S, Urdl W, Lahousen M, Winter R. Expression of oestrogen and progesterone receptors in low-grade endometrial stromal sarcomas. Br J Cancer. 2000;82(5):1030–4. https://doi.org/10.1054/bjoc.1999.1038.
184. Chu MC, Mor G, Lim C, Zheng W, Parkash V, Schwartz PE. Low-grade endometrial stromal sarcoma: hormonal aspects. Gynecol Oncol. 2003;90(1):170–6.
185. Balleine RL, Earls PJ, Webster LR, Mote PA, deFazio A, Harnett PR, Clarke CL. Expression of progesterone receptor A and B isoforms in low-grade endometrial stromal sarcoma. Int J Gynecol Pathol. 2004;23(2):138–44.
186. Kurihara S, Oda Y, Ohishi Y, Iwasa A, Takahira T, Kaneki E, Kobayashi H, Wake N, Tsuneyoshi M. Endometrial stromal sarcomas and related high-grade sarcomas: immunohistochemical and molecular genetic study of 31 cases. Am J Surg Pathol. 2008;32(8):1228–38. https://doi.org/10.1097/PAS.0b013e31816a3b42.
187. Lee CH, Marino-Enriquez A, Ou W, Zhu M, Ali RH, Chiang S, Amant F, Gilks CB, van de Rijn M, Oliva E, Debiec-Rychter M, Dal Cin P, Fletcher JA, Nucci MR. The clinicopathologic features of YWHAE-FAM22 endometrial stromal sarcomas: a histologically high-grade and clinically aggressive tumor. Am J Surg Pathol. 2012;36(5):641–53. https://doi.org/10.1097/PAS.0b013e31824a7b1a.
188. Jakate K, Azimi F, Ali RH, Lee CH, Clarke BA, Rasty G, Shaw PA, Melnyk N, Huntsman DG, Laframboise S, Rouzbahman M. Endometrial sarcomas: an immunohistochemical and JAZF1 re-arrangement study in low-grade and undifferentiated tumors. Mod Pathol. 2013;26(1):95–105. https://doi.org/10.1038/modpathol.2012.136.
189. Ng TL, Gown AM, Barry TS, Cheang MC, Chan AK, Turbin DA, Hsu FD, West RB, Nielsen TO. Nuclear beta-catenin in mesenchymal tumors. Mod Pathol. 2005;18(1):68–74. https://doi.org/10.1038/modpathol.3800272.
190. Jung CK, Jung JH, Lee A, Lee YS, Choi YJ, Yoon SK, Lee KY. Diagnostic use of nuclear beta-catenin expression for the assessment of endometrial stromal tumors. Mod Pathol. 2008;21(6):756–63. https://doi.org/10.1038/modpathol.2008.53.
191. Kildal W, Pradhan M, Abeler VM, Kristensen GB, Danielsen HE. Beta-catenin expression in uterine sarcomas and its relation to clinicopathological parameters. Eur J Cancer. 2009;45(13):2412–7. https://doi.org/10.1016/j.ejca.2009.06.017.
192. Kurihara S, Oda Y, Ohishi Y, Kaneki E, Kobayashi H, Wake N, Tsuneyoshi M. Coincident expression of beta-catenin and cyclin D1 in endometrial stromal tumors and related high-grade sarcomas. Mod Pathol. 2010;23(2):225–34. https://doi.org/10.1038/modpathol.2009.162.
193. Moinfar F, Gogg-Kamerer M, Sommersacher A, Regitnig P, Man YG, Zatloukal K, Denk H, Tavassoli FA. Endometrial stromal sarcomas frequently express epidermal growth factor receptor (EGFR, HER-1): potential basis for a new therapeutic approach. Am J Surg Pathol. 2005;29(4):485–9.
194. Cossu-Rocca P, Contini M, Uras MG, Muroni MR, Pili F, Carru C, Bosincu L, Massarelli G, Nogales FF, De Miglio MR. Tyrosine kinase receptor status in endometrial stromal sarcoma: an immunohistochemical and genetic-molecular analysis. Int J Gynecol Pathol. 2012;31(6):570–9. https://doi.org/10.1097/PGP.0b013e31824fe289.
195. Sardinha R, Hernandez T, Fraile S, Tresserra F, Vidal A, Gomez MC, Astudillo A, Hernandez N, Saenz de Santamaria J, Ordi J, Goncalves L, Ramos R, Balana C, de Alava E. Endometrial stromal tumors: immunohistochemical and molecular analysis of potential targets of tyrosine kinase inhibitors. Clin Sarcoma Res. 2013;3(1):3. https://doi.org/10.1186/2045-3329-3-3.
196. Cheng X, Yang G, Schmeler KM, Coleman RL, Tu X, Liu J, Kavanagh JJ. Recurrence patterns and prognosis of endometrial stromal sarcoma and the potential of tyrosine kinase-inhibiting therapy. Gynecol Oncol. 2011;121(2):323–7. https://doi.org/10.1016/j.ygyno.2010.12.360.
197. Moinfar F, Regitnig P, Tabrizi AD, Denk H, Tavassoli FA. Expression of androgen receptors in benign and malignant endometrial stromal neoplasms. Virchows Arch. 2004;444(5):410–4. https://doi.org/10.1007/s00428-004-0981-9.
198. Reich O, Regauer S. Aromatase expression in low-grade endometrial stromal sarcomas: an immunohistochemical study. Mod Pathol. 2004;17(1):104–8. https://doi.org/10.1038/sj.modpathol.3800031.
199. Liegl B, Reich O, Nogales FF, Regauer S. PDGF-alpha and PDGF-beta are expressed in endometrial stromal sarcoma: a potential therapeutic target for tyrosine kinase inhibitors? Histopathology. 2006;49(5):545–6. https://doi.org/10.1111/j.1365-2559.2006.02529.x.
200. Chiang S, Oliva E. Recent developments in uterine mesenchymal neoplasms. Histopathology. 2013;62(1):124–37. https://doi.org/10.1111/his.12048.
201. Micci F, Brunetti M, Dal Cin P, Nucci MR, Gorunova L, Heim S, Panagopoulos I. Fusion of the genes BRD8 and PHF1 in endometrial stromal sarcoma. Genes Chromosomes Cancer. 2017;56(12):841–5. https://doi.org/10.1002/gcc.22485.
202. Koontz JI, Soreng AL, Nucci M, Kuo FC, Pauwels P, van Den Berghe H, Dal Cin P, Fletcher JA, Sklar J. Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. Proc Natl Acad Sci U S A. 2001;98(11):6348–53. https://doi.org/10.1073/pnas.101132598.
203. Huang HY, Ladanyi M, Soslow RA. Molecular detection of JAZF1-JJAZ1 gene fusion in endometrial stromal neoplasms with classic and variant histology: evidence for genetic heterogeneity. Am J Surg Pathol. 2004;28(2):224–32.
204. Hrzenjak A, Moinfar F, Tavassoli FA, Strohmeier B, Kremser ML, Zatloukal K, Denk H. JAZF1/JJAZ1 gene fusion in endometrial stroemal sarcomas: molecular analysis by reverse transcriptase-polymerase chain reaction optimized for paraffin-embedded tissue. J Mol Diagn. 2005;7(3):388–95. https://doi.org/10.1016/S1525-1578(10)60568-5.
205. Oliva E, de Leval L, Soslow RA, Herens C. High frequency of JAZF1-JJAZ1 gene fusion in endometrial stromal tumors with smooth muscle differentiation by interphase FISH detection. Am J Surg Pathol. 2007;31(8):1277–84. https://doi.org/10.1097/PAS.0b013e318031f012.
206. Nucci MR, Harburger D, Koontz J, Dal Cin P, Sklar J. Molecular analysis of the JAZF1-JJAZ1 gene fusion by RT-PCR and fluorescence in situ hybridization in endometrial stromal neoplasms. Am J Surg Pathol. 2007;31(1):65–70. https://doi.org/10.1097/01.pas.0000213327.86992.d1.
207. Chiang S, Ali R, Melnyk N, McAlpine JN, Huntsman DG, Gilks CB, Lee CH, Oliva E. Frequency of known gene rearrangements in

endometrial stromal tumors. Am J Surg Pathol. 2011;35(9):1364–72. https://doi.org/10.1097/PAS.0b013e3182262743.
208. D'Angelo E, Ali RH, Espinosa I, Lee CH, Huntsman DG, Gilks B, Prat J. Endometrial stromal sarcomas with sex cord differentiation are associated with PHF1 rearrangement. Am J Surg Pathol. 2013;37(4):514–21. https://doi.org/10.1097/PAS.0b013e318272c612.
209. Micci F, Walter CU, Teixeira MR, Panagopoulos I, Bjerkehagen B, Saeter G, Heim S. Cytogenetic and molecular genetic analyses of endometrial stromal sarcoma: nonrandom involvement of chromosome arms 6p and 7p and confirmation of JAZF1/JJAZ1 gene fusion in t(7;17). Cancer Genet Cytogenet. 2003;144(2):119–24.
210. Panagopoulos I, Micci F, Thorsen J, Gorunova L, Eibak AM, Bjerkehagen B, Davidson B, Heim S. Novel fusion of MYST/Esa1-associated factor 6 and PHF1 in endometrial stromal sarcoma. PLoS One. 2012;7(6):e39354. https://doi.org/10.1371/journal.pone.0039354.
211. Panagopoulos I, Thorsen J, Gorunova L, Haugom L, Bjerkehagen B, Davidson B, Heim S, Micci F. Fusion of the ZC3H7B and BCOR genes in endometrial stromal sarcomas carrying an X;22-translocation. Genes Chromosomes Cancer. 2013;52(7):610–8. https://doi.org/10.1002/gcc.22057.
212. Dewaele B, Przybyl J, Quattrone A, Finalet Ferreiro J, Vanspauwen V, Geerdens E, Gianfelici V, Kalender Z, Wozniak A, Moerman P, Sciot R, Croce S, Amant F, Vandenberghe P, Cools J, Debiec-Rychter M. Identification of a novel, recurrent MBTD1-CXorf67 fusion in low-grade endometrial stromal sarcoma. Int J Cancer. 2014;134(5):1112–22. https://doi.org/10.1002/ijc.28440.
213. Li H, Ma X, Wang J, Koontz J, Nucci M, Sklar J. Effects of rearrangement and allelic exclusion of JJAZ1/SUZ12 on cell proliferation and survival. Proc Natl Acad Sci U S A. 2007;104(50):20001–6. https://doi.org/10.1073/pnas.0709986104.
214. Li H, Wang J, Mor G, Sklar J. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. Science. 2008;321(5894):1357–61. https://doi.org/10.1126/science.1156725.
215. Amant F, Tousseyn T, Coenegrachts L, Decloedt J, Moerman P, Debiec-Rychter M. Case report of a poorly differentiated uterine tumour with t(10;17) translocation and neuroectodermal phenotype. Anticancer Res. 2011;31(6):2367–71.
216. Lee CH, Ou WB, Marino-Enriquez A, Zhu M, Mayeda M, Wang Y, Guo X, Brunner AL, Amant F, French CA, West RB, McAlpine JN, Gilks CB, Yaffe MB, Prentice LM, McPherson A, Jones SJ, Marra MA, Shah SP, van de Rijn M, Huntsman DG, Dal Cin P, Debiec-Rychter M, Nucci MR, Fletcher JA. 14-3-3 fusion oncogenes in high-grade endometrial stromal sarcoma. Proc Natl Acad Sci U S A. 2012;109(3):929–34. https://doi.org/10.1073/pnas.1115528109.
217. Croce S, Hostein I, Ribeiro A, Garbay D, Velasco V, Stoeckle E, Guyon F, Floquet A, Neuville A, Coindre JM, Macgrogan G, Chibon F. YWHAE rearrangement identified by FISH and RT-PCR in endometrial stromal sarcomas: genetic and pathological correlations. Mod Pathol. 2013; https://doi.org/10.1038/modpathol.2013.69.
218. Attygalle AD, Vroobel K, Wren D, Barton DP, Hazell SJ, Cin PD, Koelble K, McCluggage WG. An unusual case of YWHAE-NUTM2A/B endometrial stromal sarcoma with confinement to the endometrium and lack of high-grade morphology. Int J Gynecol Pathol. 2017;36(2):165–71. https://doi.org/10.1097/PGP.0000000000000286.
219. Aisagbonhi O, Harrison B, Zhao L, Osgood R, Chebib I, Oliva E. YWHAE rearrangement in a purely conventional low-grade endometrial stromal sarcoma that transformed over time to high-grade sarcoma: importance of molecular testing. Int J Gynecol Pathol. 2017; https://doi.org/10.1097/PGP.0000000000000451.
220. Lee CH, Ali RH, Rouzbahman M, Marino-Enriquez A, Zhu M, Guo X, Brunner AL, Chiang S, Leung S, Nelnyk N, Huntsman DG, Blake Gilks C, Nielsen TO, Dal Cin P, van de Rijn M, Oliva E, Fletcher JA, Nucci MR. Cyclin D1 as a diagnostic immunomarker for endometrial stromal sarcoma with YWHAE-FAM22 rearrangement. Am J Surg Pathol. 2012;36(10):1562–70. https://doi.org/10.1097/PAS.0b013e31825fa931.
221. Chiang S, Lee CH, Stewart CJR, Oliva E, Hoang LN, Ali RH, Hensley ML, Arias-Stella JA 3rd, Frosina D, Jungbluth AA, Benayed R, Ladanyi M, Hameed M, Wang L, Kao YC, Antonescu CR, Soslow RA. BCOR is a robust diagnostic immunohistochemical marker of genetically diverse high-grade endometrial stromal sarcoma, including tumors exhibiting variant morphology. Mod Pathol. 2017;30(9):1251–61. https://doi.org/10.1038/modpathol.2017.42.
222. Fehr A, Hansson MC, Kindblom LG, Stenman G. YWHAE-FAM22 gene fusion in clear cell sarcoma of the kidney. J Pathol. 2012;227(4):e5–7. https://doi.org/10.1002/path.4040.
223. O'Meara E, Stack D, Lee CH, Garvin AJ, Morris T, Argani P, Han JS, Karlsson J, Gisselson D, Leuschner I, Gessler M, Graf N, Fletcher JA, O'Sullivan MJ. Characterization of the chromosomal translocation t(10;17)(q22;p13) in clear cell sarcoma of kidney. J Pathol. 2012;227(1):72–80. https://doi.org/10.1002/path.3985.
224. Lewis N, Soslow RA, Delair DF, Park KJ, Murali R, Hollmann TJ, Davidson B, Micci F, Panagopoulos I, Hoang LN, Arias-Stella JA 3rd, Oliva E, Young RH, Hensley ML, Leitao MM Jr, Hameed M, Benayed R, Ladanyi M, Frosina D, Jungbluth AA, Antonescu CR, Chiang S. ZC3H7B-BCOR high-grade endometrial stromal sarcomas: a report of 17 cases of a newly defined entity. Mod Pathol. 2018;31(4):674–84. https://doi.org/10.1038/modpathol.2017.162.
225. Hoang LN, Aneja A, Conlon N, Delair DF, Middha S, Benayed R, Hensley ML, Park KJ, Hollmann TJ, Hameed MR, Antonescu CR, Soslow RA, Chiang S. Novel high-grade endometrial stromal sarcoma: a morphologic mimicker of myxoid leiomyosarcoma. Am J Surg Pathol. 2017;41(1):12–24. https://doi.org/10.1097/PAS.0000000000000721.
226. Marino-Enriquez A, Lauria A, Przybyl J, Ng TL, Kowalewska M, Debiec-Rychter M, Ganesan R, Sumathi V, George S, McCluggage WG, Nucci MR, Lee CH, Fletcher JA. BCOR internal tandem duplication in high-grade uterine sarcomas. Am J Surg Pathol. 2018;42(3):335–41. https://doi.org/10.1097/PAS.0000000000000993.
227. Huynh KD, Fischle W, Verdin E, Bardwell VJ. BCoR, a novel corepressor involved in BCL-6 repression. Genes Dev. 2000;14(14):1810–23.
228. Fan Z, Yamaza T, Lee JS, Yu J, Wang S, Fan G, Shi S, Wang CY. BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. Nat Cell Biol. 2009;11(8):1002–9. https://doi.org/10.1038/ncb1913.
229. Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tifft C, Zhang L, Wilkie AO, van der Smagt JJ, Gorlin RJ, Burgess SM, Bardwell VJ, Black GC, Biesecker LG. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. Nat Genet. 2004;36(4):411–6. https://doi.org/10.1038/ng1321.
230. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, Spanhol-Rosseto A, Klein HU, Dugas M, Schindela S, Trifonov V, Schnittger S, Haferlach C, Bassan R, Wells VA, Spinelli O, Chan J, Rossi R, Baldoni S, De Carolis L, Goetze K, Serve H, Peceny R, Kreuzer KA, Oruzio D, Specchia G, Di Raimondo F, Fabbiano F, Sborgia M, Liso A, Farinelli L, Rambaldi A, Pasqualucci L, Rabadan R, Haferlach T, Falini B. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. Blood. 2011;118(23):6153–63. https://doi.org/10.1182/blood-2011-07-365320.

231. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, Pigneux A, Wetzler M, Stuart RK, Erba HP, Damon LE, Powell BL, Lindeman N, Steensma DP, Wadleigh M, DeAngelo DJ, Neuberg D, Stone RM, Ebert BL. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood. 2015;125(9):1367–76. https://doi.org/10.1182/blood-2014-11-610543.
232. Damm F, Chesnais V, Nagata Y, Yoshida K, Scourzic L, Okuno Y, Itzykson R, Sanada M, Shiraishi Y, Gelsi-Boyer V, Renneville A, Miyano S, Mori H, Shih LY, Park S, Dreyfus F, Guerci-Bresler A, Solary E, Rose C, Cheze S, Prebet T, Vey N, Legentil M, Duffourd Y, de Botton S, Preudhomme C, Birnbaum D, Bernard OA, Ogawa S, Fontenay M, Kosmider O. BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. Blood. 2013;122(18):3169–77. https://doi.org/10.1182/blood-2012-11-469619.
233. Dobashi A, Tsuyama N, Asaka R, Togashi Y, Ueda K, Sakata S, Baba S, Sakamoto K, Hatake K, Takeuchi K. Frequent BCOR aberrations in extranodal NK/T-Cell lymphoma, nasal type. Genes Chromosomes Cancer. 2016;55(5):460–71. https://doi.org/10.1002/gcc.22348.
234. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, Ambrogio L, Auclair D, Wang J, Song YK, Tolman C, Hurd L, Liao H, Zhang S, Bogen D, Brohl AS, Sindiri S, Catchpoole D, Badgett T, Getz G, Mora J, Anderson JR, Skapek SX, Barr FG, Meyerson M, Hawkins DS, Khan J. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov. 2014;4(2):216–31. https://doi.org/10.1158/2159-8290.CD-13-0639.
235. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, Ulyanov A, Wu G, Wilson M, Wang J, Brennan R, Rusch M, Manning AL, Ma J, Easton J, Shurtleff S, Mullighan C, Pounds S, Mukatira S, Gupta P, Neale G, Zhao D, Lu C, Fulton RS, Fulton LL, Hong X, Dooling DJ, Ochoa K, Naeve C, Dyson NJ, Mardis ER, Bahrami A, Ellison D, Wilson RK, Downing JR, Dyer MA. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature. 2012;481(7381):329–34. https://doi.org/10.1038/nature10733.
236. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, Greulich H, Lawrence MS, Lennon NJ, McKenna A, Meldrim J, Ramos AH, Ross MG, Russ C, Shefler E, Sivachenko A, Sogoloff B, Stojanov P, Tamayo P, Mesirov JP, Amani V, Teider N, Sengupta S, Francois JP, Northcott PA, Taylor MD, Yu F, Crabtree GR, Kautzman AG, Gabriel SB, Getz G, Jager N, Jones DT, Lichter P, Pfister SM, Roberts TM, Meyerson M, Pomeroy SL, Cho YJ. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature. 2012;488(7409):106–10. https://doi.org/10.1038/nature11329.
237. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DTW, Capper D, Sill M, Buchhalter I, Northcott PA, Leis I, Ryzhova M, Koelsche C, Pfaff E, Allen SJ, Balasubramanian G, Worst BC, Pajtler KW, Brabetz S, Johann PD, Sahm F, Reimand J, Mackay A, Carvalho DM, Remke M, Phillips JJ, Perry A, Cowdrey C, Drissi R, Fouladi M, Giangaspero F, Lastowska M, Grajkowska W, Scheurlen W, Pietsch T, Hagel C, Gojo J, Lotsch D, Berger W, Slavc I, Haberler C, Jouvet A, Holm S, Hofer S, Prinz M, Keohane C, Fried I, Mawrin C, Scheie D, Mobley BC, Schniederjan MJ, Santi M, Buccoliero AM, Dahiya S, Kramm CM, von Bueren AO, von Hoff K, Rutkowski S, Herold-Mende C, Fruhwald MC, Milde T, Hasselblatt M, Wesseling P, Rossler J, Schuller U, Ebinger M, Schittenhelm J, Frank S, Grobholz R, Vajtai I, Hans V, Schneppenheim R, Zitterbart K, Collins VP, Aronica E, Varlet P, Puget S, Dufour C, Grill J, Figarella-Branger D, Wolter M, Schuhmann MU, Shalaby T, Grotzer M, van Meter T, Monoranu CM, Felsberg J, Reifenberger G, Snuderl M, Forrester LA, Koster J, Versteeg R, Volckmann R, van Sluis P, Wolf S, Mikkelsen T, Gajjar A, Aldape K, Moore AS, Taylor MD, Jones C, Jabado N, Karajannis MA, Eils R, Schlesner M, Lichter P, von Deimling A, Pfister SM, Ellison DW, Korshunov A, Kool M. New brain tumor entities emerge from molecular classification of CNS-PNETs. Cell. 2016;164(5):1060–72. https://doi.org/10.1016/j.cell.2016.01.015.
238. Appay R, Macagno N, Padovani L, Korshunov A, Kool M, Andre N, Scavarda D, Pietsch T, Figarella-Branger D. HGNET-BCOR tumors of the cerebellum: clinicopathologic and molecular characterization of 3 cases. Am J Surg Pathol. 2017;41(9):1254–60. https://doi.org/10.1097/PAS.0000000000000866.
239. Astolfi A, Melchionda F, Perotti D, Fois M, Indio V, Urbini M, Genovese CG, Collini P, Salfi N, Nantron M, D'Angelo P, Spreafico F, Pession A. Whole transcriptome sequencing identifies BCOR internal tandem duplication as a common feature of clear cell sarcoma of the kidney. Oncotarget. 2015;6(38):40934–9. https://doi.org/10.18632/oncotarget.5882.
240. Karlsson J, Valind A, Gisselsson D. BCOR internal tandem duplication and YWHAE-NUTM2B/E fusion are mutually exclusive events in clear cell sarcoma of the kidney. Genes Chromosomes Cancer. 2016;55(2):120–3. https://doi.org/10.1002/gcc.22316.
241. Ueno-Yokohata H, Okita H, Nakasato K, Akimoto S, Hata J, Koshinaga T, Fukuzawa M, Kiyokawa N. Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney. Nat Genet. 2015;47(8):861–3. https://doi.org/10.1038/ng.3338.
242. Argani P, Kao YC, Zhang L, Bacchi C, Matoso A, Alaggio R, Epstein JI, Antonescu CR. Primary renal sarcomas with BCOR-CCNB3 gene fusion: a report of 2 cases showing histologic overlap with clear cell sarcoma of kidney, suggesting further link between BCOR-related sarcomas of the kidney and soft tissues. Am J Surg Pathol. 2017;41(12):1702–12. https://doi.org/10.1097/PAS.0000000000000926.
243. Wong MK, Ng CCY, Kuick CH, Aw SJ, Rajasegaran V, Lim JQ, Sudhanshi J, Loh E, Yin M, Ma J, Zhang Z, Iyer P, Loh AHP, Lian DWQ, Wang S, Goh SGH, Lim TH, Lim AST, Ng T, Goytain A, Loh AHL, Tan PH, Teh BT, Chang KTE. Clear cell sarcomas of the kidney are characterised by BCOR gene abnormalities, including exon 15 internal tandem duplications and BCOR-CCNB3 gene fusion. Histopathology. 2018;72(2):320–9. https://doi.org/10.1111/his.13366.
244. Specht K, Zhang L, Sung YS, Nucci M, Dry S, Vaiyapuri S, Richter GH, Fletcher CD, Antonescu CR. Novel BCOR-MAML3 and ZC3H7B-BCOR gene fusions in undifferentiated small blue round cell sarcomas. Am J Surg Pathol. 2016;40(4):433–42. https://doi.org/10.1097/PAS.0000000000000591.
245. Peters TL, Kumar V, Polikepahad S, Lin FY, Sarabia SF, Liang Y, Wang WL, Lazar AJ, Doddapaneni H, Chao H, Muzny DM, Wheeler DA, Okcu MF, Plon SE, Hicks MJ, Lopez-Terrada D, Parsons DW, Roy A. BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children. Mod Pathol. 2015;28(4):575–86. https://doi.org/10.1038/modpathol.2014.139.
246. Kao YC, Owosho AA, Sung YS, Zhang L, Fujisawa Y, Lee JC, Wexler L, Argani P, Swanson D, Dickson BC, Fletcher CDM, Antonescu CR. BCOR-CCNB3 fusion positive sarcomas: a clinicopathologic and molecular analysis of 36 cases with comparison to morphologic spectrum and clinical behavior of other round cell sarcomas. Am J Surg Pathol. 2018;42(5):604–15. https://doi.org/10.1097/PAS.0000000000000965.
247. Matsuyama A, Shiba E, Umekita Y, Nosaka K, Kamio T, Yanai H, Miyasaka C, Watanabe R, Ito I, Tamaki T, Hayashi S, Hisaoka M. Clinicopathologic diversity of undifferentiated sarcoma with BCOR-CCNB3 fusion: analysis of 11 cases with a reappraisal

of the utility of immunohistochemistry for BCOR and CCNB3. Am J Surg Pathol. 2017;41(12):1713–21. https://doi.org/10.1097/PAS.0000000000000934.
248. Antonescu CR, Sung YS, Chen CL, Zhang L, Chen HW, Singer S, Agaram NP, Sboner A, Fletcher CD. Novel ZC3H7B-BCOR, MEAF6-PHF1, and EPC1-PHF1 fusions in ossifying fibromyxoid tumors–molecular characterization shows genetic overlap with endometrial stromal sarcoma. Genes Chromosomes Cancer. 2014;53(2):183–93. https://doi.org/10.1002/gcc.22132.
249. Tanner EJ, Garg K, Leitao MM Jr, Soslow RA, Hensley ML. High grade undifferentiated uterine sarcoma: surgery, treatment, and survival outcomes. Gynecol Oncol. 2012;127(1):27–31. https://doi.org/10.1016/j.ygyno.2012.06.030.
250. Malouf GG, Lhomme C, Duvillard P, Morice P, Haie-Meder C, Pautier P. Prognostic factors and outcome of undifferentiated endometrial sarcoma treated by multimodal therapy. Int J Gynaecol Obstet. 2013;122(1):57–61. https://doi.org/10.1016/j.ijgo.2013.01.025.
251. Jin Y, Pan L, Wang X, Dai Z, Huang H, Guo L, Shen K, Lian L. Clinical characteristics of endometrial stromal sarcoma from an academic medical hospital in China. Int J Gynecol Cancer. 2010;20(9):1535–9.
252. Sciallis AP, Bedroske PP, Schoolmeester JK, Sukov WR, Keeney GL, Hodge JC, Bell DA. High-grade endometrial stromal sarcomas: a clinicopathologic study of a group of tumors with heterogenous morphologic and genetic features. Am J Surg Pathol. 2014;38(9):1161–72. https://doi.org/10.1097/PAS.0000000000000256.
253. Cotzia P, Benayed R, Mullaney K, Oliva E, Felix A, Ferreira J, Soslow R, Antonescu CR, Ladanyi M, Chiang S. Undifferentiated uterine sarcomas represent underrecognized high-grade endometrial stromal sarcomas. Lab Invest. 2018;98:415–6.
254. Chiang S, Cotzia P, Hyman DM, Drilon A, Tap WD, Zhang L, Hechtman JF, Frosina D, Jungbluth AA, Murali R, Park KJ, Soslow RA, Oliva E, Iafrate AJ, Benayed R, Ladanyi M, Antonescu CR. NTRK fusions define a novel uterine sarcoma subtype with features of fibrosarcoma. Am J Surg Pathol. 2018;42(6):791–8. https://doi.org/10.1097/PAS.0000000000001055.
255. Kaplan DR, Martin-Zanca D, Parada LF. Tyrosine phosphorylation and tyrosine kinase activity of the trk proto-oncogene product induced by NGF. Nature. 1991;350(6314):158–60. https://doi.org/10.1038/350158a0.
256. Barbacid M. Structural and functional properties of the TRK family of neurotrophin receptors. Ann N Y Acad Sci. 1995;766:442–58.
257. Kaplan DR, Miller FD. Neurotrophin signal transduction in the nervous system. Curr Opin Neurobiol. 2000;10(3):381–91.
258. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, Nathenson M, Doebele RC, Farago AF, Pappo AS, Turpin B, Dowlati A, Brose MS, Mascarenhas L, Federman N, Berlin J, El-Deiry WS, Baik C, Deeken J, Boni V, Nagasubramanian R, Taylor M, Rudzinski ER, Meric-Bernstam F, Sohal DPS, Ma PC, Raez LE, Hechtman JF, Benayed R, Ladanyi M, Tuch BB, Ebata K, Cruickshank S, Ku NC, Cox MC, Hawkins DS, Hong DS, Hyman DM. Efficacy of Larotrectinib in TRK fusion-positive cancers in adults and children. N Engl J Med. 2018;378(8):731–9. https://doi.org/10.1056/NEJMoa1714448.
259. Drilon A, Nagasubramanian R, Blake JF, Ku N, Tuch BB, Ebata K, Smith S, Lauriault V, Kolakowski GR, Brandhuber BJ, Larsen PD, Bouhana KS, Winski SL, Hamor R, Wu WI, Parker A, Morales TH, Sullivan FX, DeWolf WE, Wollenberg LA, Gordon PR, Douglas-Lindsay DN, Scaltriti M, Benayed R, Raj S, Hanusch B, Schram AM, Jonsson P, Berger MF, Hechtman JF, Taylor BS, Andrews S, Rothenberg SM, Hyman DM. A next-generation TRK kinase inhibitor overcomes acquired resistance to prior TRK kinase inhibition in patients with TRK fusion-positive solid tumors. Cancer Discov. 2017;7(9):963–72. https://doi.org/10.1158/2159-8290.CD-17-0507.
260. Bennett JA, Nardi V, Rouzbahman M, Morales-Oyarvide V, Nielsen GP, Oliva E. Inflammatory myofibroblastic tumor of the uterus: a clinicopathological, immunohistochemical, and molecular analysis of 13 cases highlighting their broad morphologic spectrum. Mod Pathol. 2017;30(10):1489–503. https://doi.org/10.1038/modpathol.2017.69.
261. Haimes JD, Stewart CJR, Kudlow BA, Culver BP, Meng B, Koay E, Whitehouse A, Cope N, Lee JC, Ng T, McCluggage WG, Lee CH. Uterine inflammatory myofibroblastic tumors frequently harbor ALK fusions with IGFBP5 and THBS1. Am J Surg Pathol. 2017;41(6):773–80. https://doi.org/10.1097/PAS.0000000000000801.
262. Parra-Herran C, Quick CM, Howitt BE, Dal Cin P, Quade BJ, Nucci MR. Inflammatory myofibroblastic tumor of the uterus: clinical and pathologic review of 10 cases including a subset with aggressive clinical course. Am J Surg Pathol. 2015;39(2):157–68. https://doi.org/10.1097/PAS.0000000000000330.
263. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih Ie M, Kurman R, Dao F, Levine DA, Giuntoli R, Roden R, Eshleman JR, Carvalho JP, Marie SK, Papadopoulos N, Kinzler KW, Vogelstein B, Diaz LA Jr. Evaluation of DNA from the papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med. 2013;5(167):167ra164. https://doi.org/10.1126/scitranslmed.3004952.
264. Wang Y, Li L, Douville C, Cohen JD, Yen TT, Kinde I, Sundfelt K, Kjaer SK, Hruban RH, Shih IM, Wang TL, Kurman RJ, Springer S, Ptak J, Popoli M, Schaefer J, Silliman N, Dobbyn L, Tanner EJ, Angarita A, Lycke M, Jochumsen K, Afsari B, Danilova L, Levine DA, Jardon K, Zeng X, Arseneau J, Fu L, Diaz LA, Jr., Karchin R, Tomasetti C, Kinzler KW, Vogelstein B, Fader AN, Gilbert L, Papadopoulos N. Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. Sci Transl Med. 2018;10(433). https://doi.org/10.1126/scitranslmed.aap8793
265. Gibson WJ, Hoivik EA, Halle MK, Taylor-Weiner A, Cherniack AD, Berg A, Holst F, Zack TI, Werner HM, Staby KM, Rosenberg M, Stefansson IM, Kusonmano K, Chevalier A, Mauland KK, Trovik J, Krakstad C, Giannakis M, Hodis E, Woie K, Bjorge L, Vintermyr OK, Wala JA, Lawrence MS, Getz G, Carter SL, Beroukhim R, Salvesen HB. The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. Nat Genet. 2016;48(8):848–55. https://doi.org/10.1038/ng.3602.
266. Boronow RC. Endometrial cancer: not a benign disease. Obstet Gynecol. 1976;47(5):630–4.
267. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4. https://doi.org/10.1158/2159-8290.CD-12-0095.

Genomic Applications in Ovarian Cancer

32

Martin Köbel and James D. Brenton

Ovarian Cancer

Introduction

Overall survival rates for women with advanced epithelial ovarian cancer (ovarian carcinoma) have remained unchanged over the past three decades, and fewer than 40% of patients remain alive at 5 years after diagnosis [1]. High-grade serous ovarian carcinoma (HGSOC) accounts for the majority of these cases. Mortality for HGSOC has not been altered by the use of complex cytotoxic chemotherapy combinations [2–4], and the lack of progress in improving outcomes reflects its unique biology and extreme genomic complexity. Here, we review key approaches to diagnosis and stratification of HGSOC that are now needed to advance treatment options for patients. The most important clinical questions for the pathologist remain how to unequivocally classify the different histotypes of ovarian carcinoma and which additional genomic data may identify individuals with high risk of response or relapse. This section will concentrate on recent molecular insights that are likely to be highly relevant to clinical care over the next 5 years.

Cell of Origin of Epithelial Ovarian Carcinoma

Genomic classifiers for cancer are strongly determined by the cell of origin of carcinoma. Long-standing confusion about the cellular origin of ovarian carcinoma has hindered the development of strong morphological and molecular classifications. The prevailing model until 2001 was that ovarian carcinoma arose from metaplastic changes in the normal ovarian epithelium (NOE) and could arise from structural epithelial abnormalities on the surface of the ovary, such as clefts and inclusion cysts (reviewed in Ref. [5]). Transdifferentiation of the NOE into HGSOC, endometrioid and mucinous subtypes was hypothesized to follow developmental pathways of Müllerian differentiation in the fallopian tube, endocervix and endometrium, respectively. Dubeau [5] first suggested that 'ovarian' cancers might arise from the secondary Müllerian system, including endosalpingiosis, endometriosis and para-ovarian and paratubal cysts. Evidence that preinvasive lesions of HGSOC were present in the distal fallopian tube was discovered by meticulous pathological studies on the fimbria of women with *BRCA1* or *BRCA2* mutations collected after risk-reducing surgery [6–12]. Similar studies on sporadic HGSOC cases showed that serous tubal intraepithelial carcinomas (STICs) could be identified in 75% of cases and that these fallopian tube lesions were under-diagnosed by standard histopathological examination [13]. Identical *TP53* mutations were demonstrated in STIC and ovarian metastases from the same patient, providing strong evidence for a single clonal origin for HGSOC arising in the tubal epithelium [10, 13]. These studies now strongly implicate STIC in the fallopian tube (and potentially in endosalpingiosis tissues) as the cells of origin for HGSOC.

The epidemiological risk factors for endometrioid (EOC) and clear-cell ovarian cancers (OCCA) are distinctly different from HGSOC and, in particular, are strongly linked to endometriosis [14, 15]. *ARID1A* encodes BAF250a which is a component of the SWI/SNF nucleosome remodelling complex (reviewed in Ref. [16]) and is a common driver for endometriosis-associated ovarian carcinomas [17, 18]. Loss of ARID1A expression and *ARID1A* or *PIK3CA* mutation has also been demonstrated in endometriosis tissue contiguous with OCCA [17–20]. These data provide very strong molecular evidence that endometriosis tissue is the precursor lesion for EOC and OCCA.

M. Köbel
Pathology and Laboratory Medicine, Foothills Medical Centre, Calgary, AB, Canada

J. D. Brenton (✉)
Functional Genomics of Ovarian Cancer Laboratory, Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
e-mail: james.brenton@cruk.cam.ac.uk

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_32

Taken together, these data show that histotypes of ovarian carcinoma differ in cell of origin, epidemiology, natural history and biology. It is therefore logical to view them as distinct and different diseases, and the term 'histotype' deprecated as its use implies a common tissue of origin. As genomically targeted therapies for EOC and OCCA cancer are not currently in the clinic as standard of care treatments, the remainder of this chapter will concentrate on potential biomarkers for personalized treatment approaches in HGSOC and their relationship with platinum and PARP inhibitor sensitivity and resistance.

Molecular Classification of HGSOC

High-grade serous carcinoma was introduced as a new diagnostic category in the 2014 WHO classification. This decision reflects the strong consensus view that HGSOC is completely distinct from low-grade serous carcinoma as a separate neoplasm, not a continuum in grade. With other refinements, ovarian carcinomas are now classified into five clinically relevant histotypes: high-grade serous, low-grade serous, endometrioid, clear-cell, and mucinous carcinomas. The mainstay for diagnosis still depends on conventional histomorphology, but ancillary molecular tests are being increasingly integrated into decision-making. For example, a simple decision tree of four immunohistochemical markers assigns five histotypes with an almost 90% accuracy (Fig. 32.1) [21]. The wider use of additional immunohistochemical markers has provided a more objective basis for diagnosis making histotype assignment highly reproducible, even on small pretreatment biopsies [22].

These diagnostic advances have been enabled by sequential optimization and refinement of interpretation of classical immunohistochemical markers such as p53. Historically, p53 immunohistochemistry has been interpreted as negative or positive using arbitrary cut-offs without detailed comparison to the gold standard of *TP53* mutation. With the advent of highly sensitive immunohistochemical platforms, it has become possible to consistently recognize additional abnormal patterns of p53 expression. Nonsynonymous *TP53* 'gain-of-function' mutations cause strong nuclear p53 staining ('overexpression') because of impaired MDM2 degradation and genotoxic stress. Loss of function mutations, including stopgain, splicing or indels, are present in one third of HGSOC and cause a complete absence of p53 protein in tumour cells owing to degradation of mRNA by nonsense-mediated RNA decay. Using newer and more sensitive p53 immunohistochemistry, it is now feasible to consistently distinguish complete absence of p53 staining from low-level wild-type p53 expression by comparison to normal cells, such as fibroblasts or lymphocytes, as internal controls. In addition, cytoplasmic p53 staining is also now increasingly identified and occurs when mutations disrupt the nuclear localization domain.

These three abnormal p53 patterns contrast with the normal p53 wild-type expression, which is defined as variable (with respect to distribution and intensity) nuclear p53 expression between overexpression and complete absence. When compared to the 'gold standard' of next-generation sequencing, the overall accuracy of p53 immunohistochemistry to predict a *TP53* mutation is 97% with a specificity of 100% [23]. This has direct clinical relevance for interpretation of small biopsies—as if p53 staining is abnormal on well-processed tissue, there is almost certainly an underlying *TP53* mutation—a critical diagnostic feature for HGSOC. However, not all *TP53* mutations result in abnormal staining, and 4% of mutated cases show low-level wild-type staining as a false-negative result. This staining pattern occurs in cases with mutations in the 3′ portion of *TP53* which may result in more stable RNA or impair nonsense-mediated RNA decay. The main clinical value of p53 immunohistochemistry is the distinction of low-grade serous tumours (including low-grade serous carcinoma and serous borderline tumour, which do not contain *TP53* mutations) from high-grade serous carcinomas, which ubiquitously harbour *TP53* mutations. Due to its perfect specificity, an abnormal p53 immunohistochemistry practically excludes the possibility of a low-grade serous tumour. Although the sensitivity is >95%, there are a few circumstances in which p53 immunohistochemistry will result in wild-type pattern despite underlying mutation. Pathologists should be aware of this possibility and integrate with morphology. In such instances, further testing, e.g. p16 IHC, MAPK pathway mutation testing or direct *TP53* sequencing, should be considered [24, 25].

TP53 mutations also occur in mucinous, endometrioid and clear-cell carcinoma, albeit at much lower frequency in endometrioid (9%) and clear-cell carcinomas (13%). Here, the serous cell lineage marker WT1, which is expressed in the normal fallopian tube and the mesothelial lining of the peritoneal cavity and virtually 100% of high-grade serous carcinomas, becomes important. WT1 expression is virtually absent in all clear or mucinous carcinomas but can be expressed in a minority of endometrioid carcinomas (10–15%). Hence, sole reliance on WT1 will not accurately distinguish endometrioid from high-grade serous carcinomas. However, if WT1 expression is positive together with abnormal p53 expression, this has a high (~99%) specificity for a high-grade serous carcinoma. There is now strong consensus among gynaecological pathologists that tumours with this combination of expression profiles should be classified as HGSOC. The expression of WT1 and p53 is extremely stable in spatial and temporal manner making them robust diagnostic markers [26]. Of note, quality assurance programmes are required to monitor the performance of these tests if the diagnosis relies heavily on them [27]. With the realization

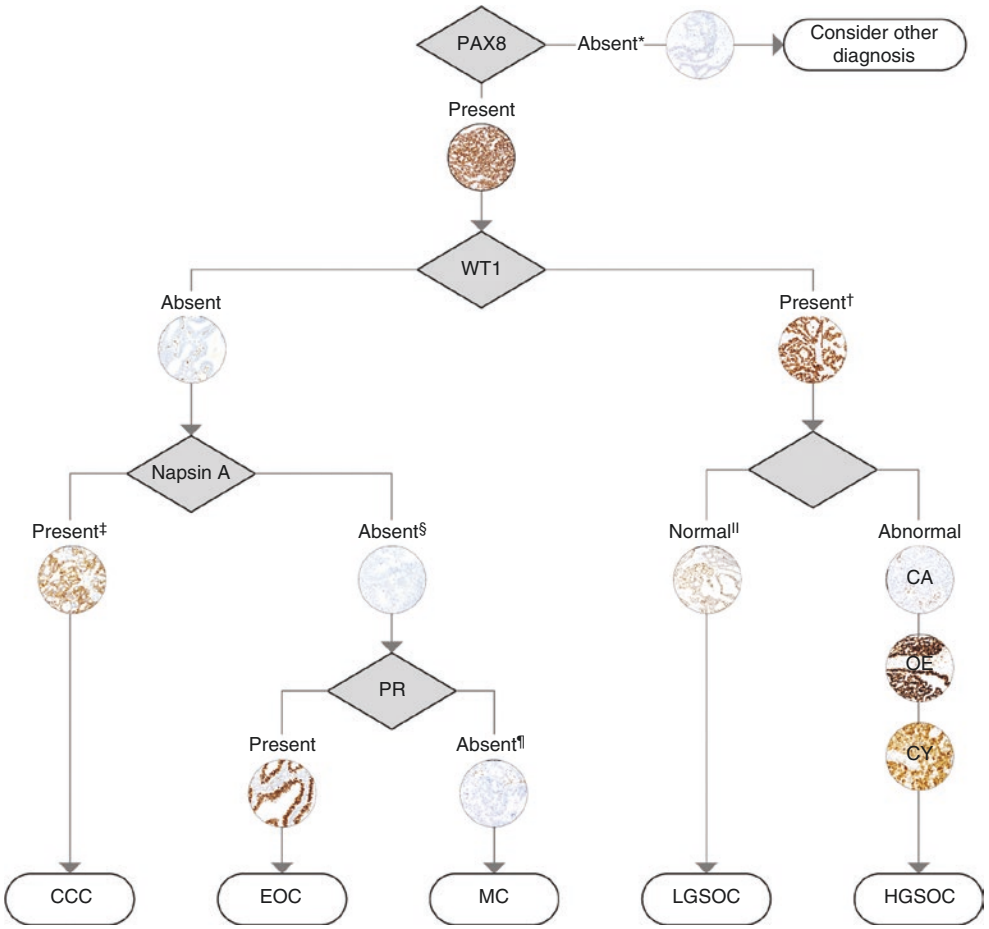


Fig. 32.1 Immunohistochemical decision tree for ovarian carcinoma histotyping. The overall accuracy of this decision tree is close to 90%. Note that H&E morphology should be integrated with the immunohistochemical staining results (see footnotes below). PAX8 may be used as an entry marker to confirm Müllerian cell lineage (*CCC* clear-cell carcinoma, *EOC* endometrioid ovarian carcinoma, *MC* mucinous carcinoma, *LGSOC* low-grade serous ovarian carcinoma, *HGSOC* high-grade serous ovarian carcinoma, *CA* complete absence with retained internal control, *OE* overexpression, *CY* cytoplasmic expression)

*PAX8 is absent in 15% of endometrioid and slightly more than half of mucinous carcinoma. WT1 segregated serous from non-serous cases

†WT1 is present in about 10% of endometrioid carcinomas. *TP53* mutation segregates high-grade from low-grade serous carcinoma

‡Napsin A is present in up to 10% of endometrioid carcinomas

§Napsin A is absent in 10–20% of clear-cell carcinomas on small tissue specimens because of focal Napsin A expression in clear-cell carcinomas

‖2–5% of high-grade serous carcinomas can show p53 normal wild-type expression and still harbour a *TP53* mutation

¶PR is absent in about 15% of endometrioid carcinomas

that many high-grade endometrioid carcinomas and all transitional cell carcinomas were morphological variants of high-grade serous carcinomas [28], there is now a trend to use high-grade serous carcinoma as a diagnostic wastebasket category. Immunohistochemistry and particularly the combination of WT1 and p53 can identify the rare examples of high-grade endometrioid carcinomas [29]. Yet, there remain rare examples, in which immunohistochemistry will contradict morphology. Since this has potential ramifications for genetic counselling (high-grade serous carcinoma for BRCA1/2, endometrioid for Lynch syndrome) and targeted therapy (high-grade serous for PARP inhibitors, endometrioid with mismatch repair deficiency for PDL1 inhibitors), additional molecular tests should be performed in such cases.

An uncommon diagnostic problem is the presence of serous carcinoma both within the endometrium and the tubo-ovarian region. The differential diagnosis includes synchronous endometrial serous and tubo-ovarian high-grade serous, drop-down metastasis from a tubo-ovarian high-grade serous carcinomas arising in the fallopian tube versus an endometrial serous carcinoma with metastatic tubo-ovarian involvement [30–32]. Since the morphology is identical and *TP53* mutations are ubiquitous in both, the most reliable marker is WT1, which is expressed in virtually all tubo-ovarian high-grade serous carcinomas but only in 34% of endometrial serous carcinomas [30]. While positivity is not specific, negative staining strongly suggests an endometrial primary.

With the increasing use of ancillary techniques, it became also clear that the vast majority of ovarian carcinomas that were previously considered to be composed of mixed components are in fact clonal and should be assigned to one of the five major histotypes. A characteristic feature of HGSOC is intratumoural heterogeneity owing to underlying chromosomal instability. This may often result in areas of morphological mimicry of other histotypes historically classified as mixed. However, a recent study demonstrated that true mixed carcinomas, defined by evidence of a common ancestral clone (in contrast to independent collision tumours) and evidence of divergent cell lineage are exceedingly rare [32].

As a consequence of improved classification, the association with hereditary cancer syndromes becomes clearer as only high-grade serous carcinomas are associated with germline BRCA1/2 mutation but not with Lynch syndrome [33].

Expression Analysis

Initial attempts to define prognostic and predictive signatures for chemoresistance in ovarian carcinoma focused on the use of expression microarray studies [34–44]. These studies have had little success in identifying useful biomarkers [45–47]. This is in marked contrast to similar efforts in breast cancer, where expression profiling of a relatively small number of cases provided new molecular classifications with prognostic value that have been quickly applied in research and in the clinic [48, 49]. Expression signatures in breast cancer are dominated by the transcriptional effects of ER-, PR- and *ERBB2*-mediated pathways. Although it is assumed that HGSOC most closely represents triple negative breast cancer (see [50] and below), a large analysis of 2933 ovarian carcinoma showed that strong PR staining was independently associated with improved disease-specific survival in HGSOC, although this was limited to 7% of HGSOC [51]. This provides strong evidence for the possible utility of measuring ER and PR in HGSOC. Preclinical studies have identified features associated with functional ERα signalling and suggest that fulvestrant which is a selective ERα downregulator may be more effective than tamoxifen in blocking ERα action [52].

Expression profiling has defined four main molecular subtypes in HGSOC, described as C1 (high stromal response), C2 (high immune signature), C4 (low stromal response) and C5 (mesenchymal, low immune signature) [44, 50]. Although reproducible, the prognostic value of these signatures is weaker than that of ER/PR expression discussed above and has not had clinical impact. Substantial expansion of the classifications identified by Tothill has recently been performed by the TCGA [53]. Importantly, this has shown that individual HGSOC samples can show multiple subtype signatures, which further reduce the predictive utility of these classifications.

These expression signatures are strongly modulated by stroma and infiltrating immune cells. Quantitative and semi-quantitative analysis of the tumour microenvironment and constitutive immune cells offers potential strong prognostic information. In colorectal cancer, development of the 'immunoscore' is altering risk prediction for recurrence [54–56]. Following seminal observations by Coukos and colleagues [57], the strong prognostic effects of infiltrating T and B cells in ovarian cancer have been confirmed [58–60] together with associations with *BRCA1* loss [61] and paracrine inhibition of T-cell infiltration by endothelial Fas ligand [62]. In addition, quantitative analysis of tumour stroma and lymphocytes in H&E sections from 91 advanced primary ovarian cancers showed that the stromal percentage was independent predictor of survival [63]. In ovarian carcinomas, the strongest link between TIL infiltrate and outcome is for HGSOC [64] and that there was a near-log-linear relationship between TIL levels and overall survival. Comparing the extreme groups, HGSOC patients with high TIL levels had a median survival of 5.1 years compared to 2.8 years for patients with no TIL, suggesting that immune response has a remarkable effect on patient survival. Although CD8-positive TIL is a robustly validated prognostic marker, its clinical application is still limited as even within the group of HGSOC patients with high TIL, less than half of patients are alive after 5 years. It remains unclear whether the immune response in HGSOC is driven by specific neoantigens, such as mutant p53, or by viral mimicry arising from DNA damage repair [33, 65].

Proteomic analysis on 412 HGSOC samples with reverse phase protein arrays provided improved prognostic information as compared to expression signatures [66]. Refinement of the protein signature showed that five proteins (AR, BID, phosphorylated TAZ, phosphorylated EGFR and HSP70) were associated with longer PFS and increased expression in the low-risk group. Four proteins (STAT5α, phosphorylated PKCα, phosphorylated MEK1 and EEF2) were associated with shorter PFS and increased expression in the high-risk group. AR may be a compelling marker and should now be tested across large sample sets using tissue microarrays.

Mutational Spectrum of High-Grade Serous Ovarian Carcinoma

Data from the Cancer Genome Anatomy Project has provided a comprehensive mutation survey of fresh-frozen samples from HGSOC [50]. Whole-exome sequencing from 489 HGSOC cases showed profound mutational heterogeneity between patients and confirmed relatively few recurrent gene mutations, involving the tumour suppressor genes *TP53*, *BRCA1*, *BRCA2* and *NF1*. Of note, oncogenic mutations in *EGFR*, *PIK3CA*, *BRAF* and *KRAS* are extremely infrequent and may account for less than 1% of HGSOC cases. Indeed, the presence of a *BRAF* or *KRAS* mutation in

a serous neoplasm should always suggest the diagnosis of low-grade serous carcinoma and confirmation of lack of *TP53* mutation.

Mutations in *TP53* Are Ubiquitous in High-Grade Serous Ovarian Carcinoma

The *TP53* gene encodes the p53 tumour suppressor protein and is among the most frequently mutated genes in human cancer [67, 68]. The frequency of *TP53* mutation is different between ovarian carcinoma histotypes [69–72]. Demonstration that *TP53* mutation is ubiquitous in HGSOC was first provided by Sanger sequencing results from 145 women with serous neoplasms [73]. Whole-exome sequencing by the Cancer Genome Atlas project also confirmed near 100% mutation rate. These studies underscore the essential role that *TP53* mutation has as an early driver mutation for HGSOC and are consistent with the findings of p53 accumulation and mutation in fallopian tube p53 signature foci and STICs [12]. However, the fact that signature foci in the fallopian tube may have *TP53* mutation, but do not develop into STICs, suggests that loss of p53 function is required but not sufficient for the development of HGSOC.

Determining *TP53* mutational status should now be considered for all women with HGSOC, although, at present, this result is not predictive of outcome. Firstly, presence of *TP53* mutation provides diagnostic information and excludes low-grade serous carcinoma. Secondly, the type of mutation may give prognostic information. As in other cancers, the majority of *TP53* mutations in HGSOC are missense, but null mutations (including nonsense, frameshift and splice-site mutations) account for about one third [50, 73]. Thirdly, *TP53* mutation-specific therapy is now available or will soon be placed into early phase trials [74–77], and precise knowledge of the type of mutation may be important for eligibility or stratification.

High-Grade Serous Carcinoma Is Defined by Profound Structural DNA Aberrations

In contrast to the relative lack of recurrent oncogenic mutation, the TCGA analysis showed profound and recurrent copy number aberrations (CNAs). Recurrent focal CNAs were identified in 63 regions including *CCNE1*, *MYC* and *MECOM*. There is a hyperbolic relationship between the frequency of nucleotide substitutions and structural variants across common cancers, and HGSOC cases are the extreme outliers with the highest number of structural variants and low numbers of oncogenic mutations [78].

Amplification of *CCNE1* which encodes the cell cycle checkpoint protein cycline E1 may be a therapeutic target. A previous study of 118 HGSOC cases showed that amplification of chromosome 19q12 was a strong negative prognostic factor and correlated with platinum resistance [79–81]. As well as *CCNE1*, this region also includes the anti-apoptotic oncogene *C19orf2* (also known as *URI*). Knockdown of *CCNE1* in ovarian carcinoma cell lines with 19q12 amplification paradoxically increased cisplatin resistance in short-term assays, although it did result in reduced clonogenic survival, suggesting an oncogenic effect [82]. Studies of the effects of amplification or overexpression of *C19orf2* in vitro and in vivo showed increased cisplatin resistance, mediated by increased S6 K1-BAD survival signalling [83]. Mechanistic insight into platinum resistance in *CCNE1*-amplified cases has come from siRNA screens showing that *BRCA1* and members of the ubiquitin pathway are required for survival in cancers that have *CCNE1* amplification [84]. Use of CDK2 inhibitors or targeting the G2 checkpoint with Wee-1 inhibitors may be relevant therapeutic approaches [85]. Despite the strong prognostic effect of 19q12 amplification, CNAs are not routinely assayed in clinical practice.

Homologous Recombination and High-Grade Serous Carcinoma

BRCA1/2 Mutation Has a Strong Survival Effect

In comparison to other epithelial cancers, HGSOC shows the highest sensitivity to platinum-based chemotherapy, and initial response rates are 70–80% when surgery is combined with chemotherapy. Uniquely, a substantial proportion of patients with relapsed disease will respond to retreatment with platinum chemotherapy. The time interval between diagnosis and development of progressive disease is the strongest predictor of response rates to retreatment and is used clinically to define 'platinum-resistant' and 'platinum-sensitive' relapsed disease [86]. Despite the critical importance of platinum therapy, primary and acquired resistance is still poorly understood [87].

The relative hypersensitivity of HGSOC to treatment may be explained by high rates of intrinsic homologous recombination deficiency (HRD). Carboplatin induces inter- and intra-strand cross-linking that results in both single- and double-strand DNA breaks. In normal cells, double-strand break damage can be repaired by either error-free homologous recombination (HR) or by the error-prone non-homologous end-joining (NHEJ) pathway. Cancer cells that lack BRCA1 or BRCA2 function cannot carry out HR repair and are therefore very sensitive to platinum-induced DNA damage that induces apoptotic death.

BRCA1 and *BRCA2* mutation carriers are largely restricted to HGSOC, and previous data obtained from all ovarian carcinoma histotypes may underestimate the prevalence of these mutations in this subgroup. Recent studies of *BRCA1* and *BRCA2* in HGSOC patients have shown combined germline mutation rates of up to 23%, and additional pathogenic

somatic mutations can also be found in noncarrier cases [50, 88–90]. This is correlated with significantly longer survival in HGSOC patients with *BRCA1* or *BRCA2* germline mutations as compared with noncarriers [88, 91–93]. A large collaborative study has recently shown that 5-year overall survival was 36% for noncarriers ($n = 2666$), 44% for *BRCA1* carriers ($n = 909$) and 52% for *BRCA2* carriers ($n = 304$). Highly significant survival differences remained after additional adjustment for major prognostic factors (*BRCA1*: HR, 0.73; *BRCA2*: HR, 0.49; $P < 0.001$ for both) [94].

These findings have important implications for the development of predictive biomarkers for women with HGSOC as they suggest that cases with *BRCA1/2* mutations will have longer progression-free survival and be overrepresented in the clinically defined platinum-sensitive relapsed group (progression >6 months from primary treatment). Support for this hypothesis is provided by the AOCS study of 1001 cases, in which *BRCA1* or *BRCA2* carriers had improved rates of progression-free as well as overall survival [90]. Carriers were less frequent in patients who progressed <6 months from primary treatment (platinum-resistant group). Mutation-negative patients who responded to multiple courses of platinum-based treatment were more likely to carry somatic *BRCA1/2* mutations. Therefore, assessing the *BRCA1* and *BRCA2* mutation status by sequencing may be highly predictive for outcome and now should be included for stratification in clinical trials. It is likely that this testing will be provided by the medical genetics service in most hospitals, but somatic sequencing of *BRCA1* and *BRCA2* may be coordinated by the molecular pathologist.

Mutational Signatures of HRD May Be Strong Predictive Biomarkers for HGSOC

Recent publications have suggested that HRD may be a common mutator phenotype in HGSOC patients. As discussed above, more than 20% of HGSOC cases may have germline or somatic *BRCA1/2* mutations that will cause HRD. In addition, familial non-*BRCA1/2* cases are enriched for mutations in HR pathway genes, including *RAD51C* [95, 96], *RAD51D* [97] and *BRIP1* [98], suggesting that there is common involvement of DNA repair enzymes in the pathogenesis of HGSOC. Thirdly, using a functional assay of HR status based on RAD51 focus formation after in vitro DNA damage, direct testing was performed on 24 primary cultures of ovarian carcinoma, suggesting that 16 (64%) were HR deficient as these data were highly correlated with in vitro response to PARPi [99]. A similar rate was predicted from the analysis carried out by the TCGA using survival data and mutation status in candidate HR-related genes [50].

Functional studies have intrinsic limitations for clinical assessment, and HRD has been correlated with loss of heterozygosity and genomic imbalances [100, 101]. Detailed analyses of whole-genome sequencing data has revealed mutational signatures based on patterns of nucleotide substitution that can reveal the imprint of mutagenic processes accumulated over the lifetime of a cancer cell [102]. By comparison across whole-genome sequencing data from multiple cancer types, 30 distinct mutational signatures have been identified. Signature 3 is frequent in breast, ovarian and pancreatic cancers and is strongly correlated with HRD and mutation in *BRCA1* and *BRCA2*, suggesting that this may provide genomic predictor for platinum and PARP sensitivity. The predictive power of signature 3 has been further improved with the addition of other genomic features, including microhomology and deletion size [103]. Application of these methods has the potential to accurately identify a larger proportion of HGSOC with HRD and who are suitable for PARP therapy.

Intratumoural Heterogeneity in HGSOC May Contribute to Platinum Resistance

Although platinum resistance has been intensively studied using in vitro models, there is only limited evidence that mechanisms commonly seen in cell lines, involving altered apoptosis pathways, increased drug excretion or tolerance to DNA adducts, occur in clinical samples [87]. The importance of HRD in determining platinum sensitivity is underscored by studies showing that in a small proportion of *BRCA1* and *BRCA2* carriers, resistance is caused by reverting or secondary mutations that restore somatic BRCA1 or BRCA2 protein function [104–109]. Although it is unknown if these mutations predate platinum treatment, their existence strongly argues that the acquisition of a resistant phenotype involves selection effects.

It has been proposed that genetic heterogeneity could explain the development of drug resistance in HGSOC [110] based on the demonstration of divergent evolution between sensitive and resistant subclones from three cases of HGSOC [111]. The genetic changes in the cell lines derived before and after clinical resistance developed were incompatible with a simple linear model, and the most parsimonious explanation was that resistant lineages were present as a minor subpopulation of the tumour mass at the time of first therapy [110, 111]. The possibility that significant genetic heterogeneity existed within many cancers was originally proposed by Nowell in 1976 and demonstrated using cytogenetic methods in 1978 [112, 113]. However, it is only recently that the degree and types of genetic variation present within an individual's cancer could be accurately characterized through the advances of next-generation sequencing and high-accuracy SNP CGH arrays [110, 111, 114–121]. Loss of heterozygosity data has provided evidence for genetic heterogeneity in HGSOC [114, 115], and more detailed studies have shown that subclonal populations preexist in epithelial

tumours [116, 117, 122, 123] and undergo treatment-related selection in leukaemias [124–126] and breast [118] and renal cancer [127].

There is strong evidence that selective effects can explain drug resistance in haematological cancers [128, 129]. Point mutations conferring resistance to imatinib have been shown to be present at low frequency before treatment in both acute lymphocytic leukaemia and chronic myeloid leukaemia. At relapse, these mutations are present in high frequency in the leukaemic blasts [124–126]. In acute lymphocytic leukaemia, a higher frequency of resistant mutations in the initial presenting disease is directly correlated with shorter remission [124]. Similar changes have been shown in breast cancer using next-generation sequencing of primary and relapsed disease in a single case of lobular breast cancer [118].

Intratumoural genetic heterogeneity in HGSOC has been demonstrated both within a region of tumour and between metastatic sites [114, 115, 130–133]. These genetic differences could be expected to alter chemosensitivity. Consistent with this, variable in vitro responses to a variety of chemotherapeutic agents were observed in primary ovarian carcinoma cells obtained from different metastatic sites from the same individual, suggesting the existence of genetically or epigenetically diverse subpopulations [134]. Similar differential effects on response have also been shown during chemoradiation of advanced cervical cancer [110]. In three out of ten cases, there were distinct genetic subpopulations before treatment, and these regions showed differential responses to chemoradiotherapy, leading to mixed response and selection of resistant disease. Studies of cell lines derived before and after relapse in three cases of HGSOC have found that presentation and resistant disease are not linearly genetically related, showing that the relapsed genotype cannot have arisen by direct descent from a dominant clone at disease presentation [111]. These studies suggest a hypothesis for platinum-resistant disease in HGSOC where strong selection for a minor resistant clonal population occurs, rather than genetic progression from disease at presentation. These models may also be relevant in platinum-sensitive disease as it will be important to determine whether reverting mutations in *BRCA1* and *BRCA2* exist before treatment, as is seen for *ABL* mutations in chronic myeloid leukaemia.

Understanding Tumour Heterogeneity and Platinum Resistance in HGSOC Requires Sequential Biopsy and Plasma Studies

In order to understand how platinum resistance evolves in HGSOC, it is now essential to undertake genomic comparisons of sequential tissues and plasma ctDNA from diagnosis to relapsed/progressive disease. Image-guided core biopsy (IGCB) of ovarian cancer, either under ultrasound or CT guidance, is safe and feasible and is the standard of care for diagnosis of patients with suspected ovarian cancer, particularly prior to neoadjuvant chemotherapy [135, 136]. However, re-biopsy of relapsed HGSOC tissue is not a standard of care, and the cost for each biopsy may be significant. Although studies on malignant cells from abdominal ascites can offer a less morbid approach to assaying relapsed disease, only a small proportion of patients with relapsed disease will have ascites that can be aspirated. In addition, malignant cells in ascites do not allow examination of tumour stroma and tumour vasculature, which can both profoundly affect tumour biology.

Circulating Tumour DNA Can Be Used To Identify Mutations and Track Tumour Evolution

Although CT and serum CA125 are the standard of care for estimating disease burden and response in HGSOC, there is an urgent need for cheap and sensitive blood-based markers to provide molecular measures of response and to identify minimal residual disease. Circulating DNA in plasma and serum contains tumour-specific sequences that have recently been exploited in small numbers of patients as personalized biomarkers. Evidence that cancer patients had higher levels of circulating free DNA in serum was first shown in 1977 [137], which prompted exploratory studies using total DNA as a potential biomarker in several cancers, including ovarian carcinoma [138]. However, these assays were not specific and had confounding effects that prevented reliable clinical use.

Mutation-specific assays of circulating tumour DNA (ctDNA) can be developed by genotyping or sequencing the tumour followed by the design of allele- or mutation-specific PCR-based assays to detect tumour DNA. The feasibility of this approach has been demonstrated by accurate quantitation of dynamic changes in ctDNA from colorectal cancer, non-small cell lung cancer, breast cancer and osteosarcoma, using assays developed for tumour-specific mutations and rearrangements [123, 139–142]. These early data suggest that ctDNA dynamics compare favourably with other diagnostic modalities, including serum tumour markers and CT imaging. Use of ctDNA is potentially very important for studies in HGSOC as the detection of circulating tumour cells (CTC) has been very difficult. Using the Veridex CellSearch system, only 14% of patients with advanced ovarian carcinoma had >2 CTCs (median 0) indicating that CTC counts are too insensitive to be used as a measure of response [143].

As *TP53* mutations are ubiquitous in high-grade serous ovarian cancer [73], this important clinical feature can be used to develop noninvasive plasma assays that have the potential to measure treatment response and to monitor evolution of clonal populations [144]. Tagged-amplicon deep sequencing (TAm-Seq) uses next-generation sequencing technology to reliably sequence candidate genes and genomic regions from low amounts of potentially degraded DNA

[144]. This has been applied to reliably identify *TP53* mutations in the blood when present at allelic frequencies of 2–65% [144]. Application of Tam-Seq assays to sequential samples revealed the progressive emergence of a subclonal HGSOC population marked by a new EGFR mutation. Sequencing of archival paraffin blocks from the original surgery for this case showed that the *EGFR*-mutated clone was present at very low allele fractions in omental deposits. In another example, it was possible to discriminate relapsed disease between previously diagnosed colorectal and ovarian cancer by identifying the specific *TP53* mutation. This work demonstrates that sequencing of ctDNA offers the potential for a 'liquid biopsy' that can be used for personalized genomic profiling and to explore clonal evolution and the potential of differential response to treatment. Whole-exome sequencing can detect a wider number of mutations and clonal change in patients with high levels of circulating tumour DNA [145]. Comparison between ctDNA and matched tumour biopsies showed that the mutational readout from the blood was wider than from tissue and that copy number abnormalities were conserved. Exome analysis of plasma ctDNA could provide higher sensitivity for capturing the complexity of tumour evolution and can detect DNA mutations from multiple lesions in the same individual without requiring multiple biopsies.

ctDNA can reliably estimate response to chemotherapy in patients with relapsed HGSOC [146]. Comparison of ctDNA with volumetric analysis of CT images from women with recurrent HGSOC showed that ctDNA was strongly correlated with the burden of abdominal disease, as well as the presence of ascites [146]. In addition, a fall in ctDNA after the first cycle of chemotherapy was an independent predictor of time to progression based on multivariable analysis (hazard ratio 0.2, 95% CI 0.07–0.67, $P = 0.008$). Sequential ctDNA measurements during treatment can provide a highly specific, early response marker in HGSOC, and there has been rapid uptake of ctDNA assays being included as tertiary endpoints in multiple international phase 2 trials of ovarian cancer.

Summary

The emerging molecular data indicate that ovarian carcinoma is a group of distinct diseases, each with different clinical and epidemiological characteristics. Previous work to identify predictive and prognostic biomarkers has been greatly weakened by the inclusion of different ovarian histotypes. Platinum resistance in HGSOC may represent a distinct molecular subtype, but further work is required to define the driver mutations in this disease and the mechanisms of intrinsic resistance. The use of PARP inhibitors in platinum-sensitive HGSOC heralds the real possibility of curative strategies based on maintenance treatment. However, understanding how platinum sensitivity may be lost in HGSOC will require genomic comparison of sequential tissues taken at diagnosis and at time of relapse. Stratified medicine for ovarian carcinoma, particularly for platinum-sensitive disease, is now feasible but will require much wider access to *BRCA1* and *BRCA2* testing for women with HGSOC and improved genomic tests of HRD in tumour tissue.

References

1. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the international cancer benchmarking partnership): an analysis of population-based cancer registry data. Lancet. 2011;377(9760):127–38.
2. Bookman MA, Brady MF, McGuire WP, Harper PG, Alberts DS, Friedlander M, et al. Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a phase III trial of the gynecologic cancer intergroup. J Clin Oncol. 2009;27(9):1419–25.
3. du Bois A, Weber B, Rochon J, Meier W, Goupil A, Olbricht S, et al. Addition of epirubicin as a third drug to carboplatin-paclitaxel in first-line treatment of advanced ovarian cancer: a prospectively randomized gynecologic cancer intergroup trial by the Arbeitsgemeinschaft Gynaekologische Onkologie Ovarian Cancer Study Group and the Groupe d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens. J Clin Oncol. 2006;24(7):1127–35.
4. du Bois A, Herrstedt J, Hardy-Bessard AC, Muller HH, Harter P, Kristensen G, et al. Phase III trial of carboplatin plus paclitaxel with or without gemcitabine in first-line treatment of epithelial ovarian cancer. J Clin Oncol. 2010;28(27):4162–9.
5. Dubeau L. The cell of origin of ovarian epithelial tumors and the ovarian surface epithelium dogma: does the emperor have no clothes? Gynecol Oncol. 1999;72(3):437–42.
6. Piek JM, van Diest PJ, Zweemer RP, Kenemans P, Verheijen RH. Tubal ligation and risk of ovarian cancer. Lancet. 2001;358(9284):844.
7. Piek JM, van Diest PJ, Zweemer RP, Jansen JW, Poort-Keesom RJ, Menko FH, et al. Dysplastic changes in prophylactically removed fallopian tubes of women predisposed to developing ovarian cancer. J Pathol. 2001;195(4):451–6.
8. Leeper K, Garcia R, Swisher E, Goff B, Greer B, Paley P. Pathologic findings in prophylactic oophorectomy specimens in high-risk women. Gynecol Oncol. 2002;87(1):52–6.
9. Piek JM, Verheijen RH, Kenemans P, Massuger LF, Bulten H, van Diest PJ. BRCA1/2-related ovarian cancers are of tubal origin: a hypothesis. Gynecol Oncol. 2003;90(2):491.
10. Lee Y, Miron A, Drapkin R, Nucci MR, Medeiros F, Saleemuddin A, et al. A candidate precursor to serous carcinoma that originates in the distal fallopian tube. J Pathol. 2007;211(1):26–35.
11. Crum CP, Drapkin R, Kindelberger D, Medeiros F, Miron A, Lee Y. Lessons from BRCA: the tubal fimbria emerges as an origin for pelvic serous cancer. Clin Med Res. 2007;5(1):35–44.
12. Callahan MJ, Crum CP, Medeiros F, Kindelberger DW, Elvin JA, Garber JE, et al. Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. J Clin Oncol. 2007;25(25):3985–90.
13. Kindelberger DW, Lee Y, Miron A, Hirsch MS, Feltmate C, Medeiros F, et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. Am J Surg Pathol. 2007;31(2):161–9.
14. Vigano P, Somigliana E, Chiodo I, Abbiati A, Vercellini P. Molecular mechanisms and biological plausibility underlying

the malignant transformation of endometriosis: a critical analysis. Hum Reprod Update. 2006;12(1):77–89.
15. Ness RB. Endometriosis and ovarian cancer: thoughts on shared pathophysiology. Am J Obstet Gynecol. 2003;189(1):280–94.
16. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011;11(7):481–92.
17. Jones S, Wang TL, Shih Ie M, Mao TL, Nakayama K, Roden R, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science. 2010;330(6001):228–31.
18. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med. 2010;363(16):1532–43.
19. Yamamoto S, Tsuda H, Takano M, Iwaya K, Tamai S, Matsubara O. PIK3CA mutation is an early event in the development of endometriosis-associated ovarian clear cell adenocarcinoma. J Pathol. 2011;225(2):189–94.
20. Anglesio MS, Papadopoulos N, Ayhan A, Nazeran TM, Noe M, Horlings HM, et al. Cancer-associated mutations in endometriosis without cancer. N Engl J Med. 2017;376(19):1835–48.
21. Kobel M, Rahimi K, Rambau PF, Naugler C, Le Page C, Meunier L, et al. An immunohistochemical algorithm for ovarian carcinoma typing. Int J Gynecol Pathol. 2016;35(5):430–41.
22. Kobel M, Bak J, Bertelsen BI, Carpen O, Grove A, Hansen ES, et al. Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. Histopathology. 2014;64(7):1004–13.
23. Kobel M, Piskorz AM, Lee S, Lui S, LePage C, Marass F, et al. Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. J Pathol Clin Res. 2016;2(4):247–58.
24. Altman AD, Nelson GS, Ghatage P, McIntyre JB, Capper D, Chu P, et al. The diagnostic utility of TP53 and CDKN2A to distinguish ovarian high-grade serous carcinoma from low-grade serous ovarian tumors. Mod Pathol. 2013;26(9):1255–63.
25. McIntyre JB, Rambau PF, Chan A, Yap S, Morris D, Nelson GS, et al. Molecular alterations in indolent, aggressive and recurrent ovarian low-grade serous carcinoma. Histopathology. 2017;70(3):347–58.
26. Casey L, Kobel M, Ganesan R, Tam S, Prasad R, Bohm S, et al. A comparison of p53 and WT1 immunohistochemical expression patterns in tubo-ovarian high-grade serous carcinoma before and after neoadjuvant chemotherapy. Histopathology. 2017;71(5):736–42.
27. Lee S, Piskorz AM, Le Page C, Mes Masson AM, Provencher D, Huntsman D, et al. Calibration and optimization of p53, WT1, and napsin a immunohistochemistry ancillary tests for histotyping of ovarian carcinoma: Canadian immunohistochemistry quality control (CIQC) experience. Int J Gynecol Pathol. 2016;35(3):209–21.
28. Bromley AB, Altman AD, Chu P, Nation JG, Nelson GS, Ghatage P, et al. Architectural patterns of ovarian/pelvic high-grade serous carcinoma. Int J Gynecol Pathol. 2012;31(5):397–404.
29. Assem H, Rambau PF, Lee S, Ogilvie T, Sienko A, Kelemen LE, et al. High-grade endometrioid carcinoma of the ovary: a clinicopathologic study of 30 cases. Am J Surg Pathol. 2018;42(4):534–44.
30. Chen W, Husain A, Nelson GS, Rambau PF, Liu S, Lee CH, et al. Immunohistochemical profiling of endometrial serous carcinoma. Int J Gynecol Pathol. 2017;36(2):128–39.
31. Kommoss F, Faruqi A, Gilks CB, Lamshang Leen S, Singh N, Wilkinson N, et al. Uterine serous carcinomas frequently metastasize to the fallopian tube and can mimic serous tubal intraepithelial carcinoma. Am J Surg Pathol. 2017;41(2):161–70.
32. Mackenzie R, Talhouk A, Eshragh S, Lau S, Cheung D, Chow C, et al. Morphologic and molecular characteristics of mixed epithelial ovarian Can cers. Am J Surg Pathol. 2015;39(11):1548–57.
33. McAlpine JN, Porter H, Kobel M, Nelson BH, Prentice LM, Kalloger SE, et al. BRCA1 and BRCA2 mutations correlate with TP53 abnormalities and presence of immune cell infiltrates in ovarian high-grade serous carcinoma. Mod Pathol. 2012;25(5):740–50.
34. Spentzos D, Levine DA, Ramoni MF, Joseph M, Gu X, Boyd J, et al. Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol. 2004;22(23):4700–10.
35. Selvanayagam ZE, Cheung TH, Wei N, Vittal R, Lo KW, Yeo W, et al. Prediction of chemotherapeutic response in ovarian cancer with DNA microarray expression profiling. Cancer Genet Cytogenet. 2004;154(1):63–6.
36. Spentzos D, Levine DA, Kolia S, Otu H, Boyd J, Libermann TA, et al. Unique gene expression profile based on pathologic response in epithelial ovarian cancer. J Clin Oncol. 2005;23(31):7911–8.
37. Berchuck A, Iversen ES, Lancaster JM, Pittman J, Luo J, Lee P, et al. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. Clin Cancer Res. 2005;11(10):3686–96.
38. Hartmann LC, Lu KH, Linette GP, Cliby WA, Kalli KR, Gershenson D, et al. Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy. Clin Cancer Res. 2005;11(6):2149–55.
39. Bernardini M, Lee CH, Beheshti B, Prasad M, Albert M, Marrano P, et al. High-resolution mapping of genomic imbalance and identification of gene expression profiles associated with differential chemotherapy response in serous epithelial ovarian cancer. Neoplasia. 2005;7(6):603–13.
40. Jazaeri AA, Awtrey CS, Chandramouli GV, Chuang YE, Khan J, Sotiriou C, et al. Gene expression profiles associated with response to chemotherapy in epithelial ovarian cancers. Clin Cancer Res. 2005;11(17):6300–10.
41. Helleman J, Jansen MP, Span PN, van Staveren IL, Massuger LF, Meijer-van Gelder ME, et al. Molecular profiling of platinum resistant ovarian cancer. Int J Cancer. 2006;118(8):1963–71.
42. Dressman HK, Berchuck A, Chan G, Zhai J, Bild A, Sayer R, et al. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. J Clin Oncol. 2007;25(5):517–25.
43. Crijns AP, Fehrmann RS, de Jong S, Gerbens F, Meersma GJ, Klip HG, et al. Survival-related profile, pathways, and transcription factors in ovarian cancer. PLoS Med. 2009;6(2):e24.
44. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res. 2008;14(16):5198–208.
45. Fehrmann RS, Li XY, van der Zee AG, de Jong S, Te Meerman GJ, de Vries EG, et al. Profiling studies in ovarian cancer: a review. Oncologist. 2007;12(8):960–6.
46. Agarwal R, Kaye SB. Expression profiling and individualisation of treatment for ovarian cancer. Curr Opin Pharmacol. 2006;6(4):345–9.
47. Na YJ, Farley J, Zeh A, del Carmen M, Penson R, Birrer MJ. Ovarian cancer: markers of response. Int J Gynecol Cancer. 2009;19(Suppl 2):S21–9.
48. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869–74.
49. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A. 2003;100(18):10393–8.
50. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.

51. Sieh W, Kobel M, Longacre TA, Bowtell DD, deFazio A, Goodman MT, et al. Hormone-receptor expression and ovarian cancer survival: an ovarian tumor tissue analysis consortium study. Lancet Oncol. 2013;14(9):853–62.
52. Andersen CL, Sikora MJ, Boisen MM, Ma T, Christie A, Tseng G, et al. Active estrogen receptor-alpha signaling in ovarian cancer models and clinical specimens. Clin Cancer Res. 2017;23(14):3802–12.
53. Verhaak RG, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J Clin Invest. 2013;123(1):517–25.
54. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. J Pathol. 2014;232(2):199–209.
55. Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, et al. Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. Immunity. 2016;44(3):698–711.
56. Mlecnik B, Bindea G, Kirilovsky A, Angell HK, Obenauf AC, Tosolini M, et al. The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. Sci Transl Med. 2016;8(327):327ra26.
57. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med. 2003;348(3):203–13.
58. Nielsen JS, Sahota RA, Milne K, Kost SE, Nesslinger NJ, Watson PH, et al. CD20+ tumor-infiltrating lymphocytes have an atypical CD27- memory phenotype and together with CD8+ T cells promote favorable prognosis in ovarian cancer. Clin Cancer Res. 2012;18(12):3281–92.
59. Webb JR, Milne K, Watson P, Deleeuw RJ, Nelson BH. Tumor-infiltrating lymphocytes expressing the tissue resident memory marker CD103 are associated with increased survival in high-grade serous ovarian cancer. Clin Cancer Res. 2014;20(2):434–44.
60. Kroeger DR, Milne K, Nelson BH. Tumor-infiltrating plasma cells are associated with tertiary lymphoid structures, cytolytic T-cell responses, and superior prognosis in ovarian cancer. Clin Cancer Res. 2016;22(12):3005–15.
61. Clarke B, Tinker AV, Lee CH, Subramanian S, van de Rijn M, Turbin D, et al. Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. Mod Pathol. 2009;22(3):393–402.
62. Motz GT, Santoro SP, Wang LP, Garrabrant T, Lastra RR, Hagemann IS, et al. Tumor endothelium FasL establishes a selective immune barrier promoting tolerance in tumors. Nat Med. 2014;20(6):607–15.
63. Lan C, Heindl A, Huang X, Xi S, Banerjee S, Liu J, et al. Quantitative histology analysis of the ovarian tumour microenvironment. Sci Rep. 2015;5:16317.
64. Ovarian Tumor Tissue Analysis C, Goode EL, Block MS, Kalli KR, Vierkant RA, Chen W, et al. Dose-response association of CD8+ tumor-infiltrating lymphocytes and survival time in high-grade serous ovarian cancer. JAMA Oncol. 2017;3:e173290.
65. Parkes EE, Walker SM, Taggart LE, McCabe N, Knight LA, Wilkinson R, et al. Activation of STING-dependent innate immune signaling by S-phase-specific DNA damage in breast cancer. J Natl Cancer Inst 2017;109(1). https://doi.org/10.1093/jnci/djw199.
66. Yang JY, Yoshihara K, Tanaka K, Hatae M, Masuzaki H, Itamochi H, et al. Predicting time to ovarian carcinoma recurrence using protein markers. J Clin Invest. 2013;123(9):3740–50.
67. Nigro JM, Baker SJ, Preisinger AC, Jessup JM, Hostetter R, Cleary K, et al. Mutations in the p53 gene occur in diverse human tumour types. Nature. 1989;342(6250):705–8.
68. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004;10(8):789–99.
69. de Graeff P, Crijns AP, de Jong S, Boezen M, Post WJ, de Vries EG, et al. Modest effect of p53, EGFR and HER-2/neu on prognosis in epithelial ovarian cancer: a meta-analysis. Br J Cancer. 2009;101(1):149–59.
70. Stewart RL, Royds JA, Burton JL, Heatley MK, Wells M. Direct sequencing of the p53 gene shows absence of mutations in endometrioid endometrial adenocarcinomas expressing p53 protein. Histopathology. 1998;33(5):440–5.
71. Singer G, Stohr R, Cope L, Dehari R, Hartmann A, Cao DF, et al. Patterns of p53 mutations separate ovarian serous borderline tumors and low- and high-grade carcinomas and provide support for a new model of ovarian carcinogenesis: a mutational analysis with immunohistochemical correlation. Am J Surg Pathol. 2005;29(2):218–24.
72. Kuo KT, Mao TL, Jones S, Veras E, Ayhan A, Wang TL, et al. Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma. Am J Pathol. 2009;174(5):1597–601.
73. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol. 2010;221(1):49–56.
74. Brown CJ, Lain S, Verma CS, Fersht AR, Lane DP. Awakening guardian angels: drugging the p53 pathway. Nat Rev Cancer. 2009;9(12):862–73.
75. Cheok CF, Verma CS, Je B, Lane DP. Translating p53 into the clinic. Nat Rev Clin Oncol. 2011;8(1):25–37.
76. Carrassa L, Chila R, Lupi M, Ricci F, Celenza C, Mazzoletti M, et al. Combined inhibition of Chk1 and Wee1: in vitro synergistic effect translates to tumor growth inhibition in vivo. Cell Cycle. 2012;11(13):2507–17.
77. Emerling BM, Hurov JB, Poulogiannis G, Tsukazawa KS, Choo-Wing R, Wulf GM, et al. Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-null tumors. Cell. 2013;155(4):844–57.
78. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45(10):1127–33.
79. Nakayama K, Nakayama N, Jinawath N, Salani R, Kurman RJ, Shih Ie M, et al. Amplicon profiles in ovarian serous carcinomas. Int J Cancer. 2007;120(12):2613–7.
80. Farley J, Smith LM, Darcy KM, Sobel E, O'Connor D, Henderson B, et al. Cyclin E expression is a significant predictor of survival in advanced, suboptimally debulked ovarian epithelial cancers: a gynecologic oncology group study. Cancer Res. 2003;63(6):1235–41.
81. Etemadmoghadam D, deFazio A, Beroukhim R, Mermel C, George J, Getz G, et al. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. Clin Cancer Res. 2009;15(4):1417–27.
82. Etemadmoghadam D, George J, Cowin PA, Cullinane C, Kansara M, Group AOCS, et al. Amplicon-dependent CCNE1 expression is critical for clonogenic survival after cisplatin treatment and is correlated with 20q11 gain in ovarian cancer. PLoS One. 2010;5(11):e15498.
83. Theurillat JP, Metzler SC, Henzi N, Djouder N, Helbling M, Zimmermann AK, et al. URI is an oncogene amplified in ovarian cancer cells and is required for their survival. Cancer Cell. 2011;19(3):317–32.
84. Etemadmoghadam D, Weir BA, Au-Yeung G, Alsop K, Mitchell G, George J, et al. Synthetic lethality between CCNE1 amplification and loss of BRCA1. Proc Natl Acad Sci U S A. 2013;110(48):19489–94.
85. Etemadmoghadam D, Au-Yeung G, Wall M, Mitchell C, Kansara M, Loehrer E, et al. Resistance to CDK2 inhibitors is associated with selection of polyploid cells in CCNE1-amplified ovarian cancer. Clin Cancer Res. 2013;19(21):5960–71.

86. Markman M, Rothman R, Hakes T, Reichman B, Hoskins W, Rubin S, et al. Second-line platinum therapy in patients with ovarian cancer previously treated with cisplatin. J Clin Oncol. 1991;9(3):389–93.
87. Borst P, Rottenberg S, Jonkers J. How do real tumors become resistant to cisplatin? Cell Cycle. 2008;7(10):1353–9.
88. Trainer AH, Meiser B, Watts K, Mitchell G, Tucker K, Friedlander M. Moving toward personalized medicine: treatment-focused genetic testing of women newly diagnosed with ovarian cancer. Int J Gynecol Cancer. 2010;20(5):704–16.
89. Hennessy BTJ, Timms KM, Carey MS, Gutin A, Meyer LA, Flake n DD, et al. Somatic mutations in BRCA1 and BRCA2 could expand the number of patients that benefit from poly (ADP ribose) polymerase inhibitors in ovarian cancer. J Clin Oncol. 2010;28(22):3570–6.
90. Alsop K, Fereday S, Meldrum C, deFazio A, Emmanuel C, George J, et al. BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian ovarian cancer study group. J Clin Oncol. 2012;30(21):2654–63.
91. Chetrit A, Hirsh-Yechezkel G, Ben-David Y, Lubin F, Friedman E, Sadetzki S. Effect of BRCA1/2 mutations on long-term survival of patients with invasive ovarian cancer: the national Israeli study of ovarian cancer. J Clin Oncol. 2008;26(1):20–5.
92. Tan DS, Rothermundt C, Thomas K, Bancroft E, Eeles R, Shanley S, et al. "BRCAness" syndrome in ovarian cancer: a case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations. J Clin Oncol. 2008;26(34):5530–6.
93. Yang D, Khan S, Sun Y, Hess K, Shmulevich I, Sood AK, et al. Association of BRCA1 and BRCA2 mutations with survival, chemotherapy sensitivity, and gene mutator phenotype in patients with ovarian cancer. JAMA. 2011;306(14):1557–65.
94. Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA. 2012;307(4):382–90.
95. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010;42(5):410–4.
96. Pelttari LM, Heikkinen T, Thompson D, Kallioniemi A, Schleutker J, Holli K, et al. RAD51C is a susceptibility gene for ovarian cancer. Hum Mol Genet. 2011;20(16):3278–88.
97. Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet. 2011;43(9):879–82.
98. Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, et al. Mutations in BRIP1 confer high risk of ovarian cancer. Nat Genet. 2011;43(11):1104–7.
99. Mukhopadhyay A, Elattar A, Cerbinskaite A, Wilkinson SJ, Drew Y, Kyle S, et al. Development of a functional assay for homologous recombination status in primary cultures of epithelial ovarian tumor and correlation with sensitivity to poly(ADP-ribose) polymerase inhibitors. Clin Cancer Res. 2010;16(8):2344–51.
100. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer. 2012;107(10):1776–82.
101. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. Clin Cancer Res. 2016;22(15):3764–73.
102. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415–21.
103. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517–25.
104. Norquist B, Wurz KA, Pennil CC, Garcia R, Gross J, Sakai W, et al. Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. J Clin Oncol. 2011;29(22):3008–15.
105. Sakai W, Swisher EM, Jacquemont C, Chandramohan KV, Couch FJ, Langdon SP, et al. Functional restoration of BRCA2 protein by secondary BRCA2 mutations in BRCA2-mutated ovarian carcinoma. Cancer Res. 2009;69(16):6381–6.
106. Swisher EM, Sakai W, Karlan BY, Wurz K, Urban N, Taniguchi T. Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance. Cancer Res. 2008;68(8):2581–6.
107. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, et al. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature. 2008;451(7182):1116–20.
108. Edwards SL, Brough R, Lord CJ, Natrajan R, Vatcheva R, Levine DA, et al. Resistance to therapy caused by intragenic deletion in BRCA2. Nature. 2008;451(7182):1111–5.
109. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015;521(7553):489–94.
110. Cooke SL, Temple J, Macarthur S, Zahra MA, Tan LT, Crawford RA, et al. Intra-tumour genetic heterogeneity and poor chemoradiotherapy response in cervical cancer. Br J Cancer. 2011;104(2):361–8.
111. Cooke SL, Ng CK, Melnyk N, Garcia MJ, Hardcastle T, Temple J, et al. Genomic analysis of genetic heterogeneity and evolution in high-grade serous ovarian carcinoma. Oncogene. 2010;29(35):4905–13.
112. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23–8.
113. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer Res. 1978;38(10):3174–81.
114. Khalique L, Ayhan A, Weale ME, Jacobs IJ, Ramus SJ, Gayther SA. Genetic intra-tumour heterogeneity in epithelial ovarian cancer and its implications for molecular diagnosis of tumours. J Pathol. 2007;211(3):286–95.
115. Khalique L, Ayhan A, Whittaker JC, Singh N, Jacobs IJ, Gayther SA, et al. The clonal evolution of metastases from primary serous epithelial ovarian cancers. Int J Cancer. 2009;124(7):1579–86.
116. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90–4.
117. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, et al. Inferring tumor progression from genomic heterogeneity. Genome Res. 2010;20(1):68–80.
118. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009;461(7265):809–13.
119. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. Nature. 2012;482(7386):529–33.
120. Vermaat JS, Nijman IJ, Koudijs MJ, Gerritse FL, Scherer SJ, Mokry M, et al. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clin Cancer Res. 2012;18(3):688–99.
121. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012;30(5):413–21.
122. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395–9.

123. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010;467(7319):1109–13.
124. Choi S, Henderson MJ, Kwan E, Beesley AH, Sutton R, Bahar AY, et al. Relapse in children with acute lymphoblastic leukemia involving selection of a preexisting drug-resistant subclone. Blood. 2007;110(2):632–9.
125. Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. Cancer Cell. 2002;2(2):117–25.
126. Roche-Lestienne C, Soenen-Cornu V, Grardel-Duflos N, Laï J-L, Philippe N, Facon T, et al. Several types of mutations of the Abl gene can be found in chronic myeloid leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment. Blood. 2002;100(3):1014–8.
127. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
128. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell. 2007;11(3):259–73.
129. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science. 2008;322(5906):1377–80.
130. Micci F, Haugom L, Ahlquist T, Abeler VM, Trope CG, Lothe RA, et al. Tumor spreading to the contralateral ovary in bilateral ovarian carcinoma is a late event in clonal evolution. J Oncol. 2010;2010:646340.
131. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol. 2013;231(1):21–34.
132. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet. 2016;48(7):758–67.
133. Schwarz RF, Ng CK, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015;12(2):e1001789.
134. McAlpine JN, Eisenkop SM, Spirtos NM. Tumor heterogeneity in ovarian cancer as demonstrated by in vitro chemoresistance assays. Gynecol Oncol. 2008;110(3):360–4.
135. Griffin N, Grant LA, Freeman SJ, Jimenez-Linan M, Berman LH, Earl H, et al. Image-guided biopsy in patients with suspected ovarian carcinoma: a safe and effective technique? Eur Radiol. 2009;19(1):230–5.
136. Spencer JA, Weston MJ, Saidi SA, Wilkinson N, Hall GD. Clinical utility of image-guided peritoneal and omental biopsy. Nat Rev Clin Oncol. 2010;7(11):623–31.
137. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37(3):646–50.
138. Kamat AA, Baldwin M, Urbauer D, Dang D, Han LY, Godwin A, et al. Plasma cell-free DNA in ovarian cancer: an independent prognostic biomarker. Cancer. 2010;116(8):1918–25.
139. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008;14(9):985–90.
140. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med. 2010;2(20):20ra14.
141. Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. Clin Cancer Res. 2009;15(6):2076–84.
142. Chen Z, Feng J, Buzin CH, Liu Q, Weiss L, Kernstine K, et al. Analysis of cancer mutation signatures in blood by a novel ultrasensitive assay: monitoring of therapy or recurrence in non-metastatic breast cancer. PLoS One. 2009;4(9):e7220.
143. Poveda A, Kaye SB, McCormack R, Wang S, Parekh T, Ricci D, et al. Circulating tumor cells predict progression free survival and overall survival in patients with relapsed/recurrent advanced ovarian cancer. Gynecol Oncol. 2011;122(3):567–72.
144. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012;4(136):136ra68.
145. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013;497(7447):108–12.
146. Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, et al. Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. PLoS Med. 2016;13(12):e1002198.

Genomic Applications in Soft Tissue Sarcomas

33

Eva Wardelmann and Wolfgang Hartmann

Introduction

Sarcomas account for approximately 1% of all malignancies. The 2013 WHO classification [59] recognizes over 70 types of soft tissue neoplasms. During the last few years, genomic alterations which are of diagnostic, prognostic, and/or predictive value have been detected in sarcomas. As a result, the development of molecular methods to subclassify sarcoma subtypes became warranted because molecular signatures may pinpoint potential areas of interest for diagnostic tools, prediction of clinical outcomes, and potential response to therapeutic targets.

In general, sarcomas can be subdivided into two different morphological subgroups: tumors with a non-pleomorphic morphology and those with a pleomorphic phenotype. The non-pleomorphic sarcomas more often carry specific molecular aberrations, whereas pleomorphic sarcomas frequently have a complex karyotype. Sarcomas with such complex karyotypes account for approximately 50% of all soft tissue sarcomas.

Three major types of genomic alterations occur in sarcomas: reciprocal translocations (~15%), specific mutations (~25%), and amplifications (~10%). The chromosomal translocations most frequently lead to the formation of chimeric fusion genes the protein products of which function either as transcription factors, autocrine growth factors, or tyrosine kinases. Specific mutations are found preferentially in genes encoding tyrosine kinases. Amplifications mainly affect genes which encode important players in cell-cycle control.

This chapter presents an overview of the genomic applications that currently play an increasingly important role in the diagnostic and treatment decision algorithms of soft tissue sarcomas. Soft tissue tumors are classified in Table 33.1.

Table 33.1 Chapter content with soft tissue tumor classification

Adipocytic tumors
Atypical lipomatous tumor/well-differentiated liposarcoma; dedifferentiated liposarcoma
Myxoid liposarcoma
Pleomorphic liposarcoma
Fibroblastic/myofibroblastic tumors
Desmoid-type fibromatosis
Giant cell fibroblastoma
Dermatofibrosarcoma protuberans
Solitary fibrous tumor
Inflammatory myofibroblastic tumor
Infantile fibrosarcoma
Low-grade fibromyxoid sarcoma
Sclerosing epithelioid fibrosarcoma
So-called fibrohistiocytic tumors
Tenosynovial giant cell tumor, localized/diffuse type
Skeletal muscle tumors
Embryonal rhabdomyosarcoma
Alveolar rhabdomyosarcoma
Vascular tumors
Epithelioid hemangioendothelioma
Angiosarcoma
Tumors of uncertain differentiation
Synovial sarcoma
Epithelioid sarcoma
Alveolar soft part sarcoma
Clear cell sarcoma of soft tissue
Extraskeletal myxoid chondrosarcoma
Desmoplastic small round cell tumor
Extrarenal rhabdoid tumor
Gastrointestinal stromal tumors (GIST)

Table 33.2 depicts the various genomic aberrations that have been well recognized in sarcoma subtypes. Subtypes for which such conclusive molecular data are still lacking (e.g., smooth muscle, pericytic, and nerve sheath tumors) are not listed. The vast majority of benign soft tissue tumors are also excluded, whereas some neoplasms with intermediate biologic behavior are included.

E. Wardelmann (✉) · W. Hartmann (✉)
Gerhard-Domagk-Institute of Pathology, University Hospital Münster, Münster, Germany
e-mail: eva.wardelmann@ukmuenster.de; wolfgang.hartmann@ukmuenster.de

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_33

Table 33.2 Genomic aberrations in soft tissue tumors

Tumor entity	Genomic aberration	Fusion gene, mutated gene
Alveolar rhabdomyosarcoma (ARMS)	t(2;13)(q35;q14) t(1;13)(p36;q14) t(2;2)(p23;q36) t(X;2)(q13;q36)	*PAX3-FOXO1A* *PAX7-FOXO1A* *PAX3-NCOA1* *PAX3-FOXO4*
Alveolar soft part sarcoma (ASPS)	t(X;17)(p11;q25)	*ASPSCR1-TFE3*
Angiomatoid fibrous histiocytoma (AFH)	t(12;16)(q13;p11) t(2;22)(q33;q12) t(12;22)(q13;q12)	*TLS-ATF1* *EWSR1-CREB1* *EWSR1-ATF 1*
Angiosarcoma (ASA)	Missense mutation Amplification	*KDR*, *FLT4* *c-MYC*
Clear-cell sarcoma (CCS)	t(12;22)(q13;q12) t(2;22)(q33;q12)	*EWSR1-ATF1* *EWSR1-CREB1*
Congenital fibrosarcoma (CGFS)	t(12;15)(p13;q25)	*ETV6-NTRK3*
Dermatofibrosarcoma protuberans (DFSP)	t(17;22)(q22;q13) der(22)t(17;22) Ring chromosome	*COL1A1-PDGFB*
Desmoplastic small round cell tumor (DSRCT)	t(11;22)(p13;q12)	*EWSR1-WT1*
Endometrial stromal sarcoma (ESS)	t(7;17)(p15;q21) t(10;17)(q22;p13)	*JAZF1-JJAZ1* *YWHAE-FAM22A/B*
Epithelioid hemangioendothelioma (EHE)	t(1;3(p36.3;q25) t(X;11)	*WWTR1-CAMTA1* *YAP1-TFE3*
Epithelioid sarcoma (EWS)	Intragenic deletions	*SMARCB1/INI1*
Ewing sarcoma (ES)*	t(11;22)(q24;q12) t(21;22)(q22;q12) t(7;22)(p22;q12) t(17;22)(q12;q12) t(2;22)(q33;q12) t(16;21)(p11;q22	*EWSR1-FLI1* *EWSR1-ERG* *EWSR1-ETV1* *EWSR1-E1AF* *EWSR1-FEV* *FUS-ERG*
Extrarenal rhabdoid tumor (ERT)	Homozygous inactivation by deletion	*hSNF/INI1/SMARCB1/BAF47* or *SMARCA4 (BRG1)* loss
Extraskeletal myxoid chondrosarcoma (EMCS)	t(9;22)(q22;q12) t(9;17)(q22;q11) t(3;9)(q12;q22) t(9;17)(q22;q11)	*EWSR1-NR4A3* *TAF2N-NR4A3* *TFG-NR4A3* *TCF12-NR4A3*
Fibromatosis (desmoid type)	CTNNB1 mutations, APC mutations	Missense mutations
Gastrointestinal stromal tumor (GIST)	Mutations	*KIT*, *PDGFRA*, *SDHA-D*, *NF1*, *BRAF, MAX, MEN1*
Inflammatory myofibroblastic tumor (IMFT)	t(1;2)(q23;q23) t(2;19)(q23;q13) t(2;17)(q23;q23) t(2;2)(p23;q13) t(2;11)(p23;p15) inv2(2)(p21;p23) t(6;3)(q22;q12) t(6;17)(q22;p13) Mutations	*TPM3-ALK* *TPM4-ALK* *CLTC-ALK* *RANBP2-ALK* *CARS-ALK* *EML4-ALK* and other partners *TFG-ROS1* *YWHAE-ROS* *RAS, RET*
Low-grade fibromyxoid sarcoma (LGFS) Sclerosing epithelioid fibrosarcoma (SEF)	t(7;16)(q33–34;p11) t(11;16)(p11;p11)	*FUS-CREB3L2* *FUS-CREB3L1* *EWSR1-CREB3L1*
Myxoid/round-cell liposarcoma Myxoinflammatory fibroblastic sarcoma (MIFS)	t(12;16)(q13;p11) t(12;22)(q13;q12) t(1;10)(p22;q24) Ring chromosome	*FUS-DD1T3* *EWSR1-DD1T3* Deregulation of *FGF8* + *NPM3* Amplification of *VGLL3*
Solitary fibrous tumor/hemangiopericytoma (SFT)	der(12)(q13;q13)	*NAB2-STAT6*

Table 33.2 (continued)

Tumor entity	Genomic aberration	Fusion gene, mutated gene
Synovial sarcoma (SS)	t(X;18)(p11;q11) t(X;18)(p11;q11) t(X;18)(p11;q11) t(X;20)(p11;q13)	*SS18-SSX1* *SS18-SSX2* *SS18-SSX4* *SS181-SSX1*
Tenosynovial giant cell tumor (TGCT)	t(1;2)(p13;q37)	*CSF1-COL6A3*
Well-differentiated liposarcoma (WDLS)/atypical lipomatous tumor (ALT)	Ring chromosome/giant marker	Amplification of *MDM2*, *CDK4*, *HMGA2*, *GLI1*

*Classic Ewing sarcoma excluding round cell sarcomas with alterations of CIC or BCOR

Adipocytic Tumors

Atypical Lipomatous Tumor/Well–Differentiated Liposarcoma and Dedifferentiated Liposarcoma

Clinicopathological Features

Atypical lipomatous tumors (ALT)/well-differentiated liposarcomas (WDLS) constitute approximately 40–45% of all liposarcomas. They arise mainly in middle-aged adults [28, 52] and occur predominantly in the deep soft tissues of the limbs followed in frequency by the retroperitoneum, the paratesticular region, and the mediastinum [28]. ALT/WDLS are locally aggressive but do not carry a potential for metastasis. The distinction between ALT and WDLS is clinical where lesions arising in surgically accessible sites are referred to as ALT and those arising in deeper surgically less amenable sites (and therefore enduring more frequent local relapses) are termed WDLS. Dedifferentiated liposarcomas (DDLS) arise in the same group of patients at comparable sites with a significant predominance in the retroperitoneum. Approximately 90% of DDLS (histologically often containing a variable WDLS component) arise "de novo," i.e., without prior history of ALT/WDLS, whereas 10% develop in recurrences of ALT/WDLS [28]. Consistent with their genetic hallmark (see below), most ALT/WDLS and DDLS immunohistochemically show positive nuclear staining with antibodies against MDM2 and CDK4 [28, 184].

Genomic Alterations

ALT/WDLS and DDLS are genetically characterized by the presence of supernumerary rings and giant marker chromosomes containing amplified sequences originating from the chromosomal region 12q14–15. The amplicon displays considerable heterogeneity, containing numerous oncogenes [92, 93]. *MDM2* is consistently amplified, acting as an antagonist to p53 by targeting the protein for degradation via its ubiquitin ligase function and through inhibition of its transcriptional activation function [28]. Almost 90% of tumors display co-amplification of *CDK4*, leading to cell-cycle progression via RB phosphorylation. Tumors with *MDM2* amplification lacking co-amplification of *CDK4* have been shown to be associated with a more mature histological phenotype and a better prognosis [91]. DDLS have been reported to be genomically more complex than ALT/WDLS [200].

Prognosis and Treatment

Completely excised ALT/WDLS arising in surgically amenable sites only rarely recur, whereas retroperitoneal, mediastinal, or paratesticular lesions have a higher frequency of local recurrence [28]. ALT/WDLS are associated with a variable risk of dedifferentiation that is related to site of origin. The risk is estimated to be less than 5% in lesions arising in the limbs and higher than 20% in those arising in the retroperitoneum [28]. DDLS recur locally in more than 40% of the cases and lead to distant metastases in up to 20% [80, 125]. Overall mortality is estimated to be 30–40% at 5 years [28]. Whereas complete surgical excision with wide margins represents the treatment of choice, recent trials have documented a favorable progression-free rate in patients with *CDK4*-amplified WDLS/DDLS upon treatment with a small molecule CDK4/CDK6 inhibitor [47], and efforts to target MDM2, thereby activating the p53 pathway, are also ongoing [166].

Myxoid Liposarcoma

Clinicopathological Features

Myxoid liposarcoma (MLS) represents 15–20% of liposarcomas and arises mainly in the deep soft tissues of the extremities, particularly the thigh [104] during the fourth and fifth decades of life. Thirty to forty percent of the patients develop distant metastases, frequently involving other soft tissue sites. Presence of more than 5% tumor cells with round-cell differentiation has been used to define high histological grade and is associated with an unfavorable outcome [9, 104].

Genomic Alterations

MLS is characterized by reciprocal translocations t(12;16)(q13;p11) that result in *FUS-DDIT3* (*CHOP*) gene fusions, which are present in over 95% of cases [31, 162]. In rare instances, alternative t(12;22)(q13;q12) translocations are found resulting in *EWSR1-DDIT3* fusion oncogenes [7, 32, 148]. *FUS* and *EWSR1* encode RNA-binding proteins involved in tran-

scriptional control; DDIT3 binds C/EBP transcription factors through their highly conserved leucine zipper domain and inhibits their function in adipocytic differentiation. *FUS-DDIT3* functions by inhibiting adipogenesis and maintaining immature adipocytes in a continuous cycle of proliferation without differentiation. There is strong evidence to suggest that these translocations are the primary oncogenic event in MLS [50]. Recent large-scale genomic approaches documented activation of the PIK3/AKT signaling cascade in MLS with mutations in the *PIK3CA* gene found in 18% of cases in addition to rare inactivating mutations in the *PTEN* tumor suppressor gene. Importantly, patients whose tumors harbored mutations in *PIK3CA* had a shorter disease-specific survival [13, 45]. Alternative mechanisms leading to PIK3/AKT pathway activation include HGF/MET and RET signaling [137]. Consistent with the finding of a high prevalence of IGF1R expression in MLS [47], Trautmann et al. showed transcriptional regulation of the IGF2 gene through the chimeric FUS-DDIT3 transcriptional regulator leading to an autocrine stimulatory loop [205].

Prognosis and Treatment

Increased round-cell content, presence of necrosis, and alterations of the *TP53* tumor suppressor gene are associated with unfavorable outcome. MLS is associated with an overall favorable 5-year disease-free survival (85%). Variability of the *DDIT3* translocation does not affect prognosis [9]. High expression levels of the IGF1R and IGF2 were shown to correlate with poor metastasis-free survival [25].

Complete excision with wide tumor-free margins is the treatment of choice. MLS without round-cell differentiation is particularly radiosensitive, and patients treated with adjuvant or neoadjuvant radiotherapy achieve 98% 5-year local control. Potential novel treatment agents include Trabectedin, the cytotoxic activity of which is ascribed to binding the minor groove of DNA. Trabectedin efficacy has been shown to be advantageous in myxoid liposarcomas in a phase III trial [44].

Pleomorphic Liposarcoma

Clinicopathological Features

Pleomorphic liposarcomas (PLS) are rare tumors accounting for 5% of sarcomas with adipocytic differentiation. PLS affect patients older than 50 years [11]. Most cases arise in the deep soft tissues of the extremities, the lower extremity being involved more frequently. Thirty to fifty percent of patients develop metastases with lung and pleura representing the preferred sites of metastatic spread.

Genomic Alterations

PLS display complex genomic rearrangements with recurrent losses reported in chromosomal regions 13q14.2, 17q11.2, and 17p13.1, harboring *RB*, *NF1*, and *TP53* tumor suppressor genes are located, respectively [13, 176, 199]. Barretina and colleagues identified a shared genomic signature in PLS and myxofibrosarcoma, indicating a genomic relationship between these two entities which occasionally show transitional morphologic features [16].

Prognosis Treatment

Larger tumor size, central and deep location, and high mitotic activity are associated with a worse prognosis. Overall, 40–50% tumor-associated mortality is reported, and a 5-year survival rate of 60–65% is achieved [67, 86]. Treatment modalities include complete excision with wide tumor-free margins, chemotherapy, and radiotherapy.

Fibroblastic/Myofibroblastic Tumors

Desmoid-Type (Deep) Fibromatosis

Clinicopathological Features

Fibromatoses are myofibroblastic proliferations with infiltrative growth pattern and high recurrence rate that lack metastatic potential. Cases can occur in the context of familial adenomatous polyposis (FAP) coli syndrome where they more often behave aggressively compared to sporadic desmoids, occasionally leading to death. Their overall incidence is 2–4 new cases per 100,000/year [102]. Immunohistochemically, the spindle cells may express smooth muscle actin but not desmin. They are negative for KIT receptor (CD117), DOG1, and CD34 which is essential to differentiate fibromatosis from gastrointestinal stromal tumors (GIST). The most important diagnostic marker is β-catenin which is typically expressed not only in the cytoplasm but also in the nucleus, the latter observation being crucial [201].

Genomic Alterations

At the molecular level, a sporadic mutation in the *CTNNB1* gene which encodes β-catenin is frequently detectable by selective sequencing of exon 3 [87, 88]. Apart from constituting a subunit of the cadherin protein complex, β-catenin acts as an intracellular signal transducer in the WNT signaling pathway. In the case of a mutation in exon 3, β-catenin cannot be degraded and is translocated into the nucleus where it accumulates and acts as a transcriptional co-activator [201] (Fig. 33.1). Alternatively, germline mutations in the *APC* (adenomatous polyposis coli) gene can lead to a nuclear accumulation of β-catenin [117, 140] and thus to fibromatosis. Crago and colleagues showed that desmoid fibromatoses wild-type for *CTNNB1* may rarely harbor chromosome 6 loss or *BMI1* mutations, the latter being a negative regulator of WNT-inhibitory DKK factors [30].

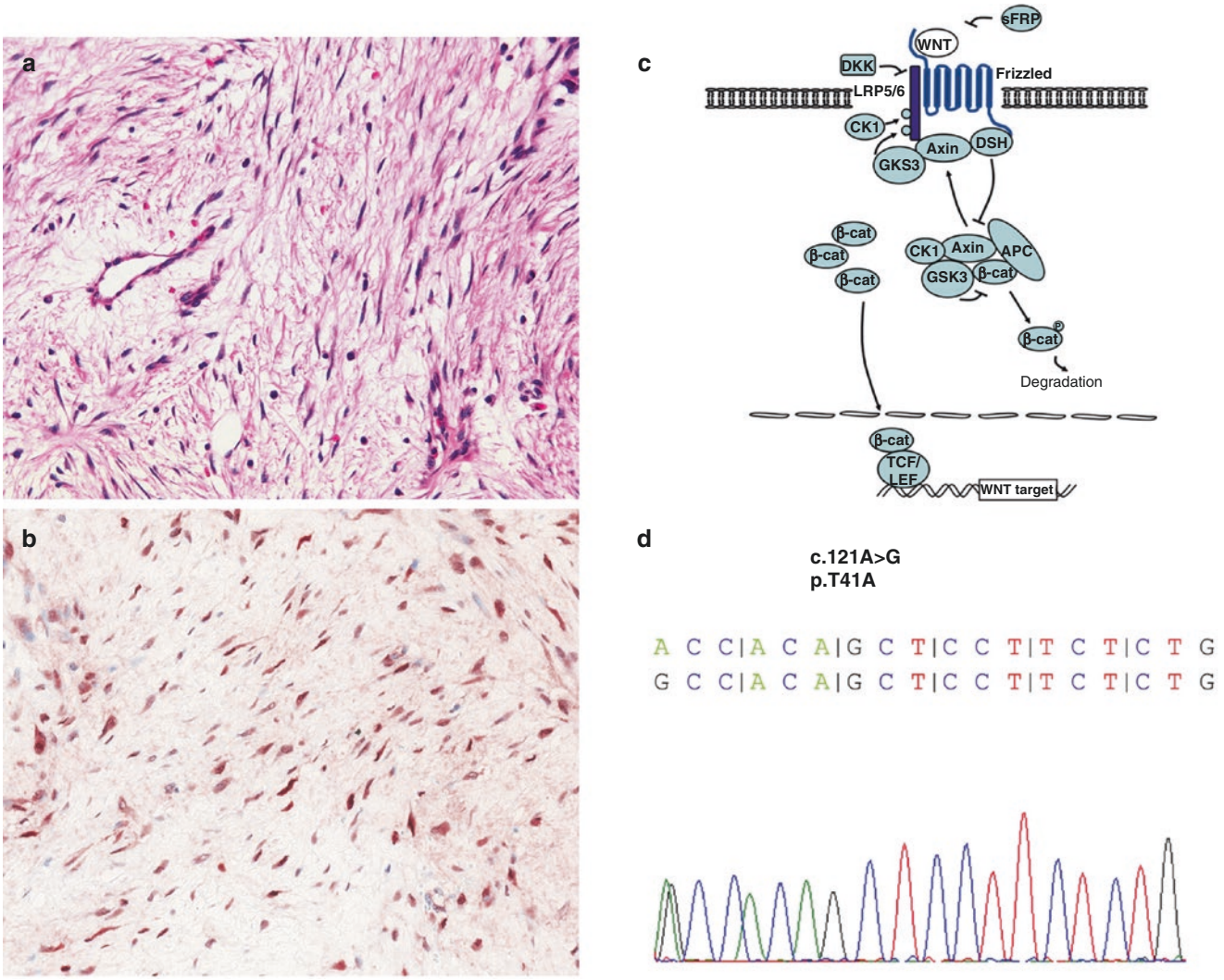


Fig. 33.1 Sporadic mutation in the *CTNNB1* gene. (**a**) Characteristic aspect of a case of desmoid-type fibromatosis composed of a relatively monomorphic spindle cell proliferation of variable density displaying (**b**) nuclear accumulation of β-catenin detectable by immunohistochemistry. (**c**) Schematic view of the WNT signaling pathway with the central effector β-catenin being subject to degradation in the OFF state of the pathway and nuclear transfer and transcriptional activation in the ON state due to WNT signals, activating mutations of the *CTNNB1* gene or inactivating mutations of components of the degradation complex. (**d**) Heterozygous point mutation affecting codon 41 of the *CTNNB1* gene encoding β-catenin leading to nuclear accumulation. (Adapted from Heinrich et al. [78])

The risk for patients with FAP to develop a fibromatosis is 2.56/1000 persons per year and increases with repeated surgical procedures.

Prognosis and Treatment

The prognosis of fibromatosis is not predictable by morphology or molecular markers. As yet, there is no evidence that the molecular subtype influences the outcome [29, 87, 88, 102, 209]. Spontaneous regression is observed in a subgroup of patients, whereas occasional tumors may culminate in death following multiple recurrences. Treatment options in fibromatoses include simple surgical resection, chemotherapy, antihormonal treatment, tyrosine kinase inhibition, and/or radiation [102]. Currently, no reliable molecular biomarker has been shown to have a role in guiding treatment strategy.

Giant-Cell Fibroblastoma/ Dermatofibrosarcoma Protuberans

Clinicopathological Findings

Giant cell fibroblastoma (GCF) is now regarded as the juvenile form of dermatofibrosarcoma protuberans (DFSP) because both dermal sarcomas carry the same translocation [70, 97, 202]. GCF occurs mainly in childhood in the age

group below 10 years and belongs to the group of fibroblastic tumors with intermediate malignant potential (locally aggressive but not metastatic). The majority of patients are male. Typically, these tumors occur in the dermis and subcutis of the trunk or extremities and less often in the head and neck region. DFSP occurs in the third or fourth decades of life with slight male preponderance. It is characterized by its locally infiltrative nature and high tendency to recur (in up to 30% of cases).

Immunohistochemically, CD34 is the most reliable marker, although not specific. Typically, it is strongly and diffusely positive. In the rare cases of fibrosarcomatous transformation in DFSP, CD34 expression may be lost. GCF and DFSP are negative for factor XIIIa but may express other histiocytic markers such as CD68, lysozyme, and CD10.

Genomic Alterations

Both GCF and DFSP are characterized by a specific reciprocal translocation t(17;22)(q22;q13) or more often as a supernumerary ring chromosome involving sequences of both chromosomes 17 and 22. These rearrangements lead to the fusion of the collagen 1A1 gene (CO*L1A1*; alpha-chain type 1 of collagen gene) and the platelet-derived growth factor B gene (*PDGFB*) (Fig. 33.2). As a result tumor cells produce high amounts of PDGFB leading to the constitutive activation of the PDGFB receptor, a type III receptor tyrosine kinase. Assays used for the analysis of the translocation include multiplex reverse transcription polymerase chain reaction and FISH assays, with *COL1A1-PDGFB* dual-color dual fusion or *PDGFB* dual-color break-apart probes showing reliable results [156, 212].

Prognosis and Treatment

Both GCF and DFSP have a high risk for recurrence. Risk for metastasis is limited to those DFSP cases exhibiting fibrosarcomatous transformation. Treatment of advanced, inoperable, or recurrent DFSP lesions is now based on targeting tyrosine kinase inhibition through PDGFR inhibition as reviewed in Llombart et al. [118]. The majority of patients show partial or even complete response to imatinib treatment with minimal toxicity [124]. As a result, imatinib is now considered the gold standard for patients with locally advanced or metastatic DFSP [169]. This therapy is also used in the neoadjuvant setting in order to reduce tumor size and to decrease morbidity prior to a surgical excision.

Solitary Fibrous Tumor

Clinicopathological Findings

Solitary fibrous tumor (SFT) is a fibroblastic neoplasm characterized by a typical vascular pattern frequently referred to as hemangiopericytic. SFT was initially described in the pleura but is now recognized to occur in nearly any location. In less than 5% of patients, SFT manifests with hypoglycemia that is thought to be mediated by insulin-like growth factors (IGFs) produced by the tumor [68]. Hypoglycemia disappears following tumor resection. SFT occurs in adults of all age groups and has equal incidence in both sexes. Rarely, an abrupt transition from SFT to high-grade sarcoma can be seen.

Genomic Alterations

Recently, a recurrent translocation has been identified in SFT. It is the fusion of two neighboring, partly overlapping

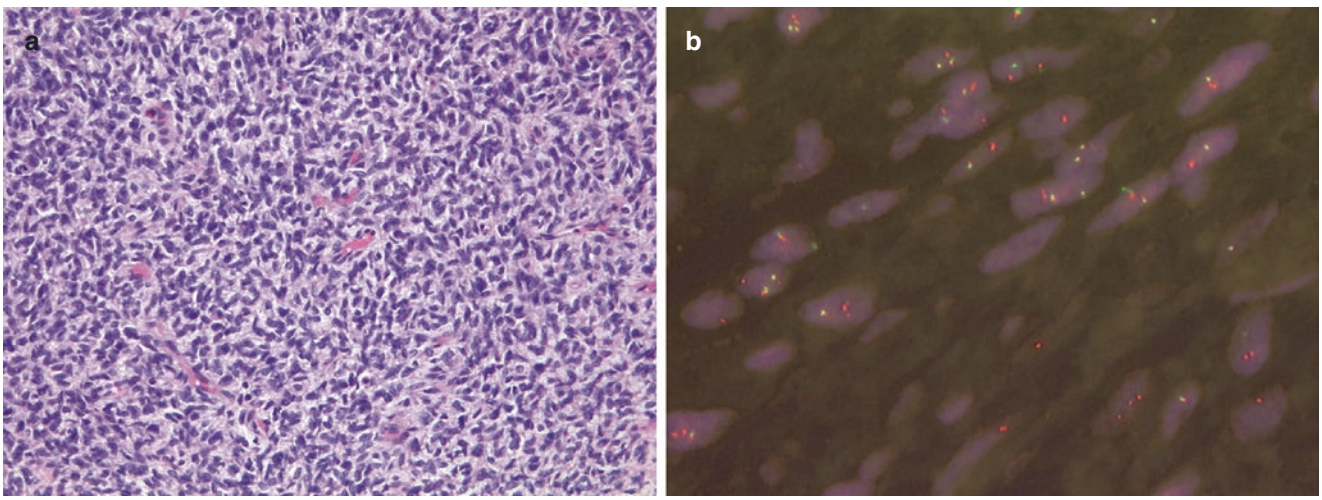


Fig. 33.2 Dermatofibrosarcoma protuberans. (**a**) Characteristic aspect of a case of dermatofibrosarcoma protuberans composed of monomorphic spindle cells growing in a storiform pattern. (**b**) Fluorescence in situ hybridization of the tumor using a *PDGFB* dual-color break-apart probe showing several tumor cells with one red-green signal indicating a normal *PDGFB* locus and 1–3 extra copies of red signals indicating a break in the *PDGFB* locus

genes in chromosome band 12q13: *NAB2* and *STAT6* [26, 133]. *NAB2* is transcribed from telomere to centromere and *STAT6* vice versa. In both genes, several breakpoints have been identified leading to diverse fusion products. Over 90% of SFT display *NAB2-STAT6* fusions. The detection rate depends on primer combinations and perhaps also on sampling. Whether deregulation of *STAT6* or of *NAB2* is the driving force for tumor development is still under debate. *NAB2-STAT6* fusion results in a chimeric protein in which the carboxy-terminal repressor domain of NAB2 (repressing *EGR1*: early growth response gene 1) is replaced with a highly variable portion of STAT6 which seems to play a pivotal role in the development of SFT. However, functional assays will be needed to fully understand the role of *NAB2* and *STAT6* in the development of SFT. Interestingly, both *NAB2* and *STAT6* expression can be detected by immunohistochemistry in SFT [147]. Most if not all SFT show a strong nuclear NAB2 expression as well as nuclear expression of STAT6. The latter is in contrast to the cytoplasmic expression of STAT6 in nonneoplastic tissue. Thus, nuclear localization of STAT6 is possibly a surrogate immunohistochemical marker for *NAB2-STAT6* fusion [179].

Prognosis and Treatment

SFT is another soft tissue neoplasm of intermediate biologic behavior with only 15% of tumors showing aggressive biologic features that include metastatic potential. Prognostic histomorphological parameters of aggressive behavior include tumor size larger than 10 cm, presence of ≥4 mitoses/10 HPFs (high-power fields) necrosis, and a strong or intermediate p53 expression in more than 5% of tumor cells [36, 173]. Barthelmeß and colleagues showed that SFT with the more common *NAB2ex4-STAT6ex2/3* fusion variant corresponded to classic pleuropulmonary SFTs with mostly benign behavior and occurred in older patients, while tumors with the *NAB2ex6-STAT6ex16/17* fusion variant were found in younger patients and represented tumors from deep soft tissue with a more aggressive phenotype and clinical behavior [14]. The treatment of choice is complete surgical resection. In advanced SFT, small series of cases have been successfully treated with tyrosine kinase inhibitors such as sunitinib, figitumumab, or pazopanib with partial response obtained in half of the cases [114, 188].

Inflammatory Myofibroblastic Tumor

Clinicopathological Features

Inflammatory myofibroblastic tumors (IMT) are soft tissue neoplasms of intermediate biologic behavior which frequently recur but only rarely metastasize. They present preferentially in the omentum and the mesentery. Systemic B-symptoms such as fever, anemia, and weight loss may be present at time of diagnosis. IMT more frequently affect children and young adults, but older patients may also develop this neoplasm [40]. Immunohistochemistry shows ALK positivity in more than half of the cases in correlation with the presence of an underlying *ALK* rearrangement. Tumor cells also exhibit smooth muscle actin expression but usually lack other myogenic markers [24].

Genomic Alterations

In up to 60% of the cases, rearrangements of the anaplastic lymphoma kinase (*ALK*) gene (chr. 2p23) have been identified, leading to aberrant constitutive activation of the ALK kinase. The fusion partners include genes encoding cytoplasmic proteins like *TPM3*, *TPM4*, *CARS*, *CLTC*, *ATIC*, and *SEC31L*1, as well as the *RANBP2* gene coding for a nuclear protein. Fusion with the former group of proteins leads to cytoplasmic ALK expression; rearrangements with *RANBP2* lead to nuclear ALK expression and a characteristic round-cell histological phenotype. Immunohistochemical detection of ALK expression reliably predicts *ALK* rearrangements in IMT, and RT-PCR assays as well as *ALK* dual-color break-apart FISH assays may be employed to prove the genomic aberration (reviewed in [69]). Recently, next-generation sequencing revealed that all in all 85% of IMT cases harbored potentially actionable kinase fusions, involving *ROS1* or *PDGFRB* in a subset of cases which were negative for *ALK* gene fusions [119, 121]. Alternatively, RET rearrangements have been reported to occur rarely in IMT [10].

Prognosis and Treatment

IMT are associated with a low risk of aggressive behavior and metastasis. Surgery is the mainstay of treatment. Treatment options in cases with advanced unresectable disease were limited until the detection of *ALK*, *ROS1*, and *PDGFRB* fusions in IMT. The first successful ALK-directed therapy was performed with the ALK/MET-inhibitor crizotinib [21]. However, as in other tumor entities with genomic *ALK* alterations, the development of secondary resistance has emerged in single cases highlighting the need of more specific and diverse *ALK* inhibitors [204].

Infantile Fibrosarcoma

Clinicopathological Features

Infantile fibrosarcomas are low-grade malignant neoplasms with a favorable prognosis that can be present at birth or develop in the first 2 years of life (occasionally up to 4 years). Synonyms are congenital or juvenile fibrosarcoma. Most often, infantile fibrosarcomas occur in the extremities, usually presenting as a rapidly enlarging mass. Rarely, infantile fibrosarcomas involve the trunk or head and neck region.

Infantile fibrosarcomas are negative for β-catenin and myogenic markers.

Genomic Alterations

The vast majority of cases carry a recurrent translocation t(12;15)(p13;q26) leading to the fusion of *NTRK3* and *ETV6*, which results in an oncogenic activation of NTRK3 tyrosine kinase [58]. Routine analysis may be performed by FISH employing ETV6 dual-color break-apart probes or by RT-PCR. Additionally, trisomies of chromosomes 8, 11, 17, and 20 are characteristic. There is a close genomic relationship to cellular congenital mesoblastic nephroma, since they share the same translocation and are very similar in morphology leading to the consensus that cellular congenital mesoblastic nephroma actually represents infantile fibrosarcoma arising in the kidney [170].

Prognosis and Treatment

Infantile fibrosarcomas have a high recurrence rate but only rarely metastasize. The mortality rate ranges from 5% to 25%. Single cases with spontaneous regression have been reported. Complete surgical excision remains the treatment of choice. Adjuvant chemotherapy has been proven effective. Depending on the size of the tumor, resection without major functional consequences may be impossible. In these cases, preoperative chemotherapy can be effective [146, 187]. To date, there are no molecularly based therapeutic approaches [145].

Low-Grade Fibromyxoid Sarcoma

Clinicopathological Features

Low-grade fibromyxoid sarcomas (LGFMS) are deep-seated tumors that are usually located in extremities (especially proximal lower limbs) or trunk. LGFMS occur in young and middle-aged adults (typically fourth decade) with a male predominance. Immunohistochemically, LGFMS frequently express CD34, EMA, and claudin-1. A particularly helpful marker for the differential diagnosis is MUC4 which is typically strongly positive in LGFMS but negative in important differential diagnoses [48]. No expression of myogenic markers, cytokeratins, or S100 protein is expected in LGFMS.

Genomic Alterations

LGFMS typically show a balanced translocation t(7;16)(q32–34;p11) or t(11;16)(p11;p11), leading to the fusion of *FUS* and *CREB3L2* or *CREB3L1*. A *FUS* dual-color break-apart FISH assay is suitable for the detection of the aberration. Rare cases have been reported to carry an *EWSR1-CREB3L1* translocation; those cases may be detected by an *EWSR1* dual-color break-apart assay in cases without a translocation of *FUS*. There is no correlation between the presence and type of the translocation and clinical outcome or morphologic characteristics [48].

Prognosis and Treatment

Recurrences occur in up to 20% of deeply located lesions after time intervals of up to 15 years (median 3.5 years). Metastases occur in about 30% of cases [56], most commonly to the lungs and pleura. In recurrent cases, LGFMS can progress to frank high-grade spindle cell sarcoma. Surgical excision with wide margins is the treatment of choice. As late occurrence of metastases is frequent, long-term follow-up is recommended [56].

Sclerosing Epithelioid Fibrosarcoma

Clinicopathological Features

As implied in the nomenclature, sclerosing epithelioid fibrosarcoma (SEF) is a variant of fibrosarcoma which is characterized, apart from a multinodular growth pattern, by at least focal epithelioid morphology and regions with dense fibrosis. SEF occurs preferentially on limbs, limb girdles, and the trunk and rarely in visceral sites. Often, there is a close connection to periosteum or fascia. Adjacent bone can be involved. SEF can mimic epithelial neoplasms such as lobular breast carcinoma. Furthermore, areas resembling LGFMS or adult-type fibrosarcoma may be encountered in SEF. Immunohistochemistry can help in the differential with metastatic epithelial neoplasms because SEF are negative for cytokeratins. The fact that both SEF and LGFMS strongly express MUC4 in the majority of cases has led to the speculation of whether both lesions belong to a spectrum of one tumor entity. In a study by Doyle et al. [48], all tumors displaying hybrid LGFMS and SEF zones showed strong MUC4 expression. MUC4 expression was also found in 20 of 29 pure SEF.

Genomic Alterations

Like LGFMS, SEF may harbor translocations leading to the fusion of *FUS* and *CREB3L1* or *CREB3L2* [48, 71]. Furthermore, hybrid tumors exhibiting both phenotypes may carry *EWSR1* and *CREB3L1* rearrangements [213]. Wang et al. [213] found pure SEF to often lack *FUS* rearrangements, especially in the absence of MUC4 expression. Therefore, it appears that different genomic subgroups exist among pure LGFMS, pure SEF, and hybrid tumors. The prognostic and predictive value of the above molecular observations remains to be determined. Detection of *FUS* or *EWSR1* rearrangements can be a strong aid to achieving the correct diagnosis in SEF. Detection of the translocation may be performed by *FUS* or *EWSR1* dual-color break-apart probes.

Prognosis and Treatment

SEFs are aggressive sarcomas with a higher than 50% local recurrence rate. Metastases have been reported in 43–86% of cases. A 43–75% range of 5-year survival rates is documented [126].

Complete surgical resection is the treatment of choice. In tumors involving bone, amputation may be required. There is no established role for adjuvant chemotherapy or radiation at this time.

Fibrohistiocytic Tumors

Tenosynovial Giant Cell Tumor

Clinicopathological Features

Tenosynovial giant cell tumors (TGCT) of the tendon sheath arise from the synovium of joints, bursae, and tendon sheaths and are subdivided according to their growth pattern into localized and diffuse subtypes. The former is the more frequent subtype and predominantly occurs in the hand, particularly the fingers. Diffuse TGCT primarily affect the knee, hip, and foot [178]. Both subtypes of TGCT can occur at any age, with a peak incidence in the fourth decade. Diffuse-type TGCT form larger villous or nodular masses; localized TGCT mostly are well-circumscribed nodules with fibrous septae. Histologically, both tumors are composed of a mixture of stromal cells, macrophages, and osteoclast-like giant cells [208].

Genomic Alterations

A balanced translocation t(1;2)(p13;q37) is present in most tumors. Most often, it leads to a chromosomal rearrangement of the *CSF1* (colony-stimulating factor 1) gene locus, partially resulting in a fusion with *COL6A3*, going along with high levels of intralesional CSF1 expression [223]. The gene fusion is present in a minority of the intratumoral cells, whereas the majority of cells express the CSF1 receptor (CSF1R), suggesting a tumor-landscaping effect with aberrant CSF1 expression in the neoplastic cells, leading to the abnormal accumulation of nonneoplastic cells that contribute to the formation of a tumorous mass.

Prognosis and Treatment

Less than one-third of localized-type TGCT recur locally, and these are usually cured by surgical excision. Diffuse-type TGCT are more likely to display locally aggressive behavior with a recurrence rate of up to 50%. Few cases of malignant "sarcomatous" diffuse-type TGCT have been described; these tumors often show a significant increase in mitotic activity, have been shown to express increased levels of cyclin A and (wild-type) p53 and to carry chromosomal losses of the region 15q22–24 [16, 134, 139, 165, 186].

Complete surgical excision is the treatment of choice. Adjuvant radiotherapy has been proposed in cases of recurrent diffuse-type TGCT. In surgically inoperable tumors and in the setting of metastatic disease, tyrosine kinase inhibitors (e.g., imatinib) have been considered [19]. Recently, novel CSF1R-directed inhibitors are being tested in clinical trials with first promising results [198].

Skeletal Muscle Tumors

Embryonal Rhabdomyosarcoma

Clinicopathological Features

Comprising 60–70% of rhabdomyosarcomas, embryonal rhabdomyosarcomas (ERMS) represent the most frequent subtype of malignant soft tissue tumors with skeletal muscle differentiation. ERMS mainly affect children up to 10 years of age with those below 5 years making up about 36% of patients [143]. The majority of these tumors arise in the head and neck and the genitourinary regions, but they may occur at any primary site. Immunohistochemically, rhabdomyoblasts express myogenic markers including desmin, myogenin, and MyoD1 [135]. As the formerly separated botryoid and anaplastic subtypes' outcomes are comparable to typical embryonal rhabdomyosarcomas when adjusting for primary site, resection and metastatic status, they have been eliminated in the last edition of the WHO classification (2013). A novel subtype of rhabdomyosarcoma now separated from ERMS is spindle cell/sclerosing rhabdomyosarcoma which has a better prognosis than typical embryonal rhabdomyosarcoma and most often occurs in the paratesticular site.

Genomic Alterations

ERMS frequently show numerical chromosomal aberrations. The imprinted chromosomal region 11p15.5 that harbors several growth-related genes including the *IGF2* (insulin-like growth factor 2) and *p57KIP2* genes is affected by preferential maternal allelic losses in most cases of ERMS [180, 226]. Deletion of *CDKN2A/B*, a key regulator of the p53 and Rb pathways, is found in the majority of ERMS tumors, whereas inactivating mutations in *TP53* occur in approximately 30% of the tumors [158, 196]. *NF1* deletions leading to activation of RAS signaling occur in 15%. Alternatively, activating *RAS* mutations may be present. These are seen in an additional 40% of cases [158]. Activation of the *FGFR4* tyrosine kinase by amplification of mutant alleles has also been observed in 20% of ERMS [158]. Gene expression profiles indicating activation of the Hedgehog pathway, partially associated with *GLI1* amplification, have been reported to confer a poor prognosis in ERMS as well as translocation-negative alveolar rhabdomyosarcoma (ARMS) [230]. Recently, copy number gains and mutations in the *ALK*

kinase, *PIK3CA*, *CTNNB1*, *FBXW7*, and *BCOR* have been found in ERMS [107, 182, 210]. Kohsaka et al. were the first to show MYOD1 (L122R) mutations in 10% cases of rhabdomyosarcomas classified as embryonal subtype, and it was the same mutation that was then detected in high prevalence in spindle cell and sclerosing rhabdomyosarcomas both in pediatric and in adult cases [3].

Prognosis and Treatment

Established prognostic factors in rhabdomyosarcomas include patient age, histological classification, stage, and site of origin. Five-year survival in patients with conventional type ERMS is 66% [17]. Lower patient age, embryonal (versus alveolar) histologic type, and the (former) botryoid variant are associated with improved outcome, although the latter observation appears to be rather a result of location than of histological subtype. On the other hand, the presence of histologically anaplastic features in ERMS and involvement of the extremities or parameningeal sites have been linked with a worse prognosis [100, 163]. Larger studies casted doubt on the prognostic relevance of anaplasia as a poor prognostic factor [161]. Recently, *ALK* copy number gains were reported to be associated with metastatic disease and poor survival [210].

Treatment is usually multimodal and "risk-adapted." It includes surgery, chemotherapy, and usually radiotherapy. Molecularly targeted approaches are currently under investigation, including substances directed against the IGF1R, mTOR, and VEGF/PDGF [77].

Alveolar Rhabdomyosarcoma

Clinicopathological Features

Alveolar rhabdomyosarcomas (ARMS) account for 20–30% of rhabdomyosarcomas. ARMS more frequently affect adolescents and younger adults and most commonly arise in the extremities followed by the head and neck and trunk regions [75, 154]. The term "alveolar" refers to their typical composition of monomorphous round cells situated in small nests, which are separated by fibrovascular septae. A solid variant also exists. Rhabdomyoblasts are less frequently encountered in ARMS compared to ERMS. Immunohistochemically, desmin, myogenin, and MyoD1 expression serve as markers of ARMS skeletal muscle differentiation [181].

Genomic Alterations

Approximately 75% of ARMS are characterized by the presence of a reciprocal translocation involving the *FOXO1* gene and a partner member of the PAX gene family of transcription factors. t(2;13)(q35;q14) translocation occurs in 60% of tumors and leads to the juxtaposition of the *PAX3* and the *FOXO1* genes, whereas t(1;13)(p36;q14) is present in an additional 10–15% of ARMS, linking the *PAX7* and *FOXO1* genes [12, 35, 66]. For routine diagnostic purposes, a *FKHR (FOXO)* dual-color break-apart FISH assay is well established; RT-PCR is comparable in terms of sensitivity but allows the detection of the translocation partner. *PAX3* and *PAX7* are members of the paired box transcription factor family, and both are involved in skeletal muscle development. *FOXO1* represents a member of the forkhead transcription factor family. The resulting chimeric proteins activate transcription at *PAX3*- and *PAX7*-binding sites, respectively, but are 10–100-fold more potent than wild-type *PAX3* and *PAX7* [226]. While *PAX3-FOXO1* expression is driven by a transcriptional mechanism, *PAX7-FOXO1* gene expression is enhanced by an amplification of the fusion gene [34]. The oncogenic nature of the *PAX-FOXO1* fusion gene has been documented in animal studies [103]. Other rare fusions include *PAX3-NCOA1* and *PAX3-INO80D* [217]. Fusion-positive ARMS frequently carry further genomic amplifications [221] including co-amplification of the *MYCN* gene on chromosome 2p24 as well as a circumscribed region on chromosome 2q13–14 that includes the *CDK4* gene; amplification of *MYCN* or *MIR17HG* has also been described. Furthermore, *ALK* copy number alterations have been shown to be associated with strong *ALK* expression and the presence of metastatic disease at the time of diagnosis.

Prognosis and Treatment

ARMS are generally more aggressive than ERMS with a 5-year survival of 53%, compared to the 66% average cited in ERMS [164]. Importantly, presence of *PAX3-FOXO1* appears to be associated with a worse outcome [77]. Fusion-negative ARMS have a similar prognosis to ERMS. Like ERMS, treatment is risk-adapted and includes surgery, chemotherapy, and usually radiotherapy. Molecularly targeted approaches are also being pursued. The use of CDK4 inhibitors has been proposed in cases with *CDK4* amplification. Two phase I clinical trials are underway evaluating different CDK4 inhibitors [65].

Vascular Tumors

Hemangioendothelioma

Hemangioendotheliomas are vascular neoplasms occupying a spectrum of biological potential ranging from tumors with intermediate to aggressive malignant potential. Kaposiform, retiform, and composite hemangioendotheliomas are among the vascular tumors of intermediate biology, whereas epithe-

lioid hemangioendothelioma (EHE) is a malignant tumor that carries metastatic potential and will be discussed below.

Clinicopathological Features

EHE is a rare tumor that usually occurs in the superficial or deep soft tissue, preferentially in the extremities and head and neck region. Visceral organs can be affected. In a subgroup of patients, EHE is characterized by multicentric growth. In these cases, it has been demonstrated that the different tumor foci are monoclonal in nature [53] and therefore represent metastatic implants of the same neoplastic clone rather than synchronous neoplasms. EHE occurs in all age groups with no gender preference. Immunohistochemically EHE tumor cells are positive for vascular markers including CD34, CD31, D2–40 (podoplanin), Fli-1, and ERG.

Genomic Alterations

A recurrent translocation t(1;3)(p36.3;q25) was initially described, in two cases, by Mendlick et al. [127]. The diagnostic relevance of this translocation was subsequently confirmed [56, 225], and it was shown that this translocation, leading to *WWTR1-CAMTA1* gene fusion, is present in a high percentage of EHE of different anatomic sites but absent in benign epithelioid hemangiomas and epithelioid angiosarcomas that may enter the differential diagnosis of EHE [54]. CAMTA1 belongs to the calmodulin-binding transcription activator family of proteins and is thought to be involved in cell-cycle regulation. WWTR1 is a transcriptional co-activator with a PDZ-binding motif but without known DNA-binding domain. Multiple interaction partners of WWTR1 have been identified, and WWTR1 is a downstream effector of the Hippo pathway. Whether different subtypes of this translocation result in diverse biological behavior in EHE is yet to be determined [56]. Recently, Antonescu et al. described a subgroup of EHE that lacks *WWTR1-CAMTA1* translocation but displays nuclear expression of TFE3 due to an underlying *TFE3* rearrangement [8].

Prognosis and Treatment

The metastatic rate of EHE is 20–30% leading to 10–20% mortality. Adverse prognostic factors include the presence of high mitotic rate (>3/50 HPF) and tumor size larger than 3 cm. The treatment of choice is complete surgical excision with wide margins.

Angiosarcoma

Clinicopathological Features

Angiosarcomas (ASA) are rare tumors that represent only 1% of all sarcomas. Males are more frequently affected. Although ASA can occur in any age group, they are far more frequently found in the elderly and are extremely rare in children. Frequent ASA locations include the soft tissue of extremities (lower more often than upper) and the trunk. A subset of ASA are "secondary" neoplasms that develop following radiotherapy (especially in the setting of adjuvant radiation for breast cancer) [128] and chronic lymphedema or due to exposure to carcinogenic agents such as thorotrast or vinyl chloride (visceral ASA). Secondary ASA may also arise in a background of a preexisting tumor such as schwannoma, *NF1*-associated malignant peripheral nerve sheath tumors, dedifferentiated liposarcoma, or germ cell tumors (for review see [159]).

Immunohistochemically, ASA strongly express endothelial markers such as CD34, CD31, D2–40 (podoplanin), Fli-1, and ERG. The potential for lack of expression of one of these markers in a given ASA exists; thus the use of a marker panel approach is favored. It is also important to remember that CD34 is not endothelial lineage specific and can be expressed in other types of sarcoma. Smooth muscle actin can be used to establish the absence of myopericytes in ASA, a feature of potential diagnostic value in difficult cases [128].

Genomic Alterations

Approximately 10% of primary as well as secondary ASA reveal *KDR* (*VEGFR2*, *FLK-1*) mutations. KDR (kinase insert domain receptor) is a type III receptor tyrosine kinase, the encoding gene of which is located on chromosome 4q11–12. In secondary ASA (following radiotherapy or chronic lymphedema), a high level amplification of *MYC* is detected in the vast majority of cases [73, 121], leading to upregulation of the miRNA cluster 17–92 (13q31.3) [94]. The miR-cluster 17–92 is responsible for the pro-angiogenetic effect of *MYC* amplification by downregulating thrombospondin and connective tissue growth factor [46]. It is suggested that the detection of *MYC* amplification can help differentiate well-differentiated ASA from "atypical vascular lesions" which may also occur in association with prior radiotherapy. Very rare cases of primary ASA and non-ASA soft tissue sarcomas have also been shown to harbor *MYC* amplification.

FLT4 amplification is yet another genomic alteration and is encountered in 25% of secondary ASA. Intriguingly, *FLT4* amplification can co-occur with *MYC* amplification but not with *KDR* mutations.

A rare subset of ASA develops in association with genetic syndromes such as Klippel-Trenaunay syndrome or Maffucci syndrome. Very recently, frequent recurrent PIK3CA mutations have been identified in Klippel-Trenaunay syndrome [120].

Prognosis and Treatment

ASA are very aggressive malignant neoplasms with poor prognosis independent of tumor grade. A dismal 5-year survival rate of 20–30% is expected. Radical surgical resection with wide tumor-free margins is the first choice of treatment.

Adjuvant chemotherapy has been utilized. More recently, the role of targeted therapy with inhibitors of angiogenesis has been explored. In vitro studies have demonstrated that ASA harboring *KDR* mutations may respond to KDR inhibitors such as sorafenib or sunitinib. In secondary ASA with *FLT4* amplification, the role of tyrosine kinase inhibitors could be promising and should be further evaluated. As investigated in a clinical trial, blockade of angiopoietin 2 with the peptibody AMG386 did not result in a significant response [33, 65].

Tumors of Uncertain Differentiation

Synovial Sarcoma

Clinicopathological Features

Synovial sarcomas (SS) represent 5–10% of all malignant soft tissue tumors. They occur more frequently in adolescents and young adults and most commonly arise in the deep soft tissue of the lower and upper extremities followed by the trunk and the head and neck regions. Monophasic and biphasic subtypes of SS are recognized. The latter displays epithelial differentiation in addition to the uniform spindle mesenchymal component of monophasic SS. Immunohistochemically, the majority of SS, regardless of subtype, at least focally express epithelial markers such as EMA and keratins as well as CD99 and TLE1 [131, 144, 203].

Genomic Alterations

At the molecular level, SS are characterized by the presence of a reciprocal translocation t(X;18)(p11;q11), linking the *SS18* (*SYT*) gene and the *SSX1*, *SSX2*, or *SSX4* gene, in order of frequency. The *SYT-SSX1* fusion occurs in approximately two-thirds of SS, whereas the *SYT-SSX2* fusion is found in almost one-third of the cases. Detection of the translocation is well established by RT-PCR and *SS18 (SYT)* dual-color break-apart FISH assays, with the latter showing a higher sensitivity [194]. Neither *SS18* nor the *SSX* genes contain a DNA-binding domain [109]; hence, the *SS18-SSX* chimeric protein exerts its oncogenic function as a part of a multiprotein complex, in which it associates with the transcription factor ATF2 and the repressor TLE1. The multiprotein complex acts by repressing the transcription of ATF2 target genes [192]. Recently, *SS18-SSX* fusion proteins have also been shown to disrupt the repressive action of SWI/SNF complexes on SOX2 expression. The latter protein expression is crucial for the proliferation control of SS cells [101].

Expression of the insulin-like growth factor receptor 1 (*IGF1R*) has been shown to be associated with an aggressive SS phenotype [227]. IGF1R and PI3K have been proposed as therapeutic targets in SS [62, 64]. Furthermore, in preclinical studies, SRC and WNT/β-catenin signals have been shown to represent other targets for therapeutic intervention [129, 206].

Prognosis and Treatment

Prognostic factors include tumor stage, tumor size, and tumor grade. Ten-year disease-specific survival rates of 75% are obtained in children and adolescents and 52% in adults [193]. Presence of *SS18-SSX2* fusion appears to be associated with a more favorable prognosis and a lower rate of metastatic disease at diagnosis [110]. Treatment is multimodal and includes surgery, chemotherapy, and radiotherapy. Novel treatment strategies using bcl2 inhibitors such as ABT-263/Navitoclax have been proposed given the strong expression of bcl2 in SS [65].

Epithelioid Sarcoma

Clinicopathological Features

Epithelioid sarcoma (ES) is a rare type of sarcoma, most frequently arising in distal extremities, especially the hand and the forearm and rarely the head and neck, penile, and vulvar regions. A proximal type of ES occurs in the proximal limb girdle, in axial locations such as the perineum, pelvis, or mediastinum and on the chest wall. ES can involve subcutaneous tissue (typically presenting as a non-healing ulcer) or deep soft tissue.

Immunohistochemically, ES co-expresses vimentin, cytokeratins, EMA, and CD34 (in 50% of cases). CD31, ERG, and S100 are not expressed in ES [185]. Unlike rhabdoid tumors, loss of INI1 expression in ES is usually not associated with *INI1* gene mutations and is thought to be due to epigenetic downregulation through promoter methylation. INI1 loss can be of utility in the differential diagnosis between ES and carcinomas [84].

Genomic Alterations

Cytogenetically, ES shows deletions of chromosome 22. In the classical ES subtype, a chromosomal translocation t(8;22)(q22;q11) is found, albeit inconsistently. Single case reports have illustrated the presence of a t(10;22) in proximal type ES, and single cases with intragenic *INI1 (SMARCB1)* deletions leading to loss of INI1 expression have been reported [60]. In case of doubt, the usual absence of *INI1* mutations may help in distinguishing ES from malignant rhabdoid tumors.

Prognosis and Treatment

Both classic and proximal ES have a high rate of recurrence and can metastasize. However, the proximal type is associated with a higher mortality rate. Metastasis to regional lymph nodes is encountered in up to one-third of cases, an occurrence that is rather unusual for a sarcoma. Hematogenous spread to the lung, bones, brain, and secondary soft tissue locations also occurs in ES. Favorable prognostic parameters include young patient age at diagnosis, tumor size below 2 cm, and female sex. Adverse prognostic factors include

proximal tumor location, presence of tumor necrosis, vascular invasion, and incomplete surgical excision.

An aggressive surgical approach is usually undertaken in ES due to the expected high recurrence rate. Amputation has to be considered in multinodular tumors of extremities. Targeted treatment approaches have not been pursued.

Alveolar Soft Part Sarcoma

Clinicopathological Features

Alveolar soft part sarcoma (ASPS) mainly affects young adults and children. It owes its designation to its alveolar-like clusters of large tumor cell morphology. Whereas the classic alveolar subtype most often occurs in the buttocks and thigh, a solid variant predominates in the tongue and the eye. The majority of ASPS are intramuscular in location. Associated distant metastasis to the lung and brain is present in up to 25% of cases [61]. Immunohistochemically, ASPS is uniquely negative for vimentin unlike most other sarcoma types. Desmin may be focally positive whereas myogenin expression is lacking. CD34, S100, and keratins are not expressed in ASPS. Strong expression of TFE3 in ASPS reflects the underlying *TFE3-ASPL* gene fusion [168].

Genomic Alterations

ASPS carries a specific unbalanced translocation der(17) t(X;17)(p11;p25) leading to the fusion of the *TFE3* gene located on Xp11.2 encoding a transcription factor and the *ASPL* gene on chromosome 17q25, which may be detected by a *TFE3* dual-color break-apart FISH [224, 229].

Prognosis and Treatment

Late recurrences and metastases are common in ASPS. The 5-, 10-, and 20-year survival rates are 60%, 38%, and 15%, respectively; this is a reflection of the rather frequent occurrence of late metastases [61]. Complete surgical resection is the treatment of choice. Recently, antiangiogenic-targeted treatment has been shown to be effective. Furthermore, given the evidence for activation of the AKT/mTOR pathway and MET activation in ASPS, treatment with mTOR inhibitors (e.g., rapamycin) and MET inhibition have been suggested [167]. ASPS has been shown to be targetable by sunitinib [65, 189].

Clear Cell Sarcoma of Soft Tissue

Clinicopathological Features

Described first by Enzinger in 1965 [51], clear cell sarcoma of soft tissue (CCS) was subsequently demonstrated to be of melanocytic differentiation by electron microscopy [82] and distinguished from other sarcomas arising in the tenosynovial soft tissue. CCS occurs in young and middle-aged adults without sex predilection. The majority of cases arises in the ankle or foot, whereas other parts of the extremities are rarely involved. Very rare cases may also occur in the head and neck region, trunk, penis, retroperitoneum, kidney, and gastrointestinal tract. CCS is located in the deep soft tissue and exhibits a relation to tendons or aponeuroses. Immunohistochemistry shows constant S100 positivity. Nearly all CCSs are additionally positive for HMB45 and often for MITF and Melan-A. Neuroendocrine markers can also be coexpressed as well as other nonlineage-specific markers such as CD57 and bcl2 [81]. Myogenic markers, CD117 and CD34, are not expressed in CCS.

Genomic Alterations

CCS often exhibits a complex karyotype. The most relevant alteration is a reciprocal translocation t(12;22)(q13;q12) leading to the fusion of the *EWS1R* and the *ATF1* (activating transcription factor 1 gene) genes [149]. Both genes encode transcription factors and the translocation leads to the fusion of the N-terminal end of *EWSR1* with the bZIP domain of ATF1. The resulting chimeric protein can activate itself in a cAMP-independent manner. Alternatively, *EWSR1* can be fused to *CREB1* (cAMP-responsive element-binding protein 1) leading to the activation of MITF and thus to a melanocytic phenotype. The latter type of translocation is primarily encountered in CCS of the gastrointestinal tract [5] but can also be occasionally seen in CCS of other locations. The detection of either translocation can be very helpful in differentiating CCS from metastatic melanoma.

CCS of the gastrointestinal tract represents a specific subtype that shares some features with its soft tissue counterpart but differs in its higher biologic aggressiveness, behaving like high-grade sarcomas [108]. In contrast to the classical type, CCS of the gastrointestinal tract (CCSLGT) expresses S100 protein but not other melanocytic markers such as HMB45, Melan A, or MITF. CD57 and/or NSE expression have also been reported [63, 228]. At the molecular level both translocation subtypes (i.e., *EWSR1* and either *ATF1* or *CREB1*) can be observed. The novel designation as gastrointestinal neuroectodermal tumor (GNET) has been proposed for this subtype to underline the different biology and the lack of melanocytic differentiation despite similar translocation subtypes as found in CCS [190].

Prognosis and Treatment

Local recurrence is common, especially following incomplete resection. Metastasis occurs in 30% of cases, often late. A 47–67% 5-year survival rate is observed. The survival rate drops to 33% and 10% at 10 and 20 years, respectively. Fifteen percent of cases develop metastases to lymph nodes. Common sites of distant metastases include the lung and bone. A tumor size larger than 5 cm is an adverse prognostic

factor. As indicated above, CCSLGT has an unfavorable prognosis. Local excision is eventually followed by adjuvant chemotherapy.

Extraskeletal Myxoid Chondrosarcoma

Clinicopathological Features

Extraskeletal myxoid chondrosarcomas (ESMC) most frequently arise in men in their fourth to sixth decade. Common intramuscular locations include the lower limb girdle and buttock as well as the distal upper extremities. Less often, ESMC arise in the retroperitoneal and the head and neck regions.

Genomic Alterations

The majority of ESMC harbor translocations involving the *NR4A3* gene (previously designated as *CHN*). *NR4A3* encodes the nuclear receptor subfamily 4 group A type III belonging to the family of steroid and thyroid hormone receptors. In 75% of ESMC, the *NR4A3* translocation involves *EWSR1* [t(9;22)]. This fusion can be demonstrated using a break-apart probe [150]. The resulting *EWSR1-NR4A3* fusion protein contains a transcriptional activation domain and a DNA-binding domain and functions as a transcription factor. Other translocation partners of *NR4A3* include *TAF15*, *TFG*, and *TCF12*. Of interest, as shown by Agaram et al., ESMC with *EWSR1*-independent *NR4A3* gene fusions show a higher incidence of rhabdoid phenotype, high-grade morphology, and a more aggressive outcome compared with the *EWSR1-NR4A3*-positive tumors [4].

Prognosis and Treatment

Typically, ESMC exhibit local recurrences and metastases, often many years following diagnosis. Cellular and high-grade tumors have aggressive behavior. Complete surgical resection with wide margins is the treatment of choice. The response to radio- or chemotherapy is poor.

Desmoplastic Small Round Cell Tumor

Clinicopathological Features

Desmoplastic small round cell tumors (DSRCT) typically arise in young males presenting as an abdominal or pelvic mass. Other locations include the pleura, paratesticular region, brain, ovaries, pancreas, soft tissue, and bones. Because of their frequent topographic relation to serosal surfaces, a mesothelial or submesothelial stem cell origin has been proposed. DSRCT are histologically composed of cords and nests of small round tumor cells separated by desmoplastic stroma, earning them their descriptive designation. Immunohistochemically, these unique tumors display multiphenotypic evidence of differentiation as indicated by their expression of epithelial markers (e.g., cytokeratins and EMA), mesenchymal markers such as desmin, and neuroectodermal markers such as NSE. WT1 expression is a consistent feature of DSRCT. CD99 is coexpressed in 20% of cases [160, 168].

Genomic Alterations

DSCRT exhibit a typical reciprocal translocation t(11;22)(p13;q12) leading to the fusion of *EWSR1* and *WT1* (Wilms tumor 1) genes, which is easily detectable in an *EWSR1* dual-color break-apart FISH approach which may be more suitable than RT-PCR assays due to some variability of the breakpoints. The resulting fusion protein acts as a transcription factor the target of which may include *PDGF-A* that could be responsible for the associated prominent desmoplasia.

Prognosis and Treatment

Overall prognosis is poor. Like other aggressive sarcomas, DSRCT treatment is multimodal in approach. To date, no effective targeted therapy has been developed.

Extrarenal Rhabdoid Tumor

Clinicopathological Features

Extrarenal rhabdoid tumors (ERT) are rare highly aggressive soft tissue tumors occurring predominantly in infancy and childhood [142]. The designation originates from the presence of morphologically and genetically identical tumors that arise in the kidney and the brain. ERT most frequently arise in deep axial locations such as the neck and paraspinal regions. Visceral manifestations (mainly in the liver) are also on record. The "rhabdoid" phenotype is due to the presence of juxtanuclear eosinophilic cytoplasmic inclusions. Immunohistochemically, ERT are positive for vimentin, EMA and keratins, CD99, as well as neuroectodermal markers (synaptophysin and NSE) [57, 106, 207]. ERT characteristically show loss of INI1 nuclear expression [85].

Genomic Alterations

Like their renal and cerebral counterparts, ERT demonstrate homozygous inactivation of the *SMARCB1* (*INI1/HSNF/BAF47*) genes, with a particularly high incidence of smaller deletions of 22q11.22–22q11.23 in soft tissue ERT [95]. The *SMARCB1* gene is a component of the mammalian SWI/SNF complex, which functions in an ATP-dependent manner to remodel chromatin. *SMARCB1* loss is associated with functional disruption of the p16INK4-CyclinD/CDK4-pRb-E2F mitotic checkpoint [211] and an activation of the Hedgehog pathway [96].

Very few cases of rhabdoid tumors retain *SMARCB1* but alternatively display a loss of another SWI/SNF member, *SMARCA4* (*BRG1*) [76, 177]. Large-scale sequencing analyses have revealed only very few additional mutations in rhab-

doid tumors further supporting the crucial oncogenic role of *SMARCB1* inactivation [113].

Prognosis and Treatment

Prognosis is poor. The 5-year overall survival has been reported to be <15% [20]. Treatment options include surgery, chemotherapy, and radiotherapy.

Gastrointestinal Stromal Tumors

Clinicopathological Features

With an estimated yearly incidence of 10–15 cases per million inhabitants [141], gastrointestinal stromal tumors (GIST) are the most common mesenchymal tumors of the gastrointestinal tract. Two-thirds are located in the stomach with the small intestine being the second most frequent tumor location. A small proportion of GIST (5%) occurs in the rectum, and rare examples affecting the esophagus or the peritoneum, the latter without clear connection to the tubular gastrointestinal tract, have been reported as well. The biologic behavior is highly correlated with its primary location with a less aggressive behavior expected in the gastric location compared to extragastric sites.

GIST are thought to be derived from a precursor of the interstitial cells of Cajal (ICCs) with which they share their characteristic expression of the stem cell proteins CD34 and the KIT receptor. Both markers are of great diagnostic utility as immunohistochemical markers of GIST. More recently, DOG1 ("discovered on GIST1") was identified as the most sensitive and specific marker for GIST. DOG1 was identified by gene expression profiling and encodes an ion channel protein with eight transmembrane domains [55, 222]. DOG1 is especially helpful in KIT-negative GISTs [115, 132]. The diagnosis in these often epithelioid gastric GIST can be further supported by the detection of mutations in the *PDGFRα* gene [157, 217].

Aggressive behavior is encountered in approximately 50% of GIST. At least three highly significant prognostic parameters are recognized. These include (a) primary location, (b) tumor size, and (c) mitotic count/5 mm^2. These three parameters are the basis for the most commonly used risk assessment system put forward by Miettinen and Lasota in 2006 [130]; see Table 33.3.

A clinical parameter which is not included in the Miettinen classification but is highly relevant for prognosis is tumor rupture (pre- or intraoperatively). A documented tumor rupture increases the risk of recurrence several folds to a greater than 90% recurrence rate [83].

Genomic Alterations

GIST are among the best examples in oncology regarding how a single somatic mutation can influence prognosis and predict treatment response. Treatment of GIST is regarded as the paradigm for molecular-targeted therapy in solid tumors. Genomic characterization is now well accepted as a central part of the diagnostic process and a prerequisite for treatment planning. Furthermore, the discovery of secondary mutations as a main cause for treatment resistance has pinpointed the most common mechanism for resistance to tyrosine kinase inhibition [215, 220].

Up to 90% of GIST carry primary activating mutations in the *KIT* gene or the *PDGFRα* gene. Both genes are located on chromosome 4 and encode type III receptor tyrosine kinases which display homology in 30% of their amino acids. The prognostic and therapeutically predictive relevance of the mutational status of these genes is now well accepted [39, 79, 112]. The reported frequencies of mutational subtypes differ considerably between anatomic locations of GIST and among different studies, probably due to case selection biases. Primary activating mutations can occur either in the extracellular domain of the receptor protein (i.e., *KIT* exon 9), in the juxtamembraneous domain (*KIT* exon 11, *PDGFRα* exon 12), in the first tyrosine kinase domain *(KIT* exon 13 and 14, *PDGFRα* exon 14), or in the second tyrosine kinase domain (*KIT* exon 17, *PDGFRα* exon 18). Mutations have been reported in *KIT* exon 8 as well, but these seem to be rare [87, 88]. Approximately 65% of all GIST carry *KIT* exon 11 mutations, whereas *KIT* exon 9 and *PDGFRα* exon 18 mutations account for about 10% of primary mutations each. Thus, at least 85% of all GIST carry a mutation at one of these three sites. Another 5% may carry mutations in exons 13, 14, or 17 of *KIT* or in exons 12 or 14 of *PDGFRα*, leading to a frequency of about 1% in each of these regions [87, 88]. As a result, the number of cases with the latter locations of mutations is low in most trials, making it difficult to draw strong conclusions concerning their prognostic and predictive value at this time. The histological phenotype and location of GIST correlate with their *KIT* or the *PDGFRα* mutation status. The vast majority of gastric GIST carry *KIT* exon 11 or *PDGFRA* exon 18 mutations, whereas *KIT* exon 11 and 9 mutations predominate in intestinal GIST [216].

The remaining 10–15% of GIST seem to lack *KIT* or *PDGFRA* mutations and are termed "wild-type GIST." Recently, several small genomic subgroups have been identified among the "wild-type GIST." One important subgroup carries inactivating mutations in the *SDHA*, *SDHB*, *SDHC*, or *SDHD* genes encoding the subunits of the succinate dehydrogenase complex. As a result of the nonfunctioning succinate dehydrogenase complex, succinate accumulates in the cytoplasm and activates the HIF1alpha pathway. *SDH* mutations have been identified both in the germline leading to Carney-Stratakis syndrome (with synchronous or metachronous paragangliomas) and somatically in sporadic GIST. SDH malfunction is also a main mechanism in another syndromic GIST subtype as part of the Carney's triad in association with pulmonary hamartomas/chondromas and

Table 33.3 Metastatic risk of gastrointestinal stromal tumors

Group	Size (cm)	Mitotic count (HPFs)	Metastatic risk			
			Stomach	Jejunum/ileum	Duodenum	Rectum
1	≤2	≤5/5 mm^2	ø	ø	ø	ø
2	>2–5	≤5/5 mm^2	Very low (1.9%)	Low (4.3%)	Low (8.3%)	Low (8.5%)
3a	>5–10	≤5/5 mm^2	Low (3.6%)	Moderate (25%)	High (34%)	High (57%)
3b	>10	≤5/5 mm^2	Moderate (12%)	High (52%)	–[a]	–[a]
4	≤2	>5/5 mm^2	ø[a]	High[a] (50%)	–	High (54%)
5	>2–5	>5/5 mm^2	Moderate (16%)	High (73%)	High (50%)	High (52%)
6a	>5–10	>5/5 mm^2	High (55%)	High (85%)	High (86%)	High (71%)
6b	>10	>5/5 mm^2	High (86%)	High (90%)	High	High

Modified According to [112]
Adapted from Miettinen and Lasota [112]
Based on previously published long-term follow-up studies on 1939 GISTs
– no cases available, *HPFs* high powered fields
[a]Tumor categories with very small numbers of cases

paragangliomas. In these cases inactivation of SDHC is a result of a hypermethylation of the promoter region of the *SDHC* gene [74]. Diagnostically, all GIST with alterations in SDH have in common an immunohistochemical loss of SDHB which is a helpful tool to identify this subgroup. The vast majority of patients are young females with multiple gastric GIST which can metastasize to regional lymph nodes [22]. Interestingly, with exception of GIST occurring as part of Carney's triad and Carney-Stratakis syndrome, lymphatic spread is exceedingly rare in sporadic GIST. It may also occur in pediatric GIST which show a strong morphological and clinical overlap with SDH-deficient GIST but lack comparable alterations [151, 155, 191].

Other subgroups among the so-called wild-type GIST rarely carry *BRAF* mutations as an underlying alteration [2, 90]. Activation of the RAS/RAF/MEK pathway occurs also as a result of mutational inactivation of the neurofibromatosis 1 protein (NF1) which is observed in patients suffering from neurofibromatosis type 1 (104) (see also the following chapter).

GIST without any of the abovementioned alterations are also called "quadruple wild-type" GIST [152]. Recently, other alternative kinase mechanisms involving FGFR1 and NTRK3 genes have also been described [183]. Also, MEN1 and MAX mutations [171] and a neuroendocrine-like molecular heterogeneity have been identified in some of them [153]. MAX inactivation, the gene located on chromosome 14q, seems to be an early event in about 20% of GIST with diverse underlying primary mutations and leads to loss of MAX protein expression. As a result, p16 inactivation and cell cycle perturbations were observed, probably resulting in increased cellular proliferation. During transition to more aggressive tumors, p16, p53, and RB1 mutations occur and may be important prognostic biomarkers [171]. As a late event during GIST progression, tumors develop dystrophin inactivation resulting in enhanced migration and invasion [214].

Mutational Status in Familial Gist

One-third of neurofibromatosis type I (NF I) patients will develop one or more GIST during their lifetime. The majority of these lesions occur in the small bowel and show low aggressive behavior [105, 136, 197]. Another familial setting of GIST is that of a rare familial disorder resulting from a germline mutation in the *KIT* gene (exons 8, 11, 13, 17; for a review see [15, 138]). The latter leads to the development of multiple GIST, in some cases in combination with systemic mastocytosis and ICC hyperplasia throughout the GI tract, and associated dysphagia. More than 30 kindreds carrying *KIT* germline mutations have been described in the literature.

Finally, kindreds with multiple GIST carrying a *PDGFRA* mutation have been described in two reports [27, 37]. Whether these gastrointestinal mesenchymal tumors are genuine GIST or rather represent inflammatory fibroid polyps (IFP) would have to be further explored by novel more specific immunohistochemical markers such as DOG1. IFP carry identical types of *PDGFRA* mutations in the same hot spots and can occur anywhere throughout the gastrointestinal tract [89, 172].

Other Genetic and Epigenetic Mechanisms in Gist Pathogenesis

Compared to other sarcoma subtypes, the majority of GIST have a low cytogenetic complexity. The most frequent alterations are losses of the long arms of chromosomes 14 and/or 22 which are found both in benign and in malignant GIST. With tumor progression the number of chromosomal losses increases with additional losses in 1p, 9p, 9q, 11p, 13q, and 15q and amplifications in 5p, 8q, 17q, and 20q. The altered chromosomal regions harbor putative tumor suppressor genes required for progression from microscopic (measuring below 1 cm) to clinically aggressive GiST. Different types of chromosomal aberrations can be correlated with primary tumor location and are of prognostic value [72].

Inactivating alterations in the tumor suppressor gene *CDKN2A* may also be at play in GIST development. The *CDKN2A* gene is located on chromosome 9p21. The encoded $p16^{INK4}$protein inhibits cyclin-dependent kinases and leads to increased cell proliferation. It has been demonstrated that different types of *CDKN2A* alterations such as promoter methylation, point mutations, or homozygous deletions leading to loss of function are associated with aggressive biological behavior of GIST [174, 175].

Prognostic Relevance of Sporadic *KIT/PDGFRα (Alpha)* Mutations

KIT/PDGFRα mutations are detected in a high proportion of the so-called micro-GISTs that are incidentally detected and measure less than 1 cm [1]. This observation demonstrates that *KIT/PDGFRα* mutations are early oncogenic events in GIST and that other additional genomic or epigenetic events modulate biologic behavior. However, several independent studies have demonstrated a strong correlation between GIST mutational status and their risk for metastasis. Our own group and others have found such a correlation between a specific 6 bp deletion in *KIT* exon 11 (on the protein level p.W557_K558del) and a high metastatic risk [122, 123, 218]. In contrast, the vast majority of GIST with *PDGFRα* mutation show a low level of aggressiveness [111, 217]. The relevance of *KIT* exon 9 mutations for biological behavior remains controversial [6] because these mutations are found almost always in non-gastric GIST, which usually behave more aggressively than gastric GIST. Several groups have proposed the inclusion of tumor mutational status as an additional prognostic parameter in a novel risk classification system.

Predictive Value of *KIT/PDGFRA* Mutations for Treatment Response

The relevance of mutational status for treatment response in metastatic GIST has become clear through multiple trials [38, 78]. In summary, GIST with *KIT* exon 11 mutation have the highest response rates (of 80–90%) to the standard daily dose of 400 mg imatinib. Tumors with *KIT* exon 9 mutation have a lower response rate of about 45%. The response rate in GIST with *PDGFRα* mutations also strongly depends on the mutational subtype [23]. A specific point mutation in *PDGFRα* exon 18 that leads to a substitution of aspartate with valine (p.D842V) results in primary imatinib resistance, whereas tumors with other mutational subtypes in the same exon respond to imatinib. "Wild-type GIST" lacking activating mutations in either *KIT* or *PDGFRα* genes behave in a heterogeneous fashion but overall seem to have a low rate of treatment response. However, their low incidence makes it difficult to draw final conclusions at present.

Adjuvant imatinib treatment for at least 3 years following complete primary resection is associated with improvement in recurrence free and overall survival rates [98, 99]. Furthermore, imatinib can be used in the neoadjuvant setting in primary inoperable GIST. The subsequent reduction of tumor size allows for secondary resection with lower morbidity (reviewed in [49]). As a result, both the European Society for Medical Oncology and the National Comprehensive Cancer Network guidelines strongly recommend molecular typing in GIST [18, 41] to identify patients with primary imatinib resistance (e.g., *PDGFRα* exon 18 mutation p.D842V) and to appropriately adjust imatinib dosage (in case of a *KIT* exon 9 mutation).

Mechanisms of Resistance in Gist

The majority of patients with metastatic GIST develop secondary resistance to imatinib. The frequency of such occurrence is estimated to be at least 80%. It is thought to be due to the development of secondary mutations in the *KIT* gene or rarely in the *PDGFRα* gene that can be demonstrated in a large percentage of progressing tumor samples. The secondary mutations are preferentially located in the kinase domain (*KIT* exons 13, 14, or 17) [219] leading to the inhibition of imatinib binding. It has recently been shown that allele-specific PCR (AS-PCR) approaches are more sensitive than denaturing high-performance liquid chromatography (DHPLC) with regard to the detection rate of resistance mutations [116]. With the increasing usage of next-generation sequencing (NGS) allowing to detect also minor resistant clones occurring only in low frequency, the number of different secondary mutations in the same patient is increasing dramatically, pinpointing to a general novel challenge when using tyrosine kinase inhibitors. Depending on the type of secondary mutation, alternative second-line treatment may be successfully pursued. Sunitinib, another tyrosine kinase inhibitor, has shown effectivity against secondary KIT exon 13 and 14 mutations and has been approved as second-line therapy in GIST resistant to imatinib. Finally, in 2013, regorafenib was approved as a therapeutic option after failure of imatinib and sunitinib. With the knowledge about pathway alterations in GIST, novel treatment strategies gain attention by the usage of novel multi-targeted tyrosine kinase inhibitors but also inhibitors of BRAF, ETV1, MEK, MET, FGFR1, and IGFR1 and checkpoint inhibitors (for review, see [195]). However, effective therapy is lacking as yet for a subgroup of cases with secondary *KIT* exon 17 mutations. In selected cases, resection of single progressing lesions can be an option. The role of pharmacokinetics in secondary imatinib resistance remains to be further clarified [42]. As NGS is now established in numerous centers, liquid biopsy will also play an increasing role in highlighting the relevance of resistant clones and monitoring them during treatment. First clinical trials have demonstrated the high concordance between mutational analysis of plasma DNA and tissue samples [43].

Conclusions

The extensive elucidation of genomic alterations in soft tissue tumors has allowed for a refinement of the morphology-based classification of these neoplasms and provided additional prognostic parameters and new targets of therapy with predictive markers of response.

References

1. Agaimy A, Wünsch P, Hofstaedter F, Blaszyk H, Rümmele P, Gaumann A, Dietmaier W, Hartmann A. Minute gastric sclerosing stromal tumors (GIST tumorlets) are common in adults and frequently show c-KIT mutations. Am J Surg Pathol. 2007;31:113–20.
2. Agaram N, Wong G, Guo T, Maki R, Singer S, Dematteo R, Besmer P, Antonescu C. Novel V600E BRAF mutations in imatinib-naive and imatinib-resistant gastrointestinal stromal tumors. Genes Chromosomes Cancer. 2008;47:853–9.
3. Agaram NP, Chen CL, Zhang L, LaQuaglia MP, Wexler L, Antonescu CR. Recurrent MYOD1 mutations in pediatric and adult sclerosing and spindle cell rhabdomyosarcomas: evidence for a common pathogenesis. Genes Chromosomes Cancer. 2014 Sep;53(9):779–87.
4. Agaram NP, Zhang L, Sung YS, Singer S, Antonescu CR. Extraskeletal myxoid chondrosarcoma with non-EWSR1-NR4A3 variant fusions correlate with rhabdoid phenotype and high-grade morphology. Hum Pathol. 2014 May;45(5):1084–91.
5. Antonescu C, Nafa K, Segal N, Dal Cin P, Ladanyi M. EWS-CREB1: a recurrent variant fusion in clear cell sarcoma – association with gastrointestinal location and absence of melanocytic differentiation. Clin Cancer Res. 2006;12:5356–62.
6. Antonescu C, Sommer G, Sarran L, Tschernyavsky S, Riedel E, Woodruff J, Robson M, Maki R, Brennan M, Ladanyi M, et al. Association of KIT exon 9 mutations with nongastric primary site and aggressive behavior: KIT mutation analysis and clinical correlates of 120 gastrointestinal stromal tumors. Clin Cancer Res. 2003;9:3329–37.
7. Antonescu CR, Elahi A, Healey JH, Brennan MF, Lui MY, Lewis J, Jhanwar SC, Woodruff JM, Ladanyi M. Monoclonality of multifocal myxoid liposarcoma: confirmation by analysis of TLS-CHOP or EWS-CHOP rearrangements. Clin Cancer Res. 2000;6:2788–93.
8. Antonescu CR, Le Loarer F, Mosquera JM, Sboner A, Zhang L, Chen CL, Chen HW, Pathan N, Krausz T, Dickson BC, et al. Novel YAP1-TFE3 fusion defines a distinct subset of epithelioid hemangioendothelioma. Genes Chromosomes Cancer. 2013;52:775–84.
9. Antonescu CR, Tschernyavsky SJ, Decuseara R, Leung DH, Woodruff JM, Brennan MF, Bridge JA, Neff JR, Goldblum JR, Ladanyi M. Prognostic impact of P53 status, TLS-CHOP fusion transcript structure, and histological grade in myxoid liposarcoma: a molecular and clinicopathologic study of 82 cases. Clin Cancer Res. 2001;7:3977–87.
10. Antonescu CR, Suurmeijer AJ, Zhang L, Sung YS, Jungbluth AA, Travis WD, Al-Ahmadie H, Fletcher CD, Alaggio R. Molecular characterization of inflammatory myofibroblastic tumors with frequent ALK and ROS1 gene fusions and rare novel RET rearrangement. Am J Surg Pathol. 2015;39(7):957–67.
11. Azumi N, Curtis J, Kempson RL, Hendrickson MR. Atypical and malignant neoplasms showing lipomatous differentiation. A study of 111 cases. Am J Surg Pathol. 1987;11:161–83.
12. Barr FG. Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. Oncogene. 2001;20:5736–46.
13. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci ND, Weir BA, Ho A, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet. 2010;42:715–21.
14. Barthelmeß S, Geddert H, Boltze C, Moskalev EA, Bieg M, Sirbu H, Brors B, Wiemann S, Hartmann A, Agaimy A, Haller F. Solitary fibrous tumors/hemangiopericytomas with different variants of the NAB2-STAT6 gene fusion are characterized by specific histomorphology and distinct clinicopathological features. Am J Pathol. 2014;184(4):1209–18.
15. Benesch M, Wardelmann E, Ferrari A, Brennan B, Verschuur A. Gastrointestinal stromal tumors (GIST) in children and adolescents: a comprehensive review of the current literature. Pediatr Blood Cancer. 2009;53:1171–9.
16. Bertoni F, Unni KK, Beabout JW, Sim FH. Malignant giant cell tumor of the tendon sheaths and joints (malignant pigmented villonodular synovitis). Am J Surg Pathol. 1997;21:153–63.
17. Betires-Alj M, Guillermo Paez J, David F, Keilhack H, Halmos B, Naoki K, Maris J, Richardson A, Bardelli A, Sugarbaker D, et al. Activating mutations of the Noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. Cancer Res. 2004;64:8816–20.
18. Blay JY, Bonvalot S, Boukovinas I, Casali PG, De Alava E, Dei Tos AP, Dirksen U, Duffaud F, Eriksson M, Fedenko A, Ferrari A, Ferrari S, del Muro XG, Gelderblom H, Grimer R, Gronchi A, Hall KS, Hassan B, Hogendoorn P, Hohenberger P, Issels R, Joensuu H, Jost L, Jurgens H, Kager L, Le Cesne A, Leyvraz S, Martin J, Merimsky O, Nishida T, Picci P, Reichardt P, Rutkowski P, Schlemmer M, Sleijfer S, Stacchiotti S, Taminiau A, Wardelmann E. Gastrointestinal stromal tumors: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. ESMO/European Sarcoma Network Working Group. Ann Oncol. 2012;23:Vii49–55.
19. Blay JY, El Sayadi H, Thiesse P, Garret J, Ray-Coquard I. Complete response to imatinib in relapsing pigmented villonodular synovitis/tenosynovial giant cell tumor (PVNS/TGCT). Ann Oncol. 2008;19:821–2.
20. Bourdeaut F, Freneaux P, Thuille B, Bergeron C, Laurence V, Brugieres L, Verite C, Michon J, Delattre O, Orbach D. Extrarenal non-cerebral rhabdoid tumours. Pediatr Blood Cancer. 2008;51:363–8.
21. Butrynski J, D'Adamo D, Hornick J, Dal Cin P, Antonescu C, Jhanwar S, Ladanyi M, Capelletti M, Rodig S, Ramaiya N, et al. Crizotinib in ALK-rearranged inflammatory myofibroblastic tumor. N Engl J Med. 2010;363:1727–33.
22. Carney J, Stratakis C. Familial paraganglioma and gastric stromal sarcoma: a new syndrome distinct from the carney triad. Am J Med Genet. 2002;108:132–9.
23. Cassier P, Fumagalli E, Rutkowski P, Schöffski P, Van Glabbeke M, Debiec-Rychter M, Emile J, Duffaud F, Martin-Broto J, Landi B, et al. Outcome of patients with platelet-derived growth factor receptor alpha-mutated gastrointestinal stromal tumors in the tyrosine kinase inhibitor era. Clin Cancer Res. 2012;18:4458–64.
24. Cessna M, Zhou H, Sanger W, Perkins S, Tripp S, Pickering D, Daines C, Coffin C. Expression of ALK1 and p80 in inflammatory myofibroblastic tumor and its mesenchymal mimics: a study of 135 cases. Mod Pathol. 2002;15:931–8.
25. Cheng H, Dodge J, Mehl E, Liu S, Poulin N, van de Rijn M, Nielsen TO. Validation of immature adipogenic status and identification of prognostic biomarkers in myxoid liposarcoma using tissue microarrays. Hum Pathol. 2009;40(9):1244–51.
26. Chmielecki J, Crago AM, Rosenberg M, O'Connor R, Walker SR, Ambrogio L, Auclair D, McKenna A, Heinrich MC, Frank DA, Meyerson M. Whole-exome sequencing identifies a recur-

rent NAB2-STAT6 fusion in solitary fibrous tumors. Nat Genet. 2013;45:131–2.
27. Chompret A, Kannengiesser C, Barrois M, Terrier P, Dahan P, Tursz T, Lenoir G, Bressac-de Paillerets B. PDGFRA germline mutation in a family with multiple cases of gastrointestinal stromal tumor. Gastroenterology. 2004;126:318–21.
28. Coindre JM, Pedeutour F, Aurias A. Well-differentiated and dedifferentiated liposarcomas. Virchows Arch. 2010;456:167–79.
29. Colombo C, Miceli R, Lazar AJ, Perrone F, Pollock RE, Le Cesne A, Hartgrink HH, Cleton-Jansen AM, Domont J, Bovée JV, Bonvalot S, Lev D, Gronchi A. CTNNB1 45F mutation is a molecular prognosticator of increased postoperative primary desmoid tumor recurrence: an independent, multicenter validation study. Cancer. 2013;119(20):3696–702.
30. Crago AM, Chmielecki J, Rosenberg M, O'Connor R, Byrne C, Wilder FG, Thorn K, Agius P, Kuk D, Socci ND, Qin LX, Meyerson M, Hameed M, Singer S. Near universal detection of alterations in CTNNB1 and Wnt pathway regulators in desmoid-type fibromatosis by whole-exome sequencing and genomic analysis. Genes Chromosomes Cancer. 2015;54(10):606–15.
31. Crozat A, Aman P, Mandahl N, Ron D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature. 1993;363:640–4.
32. Dal Cin P, Sciot R, Panagopoulos I, Aman P, Samson I, Mandahl N, Mitelman F, Van den Berghe H, Fletcher CD. Additional evidence of a variant translocation t(12;22) with EWS/CHOP fusion in myxoid liposarcoma: clinicopathological features. J Pathol. 1997;182:437–41.
33. D'Angelo SP, Mahoney MR, Van Tine BA, Adkins DR, Perdekamp MT, Condy MM, Luke JJ, Hartley EW, Antonescu CR, Tap WD, Schwartz GK. Alliance A091103 a phase II study of the angiopoietin 1 and 2 peptibody trebananib for the treatment of angiosarcoma. Cancer Chemother Pharmacol. 2015;75(3):629–38.
34. Davis RJ, Barr FG. Fusion genes resulting from alternative chromosomal translocations are overexpressed by gene-specific mechanisms in alveolar rhabdomyosarcoma. Proc Natl Acad Sci U S A. 1997;94:8047–51.
35. Davis RJ, D'Cruz CM, Lovell MA, Biegel JA, Barr FG. Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. Cancer Res. 1994;54:2869–72.
36. de Perrot M, Fischer S, Bründler M, Sekine Y, Keshavjee S. Solitary fibrous tumors of the pleura. Ann Thorac Surg. 2002;74:285–93.
37. De Raedt T, Cools J, Debiec-Rychter M, Brems H, Mentens N, Sciot R, Himpens J, De Wever I, Schöffski P, Marynen P, Legius E. Intestinal neurofibromatosis is a subtype of familial GIST and results from a dominant activating mutation in *PDGFRA*. Gastroenterology. 2006;131:1907–12.
38. Debiec-Rychter M, Dumez H, Judson I, Wasag B, Verweij J, Brown M, Dimitrijevic S, Sciot R, Stul M, Vranck H, et al. Use of c-kit/PDGFRA mutational analysis to predict the clinical response to imatinib in patients with advanced gastrointestinal stromal tumours entered on phase I and II studies of the EORTC soft tissue and bone sarcoma group. Eur J Cancer. 2004;40:689–95.
39. Debiec-Rychter M, Sciot R, Le Cesne A, Schlemmer M, Hohenberger P, van Oosterom A, Blay J, Leyvraz S, Stul M, Casali P, et al. KIT mutations and dose selection for imatinib in patients with advanced gastrointestinal stromal tumours. Eur J Cancer. 2006;42:1093–103.
40. Dehner L. Inflammatory myofibroblastic tumor. The continued definition of one type of so-called inflammatory pseudotumor. Am J Surg Pathol. 2004;28:1652–4.
41. Demetri G, von Mehren M, Antonescu C, DeMatteo R, Ganjoo K, Maki R, Pisters P, Raut C, Riedel R, Schuetze S, et al. NCCN task force report: update on the management of patients with gastrointestinal stromal tumors. J Natl Compr Cancer Netw. 2010;8:S1–41.
42. Demetri G, Wang Y, Wehrle E, Racine A, Nikolova Z, Blanke C, Joensuu H, von Mehren M. Imatinib plasma levels are correlated with clinical benefit in patients with unresectable/metastatic gastrointestinal stromal tumors. J Clin Oncol. 2009;27:3141–7.
43. Demetri GD, Jeffers M, Reichardt P, et al. Mutational analysis of plasma DNA from patients (pts) in the phase III GRID study of regorafenib (REG) versus placebo (PL) in tyrosine kinase inhibitor (TKI)-refractory GIST: correlating genotype with clinical outcomes. J Clin Oncol. 2013;31:10503.
44. Demetri GD, von Mehren M, Jones RL, Hensley ML, Schuetze SM, Staddon A, Milhem M, Elias A, Ganjoo K, Tawbi H, Van Tine BA, Spira A, Dean A, Khokhar NZ, Park YC, Knoblauch RE, Parekh TV, Maki RG, Patel SR. Efficacy and safety of Trabectedin or Dacarbazine for metastatic Liposarcoma or Leiomyosarcoma after failure of conventional chemotherapy: results of a phase III randomized multicenter clinical trial. J Clin Oncol. 2016;34(8):786–93.
45. Demicco EG, Torres KE, Ghadimi MP, Colombo C, Bolshakov S, Hoffman A, Peng T, Bovee JV, Wang WL, Lev D, Lazar AJ. Involvement of the PI3K/Akt pathway in myxoid/round cell liposarcoma. Mod Pathol. 2012;25:212–21.
46. Dews M, Homayouni A, Yu D, Murphy D, Sevignani C, Wentzel E, Furth E, Lee W, Enders G, Mendell J, Thomas-Tikhonenko A. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. Nat Genet. 2006;38:1060–5.
47. Dickson MA, Tap WD, Keohan ML, D'Angelo SP, Gounder MM, Antonescu CR, Landa J, Qin LX, Rathbone DD, Condy MM, et al. Phase II trial of the CDK4 inhibitor PD0332991 in patients with advanced CDK4-amplified well-differentiated or dedifferentiated liposarcoma. J Clin Oncol. 2013;31:2024–8.
48. Doyle L, Wang W, Dal Cin P, Lopez-Terrada D, Mertens F, Lazar A, Fletcher C, Hornick J. MUC4 is a sensitive and extremely useful marker for sclerosing epithelioid fibrosarcoma: association with FUS gene rearrangement. Am J Surg Pathol. 2012;36:1444–51.
49. Eisenberg B, Trent J. Adjuvant and neoadjuvant imatinib therapy: current role in the management of gastrointestinal stromal tumors. Int J Cancer. 2011;129:2533–42.
50. Engstrom K, Willen H, Kabjorn-Gustafsson C, Andersson C, Olsson M, Goransson M, Jarnum S, Olofsson A, Warnhammar E, Aman P. The myxoid/round cell liposarcoma fusion oncogene FUS-DDIT3 and the normal DDIT3 induce a liposarcoma phenotype in transfected human fibrosarcoma cells. Am J Pathol. 2006;168:1642–53.
51. Enzinger FM. Clear-cell sarcoma of tendons and aponeuroses. An analysis of 21 cases. Cancer. 1965;18:1163–74.
52. Enzinger FM, Winslow DJ. Liposarcoma. A study of 103 cases. Virchows Arch Pathol Anat Physiol Klin Med. 1962;335:367–88.
53. Errani C, Sung Y, Zhang L, Healey J, Antonescu C. Monoclonality of multifocal epithelioid hemangioendothelioma of the liver by analysis of WWTR1-CAMTA1 breakpoints. Cancer Genet. 2012;205:12–7.
54. Errani C, Zhang L, Sung Y, Hajdu M, Singer S, Maki R, Healey J, Antonescu C. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. Genes Chromosomes Cancer. 2011;50:644–53.
55. Espinosa I, Lee C, Kim M, Rouse B, Subramanian S, Montgomery K, Varma S, Corless C, Heinrich M, Smith K, et al. A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. Am J Surg Pathol. 2008;32:210–8.
56. Evans HL. Low-grade fibromyxoid sarcoma: a clinicopathologic study of 33 cases with long-term follow-up. Am J Surg Pathol. 2011;35:1450–62.

57. Fanburg-Smith J, Meis-Kindblom J, Fante R, Kindblom L. Malignant granular cell tumor of soft tissue: diagnostic criteria and clinicopathologic correlation. Am J Surg Pathol. 1998;22:779–94.
58. Fisher C. Soft tissue sarcomas with non-EWS translocations: molecular genetic features and pathologic and clinical correlations. Virchows Arch. 2010;456:153–66.
59. Fletcher C, Bridge J, Hogendoorn P, Mertens F, editors. WHO classification of tumours of soft tissue and bone. Lyon: IARC Press; 2013.
60. Flucke U, Slootweg P, Mentzel T, Pauwels P, Hulsebos T. Re: infrequent SMARCB1/INI1 gene alteration in epithelioid sarcoma: a useful tool in distinguishing epithelioid sarcoma from malignant rhabdoid tumor: direct evidence of mutational inactivation of SMARCB1/INI1 in epithelioid sarcoma. Hum Pathol. 2009;40:1361–2.
61. Folpe AL, Deyrup AT. Alveolar soft-part sarcoma: a review and update. J Clin Pathol. 2006;59:1127–32.
62. Friedrichs N, Küchler J, Endl E, Koch A, Czerwitzki J, Wurst P, Metzger D, Schulte J, Holst M, Heukamp L, et al. Insulin-like growth factor-1 receptor acts as a growth regulator in synovial sarcoma. J Pathol. 2008;216:428–39.
63. Friedrichs N, Testi M, Moiraghi L, Modena P, Paggen E, Plötner A, Wiechmann V, Mantovani-Löffler L, Merkelbach-Bruse S, Buettner R, Wardelmann E. Clear cell sarcoma-like tumour with giant cells in the small bowel – further evidence for a new tumor entity? Int J Surg Pathol. 2005;13:313–8.
64. Friedrichs N, Trautmann M, Endl E, Sievers E, Kindler D, Wurst P, Czerwitzki J, Steiner S, Renner M, Penzel R, Koch A, Larsson O, Tanaka S, Kawai A, Schirmacher P, Mechtersheimer G, Wardelmann E, Buettner R, Hartmann W. Phosphatidylinositol-3′-kinase/AKT signaling is essential in synovial sarcoma. Int J Cancer. 2011;129(7):1564–75.
65. Frith AE, Hirbe AC, Van Tine BA. Novel pathways and molecular targets for the treatment of sarcoma. Curr Oncol Rep. 2013;15:378–85.
66. Galili N, Davis RJ, Fredericks WJ, Mukhopadhyay S, Rauscher FJ 3rd, Emanuel BS, Rovera G, Barr FG. Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma. Nat Genet. 1993;5:230–5.
67. Gebhard S, Coindre JM, Michels JJ, Terrier P, Bertrand G, Trassard M, Taylor S, Chateau MC, Marques B, Picot V, Guillou L. Pleomorphic liposarcoma: clinicopathologic, immunohistochemical, and follow-up analysis of 63 cases: a study from the French Federation of Cancer Centers Sarcoma Group. Am J Surg Pathol. 2002;26:601–16.
68. Gengler C, Guillou L. Solitary fibrous tumour and haemangiopericytoma: evolution of a concept. Histopathology. 2006;28:67–74.
69. Gleason BC, Hornick JL. Inflammatory myofibroblastic tumours: where are we now? J Clin Pathol. 2008;61:428–37.
70. Goldblum J. Giant cell fibroblastoma: a report of three cases with histologic and immunohistochemical evidence of a relationship to dermatofibrosarcoma protuberans. Arch Pathol Lab Med. 1996;120:1052–5.
71. Guillou L, Benhattar J, Gengler C, Gallagher G, Ranchère-Vince D, Collin F, Terrier P, Terrier-Lacombe M, Leroux A, Marquès B, et al. Translocation-positive low-grade fibromyxoid sarcoma: clinicopathologic and molecular analysis of a series expanding the morphologic spectrum and suggesting potential relationship to sclerosing epithelioid fibrosarcoma: a study from the French sarcoma group. Am J Surg Pathol. 2007;31:1387–402.
72. Gunawan B, von Heydebreck A, Sander B, Schulten H, Haller F, Langer C, Armbrust T, Bollmann M, Gasparov S, Kovac D, Füzesi F. An oncogenetic tree model in gastrointestinal stromal tumours (GISTs) identifies different pathways of cytogenetic evolution with prognostic implications. J Pathol. 2007;211:463–70.
73. Guo T, Zhang L, Chang N, Singer S, Maki R, Antonescu C. Consistent MYC and FLT4 gene amplification in radiation-induced angiosarcoma but not in other radiation-associated atypical vascular lesions. Genes Chromosomes Cancer. 2011;50:25–33.
74. Haller F, Moskalev EA, Faucz FR, Barthelmeß S, Wiemann S, Bieg M, Assie G, Bertherat J, Schaefer IM, Otto C, Rattenberry E, Maher ER, Ströbel P, Werner M, Carney JA, Hartmann A, Stratakis CA, Agaimy A. Aberrant DNA hypermethylation of SDHC: a novel mechanism of tumor development in carney triad. Endocr Relat Cancer. 2014;21(4):567–77.
75. Harms D. Alveolar rhabdomyosarcoma: a prognostically unfavorable rhabdomyosarcoma type and its necessary distinction from embryonal rhabdomyosarcoma. Curr Top Pathol. 1995;89:273–96.
76. Hasselblatt M, Isken S, Linge A, Eikmeier K, Jeibmann A, Oyen F, Nagel I, Richter J, Bartelheim K, Kordes U, et al. High-resolution genomic analysis suggests the absence of recurrent genomic alterations other than SMARCB1 aberrations in atypical teratoid/rhabdoid tumors. Genes Chromosomes Cancer. 2013;52:185–90.
77. Hawkins DS, Spunt SL, Skapek SX. Children's oncology Group's 2013 blueprint for research: soft tissue sarcomas. Pediatr Blood Cancer. 2013;60:1001–8.
78. Heinrich M, Corless C, Blanke CD, Demetri G, Joensuu H, Roberts P, Eisenberg B, Von Mehren M, Fletcher C, Sandau K, et al. Molecular correlates of imatinib resistance in gastrointestinal stromal tumors. JCO. 2006;24:4764–74.
79. Heinrich M, Owzar K, Corless C, Hollis D, Borden E, Fletcher C, Ryan C, von Mehren M, Blanke C, Rankin C, et al. Correlation of kinase genotype and clinical outcome in the North American Intergroup Phase III trial of imatinib mesylate for treatment of advanced gastrointestinal stromal tumor: CALGB 150105 Study by Cancer and Leukemia Group B and Southwest Oncology Group. J Clin Oncol. 2008;26:5360–7.
80. Henricks WH, Chu YC, Goldblum JR, Weiss SW. Dedifferentiated liposarcoma: a clinicopathological analysis of 155 cases with a proposal for an expanded definition of dedifferentiation. Am J Surg Pathol. 1997;21:271–81.
81. Hisaoka M, Ishida T, Kuo T, Matsuyama A, Imamura T, Nishida K, Kuroda H, Inayama Y, Oshiro H, Kobayashi H, et al. Clear cell sarcoma of soft tissue: a clinicopathologic, immunohistochemical, and molecular analysis of 33 cases. Am J Surg Pathol. 2008;32:452–60.
82. Hoffman GJ, Carter D. Clear cell sarcoma of tendons and aponeuroses with melanin. Arch Pathol. 1973;95:22–5.
83. Hohenberger P, Ronellenfitsch U, Oladeji O, Pink D, Ströbel P, Wardelmann E, Reichardt P. Pattern of recurrence in patients with ruptured primary gastrointestinal stromal tumour. Br J Surg. 2010;97:1854–9.
84. Hollmann TJ, Hornick JL. INI1-deficient tumors: diagnostic features and molecular genetics. Am J Surg Pathol. 2011;35(10):e47–63. https://doi.org/10.1097/PAS.0b013e31822b325b.
85. Hoot AC, Russo P, Judkins AR, Perlman EJ, Biegel JA. Immunohistochemical analysis of hSNF5/INI1 distinguishes renal and extra-renal malignant rhabdoid tumors from other pediatric soft tissue tumors. Am J Surg Pathol. 2004;28:1485–91.
86. Hornick JL, Bosenberg MW, Mentzel T, McMenamin ME, Oliveira AM, Fletcher CD. Pleomorphic liposarcoma: clinicopathologic analysis of 57 cases. Am J Surg Pathol. 2004;28:1257–67.
87. Huss S, Künstlinger H, Wardelmann E, Kleine M, Binot E, Merkelbach-Bruse S, Rüdiger T, Mittler J, Hartmann W, Büttner R, Schildhaus H. A subset of gastrointestinal stromal tumors previously regarded as wild-type tumors carries somatic activating mutations in KIT exon 8 (p.D419del). Mod Pathol. 2013; https://doi.org/10.1038/modpathol.2013.47.
88. Huss S, Nehles J, Binot E, Wardelmann E, Mittler J, Kleine MA, Künstlinger H, Hartmann W, Hohenberger P, Merkelbach-Bruse S, Buettner R, Schildhaus HU. β-Catenin (CTNNB1) mutations

and clinicopathological features of mesenteric desmoid-type fibromatosis. Histopathology. 2013;62:294–304.
89. Huss S, Wardelmann E, Goltz D, Binot E, Hartmann W, Merkelbach-Bruse S, Büttner R, Schildhaus H. Activating PDGFRA mutations in inflammatory fibroid polyps occur in exons 12, 14 and 18 and are associated with tumour localization. Histopathology. 2012;61:59–68.
90. Huss S, Pasternack H, Ihle MA, Merkelbach-Bruse S, Heitkötter B, Hartmann W, Trautmann M, Gevensleben H, Büttner R, Schildhaus HU, Wardelmann E. Clinicopathological and molecular features of a large cohort of gastrointestinal stromal tumors (GISTs) and review of the literature: BRAF mutations in KIT/PDGFRA wild-type GISTs are rare events. Hum Pathol. 2017 Apr;62:206–14. https://doi.org/10.1016/j.humpath.2017.01.005.
91. Italiano A, Bianchini L, Gjernes E, Keslair F, Ranchere-Vince D, Dumollard JM, Haudebourg J, Leroux A, Mainguene C, Terrier P, et al. Clinical and biological significance of CDK4 amplification in well-differentiated and dedifferentiated liposarcomas. Clin Cancer Res. 2009;15:5696–703.
92. Italiano A, Bianchini L, Keslair F, Bonnafous S, Cardot-Leccia N, Coindre JM, Dumollard JM, Hofman P, Leroux A, Mainguene C, et al. HMGA2 is the partner of MDM2 in well-differentiated and dedifferentiated liposarcomas whereas CDK4 belongs to a distinct inconsistent amplicon. Int J Cancer. 2008;122:2233–41.
93. Italiano A, Cardot N, Dupre F, Monticelli I, Keslair F, Piche M, Mainguene C, Coindre JM, Pedeutour F. Gains and complex rearrangements of the 12q13-15 chromosomal region in ordinary lipomas: the "missing link" between lipomas and liposarcomas? Int J Cancer. 2007;121:308–15.
94. Italiano A, Thomas R, Breen M, Zhang L, Crago A, Singer S, Khanin R, Maki R, Mihailovic A, Hafner M, et al. The miR-17-92 cluster and its target THBS1 are differentially expressed in angiosarcomas dependent on MYC amplification. Genes Chromosomes Cancer. 2012;51:569–78.
95. Jackson EM, Sievert AJ, Gai X, Hakonarson H, Judkins AR, Tooke L, Perin JC, Xie H, Shaikh TH, Biegel JA. Genomic analysis using high-density single nucleotide polymorphism-based oligonucleotide arrays and multiplex ligation-dependent probe amplification provides a comprehensive analysis of INI1/SMARCB1 in malignant rhabdoid tumors. Clin Cancer Res. 2009;15:1923–30.
96. Jagani Z, Mora-Blanco EL, Sansam CG, McKenna ES, Wilson B, Chen D, Klekota J, Tamayo P, Nguyen PT, Tolstorukov M, et al. Loss of the tumor suppressor Snf5 leads to aberrant activation of the hedgehog-Gli pathway. Nat Med. 2010;16:1429–33.
97. Jha P, Moosavi C, Fanburg-Smith J. Giant cell fibroblastoma: an update and addition of 86 new cases from the Armed Forces Institute of Pathology, in honor of Dr. Franz M. Enzinger. Ann Diagn Pathol. 2007;11:81–8.
98. Joensuu H, Wardelmann E, Sihto H, Eriksson M, Sundby Hall K, Reichardt A, Hartmann JT, Pink D, Cameron S, Hohenberger P, Al-Batran SE, Schlemmer M, Bauer S, Nilsson B, Kallio R, Junnila J, Vehtari A, Reichardt P. Effect of KIT and PDGFRA mutations on survival in patients with gastrointestinal stromal tumors treated with adjuvant Imatinib: an exploratory analysis of a randomized clinical trial. JAMA Oncol. 2017; https://doi.org/10.1001/jamaoncol.2016.5751. [Epub ahead of print].
99. Joensuu H, Eriksson M, Sundby Hall K, Reichardt A, Hartmann JT, Pink D, Ramadori G, Hohenberger P, Al-Batran SE, Schlemmer M, Bauer S, Wardelmann E, Nilsson B, Sihto H, Bono P, Kallio R, Junnila J, Alvegård T, Reichardt P. Adjuvant Imatinib for high-risk GI stromal tumor: analysis of a randomized trial. J Clin Oncol. 2016;34(3):244–50.
100. Joshi D, Anderson JR, Paidas C, Breneman J, Parham DM, Crist W. Age is an independent prognostic factor in rhabdomyosarcoma: a report from the Soft Tissue Sarcoma Committee of the Children's Oncology Group. Pediatr Blood Cancer. 2004;42:64–73.
101. Kadoch C, Crabtree GR. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. Cell. 2013;153:71–85.
102. Kasper B, Ströbel P, Hohenberger P. Desmoid tumors: clinical features and treatment options for advanced disease. Oncologist. 2011;16:682–93.
103. Keller C, Arenkiel BR, Coffin CM, El-Bardeesy N, DePinho RA, Capecchi MR. Alveolar rhabdomyosarcomas in conditional Pax3:Fkhr mice: cooperativity of Ink4a/ARF and Trp53 loss of function. Genes Dev. 2004;18:2614–26.
104. Kilpatrick SE, Doyon J, Choong PF, Sim FH, Nascimento AG. The clinicopathologic spectrum of myxoid and round cell liposarcoma. A study of 95 cases. Cancer. 1996;77:1450–8.
105. Kinoshita K, Hirota S, Isozaki K, Ohashi A, Nishida T, Kitamura Y, Shinomura Y, Matsuzawa Y. Absence of c-kit gene mutations in gastrointestinal stromal tumours from neurofibromatosis type I patients. J Pathol. 2004;202:80–5.
106. Kodet R, Newton WA Jr, Sachs N, Hamoudi AB, Raney RB, Asmar L, Gehan EA. Rhabdoid tumors of soft tissues: a clinicopathologic study of 26 cases enrolled on the intergroup rhabdomyosarcoma study. Hum Pathol. 1991;22:674–84.
107. Kohsaka S, Shukla N, Ameur N, Ito T, Ng CK, Wang L, Lim D, Marchetti A, Viale A, Pirun M, Socci ND, Qin LX, Sciot R, Bridge J, Singer S, Meyers P, Wexler LH, Barr FG, Dogan S, Fletcher JA, Reis-Filho JS, Ladanyi M. A recurrent neomorphic mutation in MYOD1 defines a clinically aggressive subset of embryonal rhabdomyosarcoma associated with PI3K-AKT pathway mutations. Nat Genet. 2014;46(6):595–600.
108. Kosemehmetoglu K, Folpe A. Clear cell sarcoma of tendons and aponeuroses, and osteoclast-rich tumour of the gastrointestinal tract with features resembling clear cell sarcoma of soft parts: a review and update. J Clin Pathol. 2010;63:416–23.
109. Ladanyi M. Fusions of the SYT and SSX genes in synovial sarcoma. Oncogene. 2001;20:5755–62.
110. Ladanyi M, Antonescu CR, Leung DH, Woodruff JM, Kawai A, Healey JH, Brennan MF, Bridge JA, Neff JR, Barr FG, et al. Impact of SYT-SSX fusion type on the clinical behavior of synovial sarcoma: a multi-institutional retrospective study of 243 patients. Cancer Res. 2002;62:135–40.
111. Lasota J, Dansonka-Mieszkowska A, Sobin L, Miettinen M. A great majority of GISTs with PDGFRA mutations represent gastric tumors of low or no malignant potential. Lab Investig. 2004;84:874–83.
112. Lasota J, Miettinen M. Clinical significance of oncogenic KIT and PDGFRA mutations in gastrointestinal stromal tumours. Histopathology. 2008;53:245–66.
113. Lee RS, Stewart C, Carter SL, Ambrogio L, Cibulskis K, Sougnez C, Lawrence MS, Auclair D, Mora J, Golub TR, et al. A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. J Clin Invest. 2012;122:2983–8.
114. Levard A, Derbel O, Méeus P, Ranchère D, Ray-Coquard I, Blay J, Cassier P. Outcome of patients with advanced solitary fibrous tumors: the Centre Léon Bérard experience. BMC Cancer. 2013;13:109.
115. Liegl B, Hornick J, Corless C, Fletcher C. Monoclonal antibody DOG1.1 shows higher sensitivity than KIT in the diagnosis of gastrointestinal stromal tumors, including unusual subtypes. Am J Surg Pathol. 2009;33:437–46.
116. Liegl B, Kepten I, Le C, Zhu M, Demetri G, Heinrich M, Fletcher C, Corless C, Fletcher J. Heterogeneity of kinase inhibitor resistance mechanisms in GIST. J Pathol. 2008;216:64–74.
117. Lips DJ, Barker N, Clevers H, Hennipman A. The role of APC and beta-catenin in the aetiology of aggressive fibromatosis (desmoid tumors). Eur J Surg Oncol. 2009;35:3–10.

118. Llombart B, Serra-Guillén C, Monteagudo C, López Guerrero J, Sanmartín O. Dermatofibrosarcoma protuberans: a comprehensive review and update on diagnosis and management. Semin Diagn Pathol. 2013;30:13–28.
119. Lovly CM, Gupta A, Lipson D, Otto G, Brennan T, Chung CT, Borinstein SC, Ross JS, Stephens PJ, Miller VA, Coffin CM. Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. Cancer Discov. 2014;4(8):889–95.
120. Luks VL, Kamitaki N, Vivero MP, Uller W, Rab R, Bovée JV, Rialon KL, Guevara CJ, Alomari AI, Greene AK, Fishman SJ, Kozakewich HP, Maclellan RA, Mulliken JB, Rahbar R, Spencer SA, Trenor CC 3rd, Upton J, Zurakowski D, Perkins JA, Kirsh A, Bennett JT, Dobyns WB, Kurek KC, Warman ML, McCarroll SA, Murillo R. Lymphatic and other vascular malformative/overgrowth disorders are caused by somatic mutations in PIK3CA. J Pediatr. 2015;166(4):1048–54.
121. Manner J, Radlwimmer B, Hohenberger P, Mössinger K, Küffer S, Sauer C, Belharazem D, Zettl A, Coindre J, Hallermann C, et al. MYC high level gene amplification is a distinctive feature of angiosarcomas after irradiation or chronic lymphedema. Am J Pathol. 2010;176:34–9.
122. Martin J, Poveda A, Llombart-Bosch A, Ramos R, López-Guerrero J, Garcia del Muro J, Maurel J, Calabuig S, Gutierrez A, González de Sande J, et al. Deletions affecting codons 557–558 of the c-KIT gene indicate a poor prognosis in patients with completely resected gastrointestinal stromal tumors: a study by the Spanish Group for Sarcoma Research (GEIS). J Clin Oncol. 2005;23:6190–8.
123. Martin-Broto J, Gutierrez A, Garcia-del-Muro X, Lopez-Guerrero J, Martinez-Trufero J, de Sande L, Lainez N, Maurel J, De Juan A, Losa F, et al. Prognostic time dependence of deletions affecting codons 557 and/or 558 of KIT gene for relapse-free survival (RFS) in localized GIST: a Spanish Group for Sarcoma Research (GEIS) study. Ann Oncol. 2010;21:1552–7.
124. McArthur G, Demetri G, van Oosterom A, Heinrich M, Debiec-Rychter M, Corless C, Nikolova Z, Dimitrijevic S, Fletcher F. Molecular and clinical analysis of locally advanced dermatofibrosarcoma protuberans treated with imatinib: Imatinib Target Exploration Consortium Study B2225. J Clin Oncol. 2005;23:866–73.
125. McCormick D, Mentzel T, Beham A, Fletcher CD. Dedifferentiated liposarcoma. Clinicopathologic analysis of 32 cases suggesting a better prognostic subgroup among pleomorphic sarcomas. Am J Surg Pathol. 1994;18:1213–23.
126. Meis-Kindblom JM, Kindblom LG, Enzinger FM. Sclerosing epithelioid fibrosarcoma. A variant of fibrosarcoma simulating carcinoma. Am J Surg Pathol. 1995;19:979–93.
127. Mendlick M, Nelson M, Pickering D, Johansson S, Seemayer T, Neff J, Vergara G, Rosenthal H, Bridge J. Translocation t(1;3)(p36.3;q25) is a nonrandom aberration in epithelioid hemangioendothelioma. Am J Surg Pathol. 2001;25:684–7.
128. Mentzel T, Schildhaus HU, Palmedo G, Buttner R, Kutzner H. Postradiation cutaneous angiosarcoma after treatment of breast carcinoma is characterized by MYC amplification in contrast to atypical vascular lesions after radiotherapy and control cases: clinicopathological, immunohistochemical and molecular analysis of 66 cases. Mod Pathol. 2012;25:75–85.
129. Michels S, Trautmann M, Sievers E, Kindler D, Huss S, Renner M, Friedrichs N, Kirfel J, Steiner S, Endl E, Wurst P, Heukamp L, Penzel R, Larsson O, Kawai A, Tanaka S, Sonobe H, Schirmacher P, Mechtersheimer G, Wardelmann E, Büttner R, Hartmann W. SRC signaling is crucial in the growth of synovial sarcoma cells. Cancer Res. 2013;73(8):2518–28.
130. Miettinen M, Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. Semin Diagn Pathol. 2006;23:70–83.
131. Miettinen M, Limon J, Niezabitowski A, Lasota J. Patterns of keratin polypeptides in 110 biphasic, monophasic, and poorly differentiated synovial sarcomas. Virchows Arch. 2000;437:275–83.
132. Miettinen M, Wang Z, Lasota J. DOG1 antibody in the differential diagnosis of gastrointestinal stromal tumors: a study of 1840 cases. Am J Surg Pathol. 2009;33:1401–8.
133. Mohajeri A, Tayebwa J, Collin A, Nilsson J, Magnusson L, von Steyern FV, Brosjo O, Domanski HA, Larsson O, Sciot R, et al. Comprehensive genetic analysis identifies a pathognomonic NAB2/STAT6 fusion gene, nonrandom secondary genomic imbalances, and a characteristic gene expression profile in solitary fibrous tumor. Genes Chromosomes Cancer. 2013; https://doi.org/10.1002/gcc.22083.
134. Moore JR, Weiland AJ, Curtis RM. Localized nodular tenosynovitis: experience with 115 cases. J Hand Surg Am. 1984;9:412–7.
135. Morotti RA, Nicol KK, Parham DM, Teot LA, Moore J, Hayes J, Meyer W, Qualman SJ. An immunohistochemical algorithm to facilitate diagnosis and subtyping of rhabdomyosarcoma: the Children's Oncology Group experience. Am J Surg Pathol. 2006;30:962–8.
136. Mussi C, Schildhaus H, Gronchi A, Wardelmann E, Hohenberger P. Therapeutic consequences from molecular biology for GIST patients affected by neurofibromatosis type I. Clin Cancer Res. 2008;14:4550–5.
137. Negri T, Virdis E, Brich S, Bozzi F, Tamborini E, Tarantino E, Jocolle G, Cassinelli G, Grosso F, Sanfilippo R, et al. Functional mapping of receptor tyrosine kinases in myxoid liposarcoma. Clin Cancer Res. 2010;16:3581–93.
138. Neuhann T, Mansmann V, Merkelbach-Bruse S, Klink B, Hellinger A, Höffkes H, Wardelmann E, Schildhaus H, Tinschert S. A novel germline KIT mutation (p.L576P) in a family presenting with juvenile onset of multiple gastrointestinal stromal tumors, skin hyperpigmentations, and esophageal stenosis. Am J Surg Pathol. 2013;37:898–905.
139. Nielsen AL, Kiaer T. Malignant giant cell tumor of synovium and locally destructive pigmented villonodular synovitis: ultrastructural and immunohistochemical study and review of the literature. Hum Pathol. 1989;20:765–71.
140. Nieuwenhuis M, Lefevre J, Bülow S, Järvinen H, Bertario L, Kernéis S, Parc Y, Vasen H. Family history, surgery, and APC mutation are risk factors for desmoid tumors in familial adenomatous polyposis: an international cohort study. Dis Colon Rectum. 2011;54:1229–34.
141. Nilsson B, Bümming P, Meis-Kindblom JM, Odén A, Dortok A, Gustavsson B, Sablinska K, Kindblom LG. Gastrointestinal stromal tumors: the incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era – a population-based study in western Sweden. Cancer. 2005;103:821–9.
142. Oda Y, Tsuneyoshi M. Extrarenal rhabdoid tumors of soft tissue: clinicopathological and molecular genetic review and distinction from other soft-tissue sarcomas with rhabdoid features. Pathol Int. 2006;56:287–95.
143. Ognjanovic S, Linabery AM, Charbonneau B, Ross JA. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975–2005. Cancer. 2009;115:4218–26.
144. Olsen SH, Thomas DG, Lucas DR. Cluster analysis of immunohistochemical profiles in synovial sarcoma, malignant peripheral nerve sheath tumor, and Ewing sarcoma. Mod Pathol. 2006;19:659–68.
145. Orbach D, Rey A, Cecchetto G, Oberlin O, Casanova M, Thebaud E, Scopinaro M, Bisogno G, Carli M, Ferrari A. Infantile fibrosarcoma: management based on the European experience. J Clin Oncol. 2010;28:318–23.
146. Orbach D, Brennan B, De Paoli A, et al. Conservative strategy in infantile fibrosarcoma is possible: the European paediatric soft tissue sarcoma study group experience. Eur J Cancer. 2016;57:1–9.

147. Ouladan S, Trautmann M, Orouji E, Hartmann W, Huss S, Büttner R, Wardelmann E. Differential diagnosis of solitary fibrous tumors: a study of 454 soft tissue tumors indicating the diagnostic value of nuclear STAT6 relocation and ALDH1 expression combined with in situ proximity ligation assay. Int J Oncol. 2015;46(6):2595–605.
148. Panagopoulos I, Hoglund M, Mertens F, Mandahl N, Mitelman F, Aman P. Fusion of the EWS and CHOP genes in myxoid liposarcoma. Oncogene. 1996;12:489–94.
149. Panagopoulos I, Mertens F, Debiec-Rychter M, Isaksson M, Limon J, Kardas I, Domanski HA, Sciot R, Perek D, Crnalic S, et al. Molecular genetic characterization of the EWS/ATF1 fusion gene in clear cell sarcoma of tendons and aponeuroses. Int J Cancer. 2002;99:560–7.
150. Panagopoulos I, Mertens F, Isaksson M, Domanski HA, Brosjo O, Heim S, Bjerkehagen B, Sciot R, Dal Cin P, Fletcher JA, et al. Molecular genetic characterization of the EWS/CHN and RBP56/CHN fusion genes in extraskeletal myxoid chondrosarcoma. Genes Chromosomes Cancer. 2002;35(4):340–52.
151. Pantaleo M, Astolfi A, Indio V, Moore R, Thiessen N, Heinrich M, Gnocchi C, Santini D, Catena F, Formica S, et al. SDHA loss-of-function mutations in KIT-PDGFRA wild-type gastrointestinal stromal tumors identified by massively parallel sequencing. J Natl Cancer Inst. 2011;103:983–7.
152. Pantaleo MA, Nannini M, Corless CL, Heinrich MC. Quadruple wild-type (WT) GIST: defining the subset of GIST that lacks abnormalities of KIT, PDGFRA, SDH, or RAS signaling pathways. Cancer Med. 2015;4(1):101–3.
153. Pantaleo MA, Urbini M, Indio V, Ravegnini G, Nannini M, De Luca M, Tarantino G, Angelini S, Gronchi A, Vincenzi B, Grignani G, Colombo C, Fumagalli E, Gatto L, Saponara M, Ianni M, Paterini P, Santini D, Pirini MG, Ceccarelli C, Altimari A, Gruppioni E, Renne SL, Collini P, Stacchiotti S, Brandi G, Casali PG, Pinna AD, Astolfi A, Biasco G. Genome-wide analysis identifies MEN1 and MAX mutations and a neuroendocrine-like molecular heterogeneity in quadruple WT GIST. Mol Cancer Res. 2017;15(5):553–62.
154. Parham DM, Alaggio R, Coffin CM. Myogenic tumors in children and adolescents. Pediatr Dev Pathol. 2012;15:211–38.
155. Pasini B, McWhinney S, Bei T, Matyakhina L, Stergiopoulos S, Muchow M, Boikos S, Ferrando B, Pacak K, Assie G, et al. Clinical and molecular genetics of patients with the Carney-Stratakis syndrome and germline mutations of the genes coding for the succinate dehydrogenase subunits SDHB, SDHC, and SDHD. Eur J Hum Genet. 2008;16:79–88.
156. Patel KU, Szabo SS, Hernandez VS, Prieto VG, Abruzzo LV, Lazar AJ, Lopez-Terrada D. Dermatofibrosarcoma protuberans COL1A1-PDGFB fusion is identified in virtually all dermatofibrosarcoma protuberans cases when investigated by newly developed multiplex reverse transcription polymerase chain reaction and fluorescence in situ hybridization assays. Hum Pathol. 2008;39:184–93.
157. Pauls K, Merkelbach-Bruse S, Thal D, Büttner R, Wardelmann E. PDGFR- and c-kit mutated gastrointestinal stromal tumors (GISTs) are characterized by distinctive histological and immunohistochemical features. Histopathology. 2004;46:166–75.
158. Paulson V, Chandler G, Rakheja D, Galindo RL, Wilson K, Amatruda JF, Cameron S. High-resolution array CGH identifies common mechanisms that drive embryonal rhabdomyosarcoma pathogenesis. Genes Chromosomes Cancer. 2011;50:397–408.
159. Penel N, Marreaud S, Robin YM, Hohenberger P. Angiosarcoma: state of the art and perspectives. Crit Rev Oncol Hematol. 2011;80:257–63.
160. Philippe-Chomette P, Kabbara N, Andre N, Pierron G, Coulomb A, Laurence V, Blay JY, Delattre O, Schleiermacher G, Orbach D. Desmoplastic small round cell tumors with EWS-WT1 fusion transcript in children and young adults. Pediatr Blood Cancer. 2012;58:891–7.
161. Qualman S, Lynch J, Bridge J, et al. Prevalence and clinical impact of anaplasia in childhood rhabdomyosarcoma: a report from the Soft Tissue Sarcoma Committee of the Children's Oncology Group. Cancer. 2008;113(11):3242–7.
162. Rabbitts TH, Forster A, Larson R, Nathan P. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. Nat Genet. 1993;4:175–80.
163. Raney RB, Anderson JR, Barr FG, Donaldson SS, Pappo AS, Qualman SJ, Wiener ES, Maurer HM, Crist WM. Rhabdomyosarcoma and undifferentiated sarcoma in the first two decades of life: a selective review of intergroup rhabdomyosarcoma study group experience and rationale for Intergroup Rhabdomyosarcoma Study V. J Pediatr Hematol Oncol. 2001;23:215–20.
164. Raney RB, Anderson JR, Brown KL, Huh WW, Maurer HM, Meyer WH, Parham DM, Rodeberg DA, Wolden SL, Donaldson SS. Treatment results for patients with localized, completely resected (group I) alveolar rhabdomyosarcoma on Intergroup Rhabdomyosarcoma Study Group (IRSG) protocols III and IV, 1984–1997: a report from the Children's Oncology Group. Pediatr Blood Cancer. 2010;55:612–6.
165. Rao AS, Vigorita VJ. Pigmented villonodular synovitis (giant-cell tumor of the tendon sheath and synovial membrane). A review of eighty-one cases. J Bone Joint Surg Am. 1984;66:76–94.
166. Ray-Coquard I, Blay JY, Italiano A, Le Cesne A, Penel N, Zhi J, Heil F, Rueger R, Graves B, Ding M, et al. Effect of the MDM2 antagonist RG7112 on the P53 pathway in patients with MDM2-amplified, well-differentiated or dedifferentiated liposarcoma: an exploratory proof-of-mechanism study. Lancet Oncol 2012;13:1133–1140.
167. Reis H, Hager T, Wohlschlaeger J, Bauer S, Katenkamp K, Katenkamp D, Bubba H. Mammalian target of rapamycin pathway activity in alveolar soft part sarcoma. Hum Pathol. 2013;44(10):2266–74.
168. Rekhi B, Ingle A, Agarwal M, Puri A, Laskar S, Jambhekar N. Alveolar soft part sarcoma 'revisited': clinicopathological review of 47 cases from a tertiary cancer referral centre, including immunohistochemical expression of TFE3 in 22 cases and 21 other tumours. Pathology. 2012;44:11–7.
169. Rutkowski P, Van Glabbeke M, Rankin C, Ruka W, Rubin B, Debiec-Rychter M, Lazar A, Gelderblom H, Sciot R, Lopez-Terrada D, et al. Imatinib mesylate in advanced dermatofibrosarcoma protuberans: pooled analysis of two phase II clinical trials. J Clin Oncol. 2010;28:1772–9.
170. Sandberg A, Bridge J. Updates in the cytogenetics and molecular genetics of bone and soft tissue tumors. Gastrointestinal stromal tumors. Cancer Genet Cytogenet. 2002;135:1–22.
171. Schaefer IM, Wang Y, Liang CW, Bahri N, Quattrone A, Doyle L, Mariño-Enríquez A, Lauria A, Zhu M, Debiec-Rychter M, Grunewald S, Hechtman JF, Dufresne A, Antonescu CR, Beadling C, Sicinska ET, van de Rijn M, Demetri GD, Ladanyi M, Corless CL, Heinrich MC, Raut CP, Bauer S, Fletcher JA. MAX inactivation is an early event in GIST development that regulates p16 and cell proliferation. Nat Commun. 2017;8(8):14674.
172. Schildhaus H, Cavlar T, Binot E, Buettner R, Wardelmann E, Merkelbach-Bruse S. Inflammatory fibroid polyps harbour mutations in the platelet-derived growth factor receptor alpha (PDGFRA) gene. J Pathol. 2008;216:176–82.
173. Schirosi L, Lantuejoul S, Cavazza A, Murer B, Yves Brichon P, Migaldi M, Sartori G, Sgambato A, Rossi G. Pleuro-pulmonary solitary fibrous tumors: a clinicopathologic, immunohistochemical, and molecular study of 88 cases confirming the prognostic value of de Perrot staging system and p53 expression, and evaluating the role of c-kit, BRAF, PDGFRs (alpha/beta), c-met, and EGFR. Am J Surg Pathol. 2008;32:1627–42.

174. Schneider-Stock R, Boltze C, Lasota J, Miettinen M, Peters B, Pross M, Roessner A, Günther T. High prognostic value of p16INK4 alterations in gastrointestinal stromal tumors. J Clin Oncol. 2003;21:1688–97.
175. Schneider-Stock R, Boltze C, Lasota J, Peters B, Corless C, Ruemmele P, Terracciano L, Pross M, Insabato L, Di Vizio D, et al. Loss of p16 protein defines high-risk patients with gastrointestinal stromal tumors: a tissue microarray study. Clin Cancer Res. 2005;11:638–45.
176. Schneider-Stock R, Walter H, Radig K, Rys J, Bosse A, Kuhnen C, Hoang-Vu C, Roessner A. MDM2 amplification and loss of heterozygosity at Rb and p53 genes: no simultaneous alterations in the oncogenesis of liposarcomas. J Cancer Res Clin Oncol. 1998;124:532–40.
177. Schneppenheim R, Fruhwald MC, Gesk S, Hasselblatt M, Jeibmann A, Kordes U, Kreuz M, Leuschner I, Martin Subero JI, Obser T, et al. Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. Am J Hum Genet. 2010;86:279–84.
178. Schwartz HS, Unni KK, Pritchard DJ. Pigmented villonodular synovitis. A retrospective review of affected large joints. Clin Orthop Relat Res. 1989;247:243–55.
179. Schweizer L, Koelsche C, Sahm F, Piro R, Capper D, Reuss D, Pusch S, Habel A, Meyer J, Göck T, et al. Meningeal hemangiopericytoma and solitary fibrous tumors carry the NAB2-STAT6 fusion and can be diagnosed by nuclear expression of STAT6 protein. Acta Neuropathol. 2013;125:651–8.
180. Scrable H, Witte D, Shimada H, Seemayer T, Sheng WW, Soukup S, Koufos A, Houghton P, Lampkin B, Cavenee W. Molecular differential pathology of rhabdomyosarcoma. Genes Chromosomes Cancer. 1989;1:23–35.
181. Sebire NJ, Malone M. Myogenin and MyoD1 expression in paediatric rhabdomyosarcomas. J Clin Pathol. 2003;56:412–6.
182. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, Ambrogio L, Auclair D, Wang J, Song YK, Tolman C, Hurd L, Liao H, Zhang S, Bogen D, Brohl AS, Sindiri S, Catchpoole D, Badgett T, Getz G, Mora J, Anderson JR, Skapek SX, Barr FG, Meyerson M, Hawkins DS, Khan J. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov. 2014;4(2):216–31.
183. Shi E, Chmielecki J, Tang CM, et al. FGFR1 and NTRK3 actionable alterations in "Wild-Type" gastrointestinal stromal tumors. J Transl Med. 2016;14:339.
184. Sirvent N, Coindre JM, Maire G, Hostein I, Keslair F, Guillou L, Ranchere-Vince D, Terrier P, Pedeutour F. Detection of MDM2-CDK4 amplification by fluorescence in situ hybridization in 200 paraffin-embedded tumor samples: utility in diagnosing adipocytic lesions and comparison with immunohistochemistry and real-time PCR. Am J Surg Pathol. 2007;31:1476–89.
185. Sobanko JF, Meijer L, Nigra TP. Epithelioid sarcoma: a review and update. J Clin Aesthet Dermatol. 2009;2:49–54.
186. Somerhausen NS, Fletcher CD. Diffuse-type giant cell tumor: clinicopathologic and immunohistochemical analysis of 50 cases with extraarticular disease. Am J Surg Pathol. 2000;24:479–92.
187. Spunt SL, Million L, Coffin C. The nonrhabdomyosarcoma soft tissue sarcoma. In: Pizzo PA, Poplack DG, editors. Principles and practice of pediatric oncology. 7th ed. Philadelphia: Lippincott Williams and Wilkins; 2015. p. 827–54.
188. Stacchiotti S, Negri T, Palassini E, Conca E, Gronchi A, Morosi C, Messina A, Pastorino U, Pierotti M, Casali P, Pilotti S. Sunitinib malate and figitumumab in solitary fibrous tumor: patterns and molecular bases of tumor response. Mol Cancer Ther. 2010;9:1286–97.
189. Stacchiotti S, Tamborini E, Marrari A, Brich S, Rota SA, Orsenigo M, Crippa F, Morosi C, Gronchi A, Pierotti MA, et al. Response to sunitinib malate in advanced alveolar soft part sarcoma. Clin Cancer Res. 2009;15:1096–104.
190. Stockman DL1, Miettinen M, Suster S, Spagnolo D, Dominguez-Malagon H, Hornick JL, Adsay V, Chou PM, Amanuel B, Vantuinen P, Zambrano EV. Malignant gastrointestinal neuroectodermal tumor: clinicopathologic, immunohistochemical, ultrastructural, and molecular analysis of 16 cases with a reappraisal of clear cell sarcoma-like tumors of the gastrointestinal tract. Am J Surg Pathol. 2012;36(6):857–68.
191. Stratakis C, Carney J. The triad of paragangliomas, gastric stromal tumours and pulmonary chondromas (Carney triad), and the dyad of paragangliomas and gastric stromal sarcomas (Carney-Stratakis syndrome): molecular genetics and clinical implications. J Intern Med. 2009;266:43–52.
192. Su L, Sampaio AV, Jones KB, Pacheco M, Goytain A, Lin S, Poulin N, Yi L, Rossi FM, Kast J, et al. Deconstruction of the SS18-SSX fusion oncoprotein complex: insights into disease etiology and therapeutics. Cancer Cell. 2012;21:333–47.
193. Sultan I, Rodriguez-Galindo C, Saab R, Yasir S, Casanova M, Ferrari A. Comparing children and adults with synovial sarcoma in the surveillance, epidemiology, and end results program, 1983 to 2005: an analysis of 1268 patients. Cancer. 2009;115:3537–47.
194. Sun B, Sun Y, Wang J, Zhao X, Zhang S, Liu Y, Li X, Feng Y, Zhou H, Hao X. The diagnostic value of SYT-SSX detected by reverse transcriptase-polymerase chain reaction (RT-PCR) and fluorescence in situ hybridization (FISH) for synovial sarcoma: a review and prospective study of 255 cases. Cancer Sci. 2008;99:1355–61.
195. Szucs Z, Thway K, Fisher C, Bulusu R, Constantinidou A, Benson C, van der Graaf WTA, Jones RL. Promising novel therapeutic approaches in the management of gastrointestinal stromal tumors. Future Oncology. 2017;13(2):185–94.
196. Takahashi Y, Oda Y, Kawaguchi K, Tamiya S, Yamamoto H, Suita S, Tsuneyoshi M. Altered expression and molecular abnormalities of cell-cycle-regulatory proteins in rhabdomyosarcoma. Mod Pathol. 2004;17:660–9.
197. Takazawa Y, Sakurai S, Sakuma Y, Ikeda T, Yamaguchi J, Hashizume Y, Yokoyama S, Motegi A, Fukayama M. Gastrointestinal stromal tumors of neurofibromatosis type I (von Recklingshausen's disease). Am J Surg Pathol. 2005;29:755–63.
198. Tap WD, Wainberg ZA, Anthony SP, Ibrahim PN, Zhang C, Healey JH, Chmielowski B, Staddon AP, Cohn AL, Shapiro GI, Keedy VL, Singh AS, Puzanov I, Kwak EL, Wagner AJ, Von Hoff DD, Weiss GJ, Ramanathan RK, Zhang J, Habets G, Zhang Y, Burton EA, Visor G, Sanftner L, Severson P, Nguyen H, Kim MJ, Marimuthu A, Tsang G, Shellooe R, Gee C, West BL, Hirth P, Nolop K, van de Rijn M, Hsu HH, Peterfy C, Lin PS, Tong-Starksen S, Bollag G. Structure-guided blockade of CSF1R kinase in Tenosynovial giant-cell tumor. N Engl J Med. 2015;373(5):428–37.
199. Taylor BS, Barretina J, Socci ND, Decarolis P, Ladanyi M, Meyerson M, Singer S, Sander C. Functional copy-number alterations in cancer. PLoS One. 2008;3:e3179.
200. Taylor BS, DeCarolis PL, Angeles CV, Brenet F, Schultz N, Antonescu CR, Scandura JM, Sander C, Viale AJ, Socci ND, Singer S. Frequent alterations and epigenetic silencing of differentiation pathway genes in structurally rearranged liposarcomas. Cancer Discov. 2011;1:587–97.
201. Tejpar S, Nollet F, Li C, Wunder J, Michils G, dal Cin P, Van Cutsem E, Bapat B, van Roy F, Cassiman J, Alman B. Predominance of beta-catenin mutations and beta-catenin dysregulation in sporadic aggressive fibromatosis (desmoid tumor). Oncogene. 1999;18:6615–20.
202. Terrier-Lacombe M, Guillou L, Maire G, Terrier P, Vince D, de Saint Aubain Somerhausen N, Collin F, Pedeutour F, Coindre J. Dermatofibrosarcoma protuberans, giant cell fibroblastoma, and hybrid lesions in children: clinicopathologic comparative analysis of 28 cases with molecular data – a study from the French

Federation of Cancer Centers Sarcoma Group. Am J Surg Pathol. 2003;27:27–39.
203. Terry J, Saito T, Subramanian S, Ruttan C, Antonescu CR, Goldblum JR, Downs-Kelly E, Corless CL, Rubin BP, van de Rijn M, et al. TLE1 as a diagnostic immunohistochemical marker for synovial sarcoma emerging from gene expression profiling studies. Am J Surg Pathol. 2007;31:240–6.
204. Tothova Z, Wagner A. Anaplastic lymphoma kinase-directed therapy in inflammatory myofibroblastic tumors. Curr Opin Oncol. 2012;24:409–13.
205. Trautmann M, Menzel J, Bertling C, Cyra M, Isfort I, Steinestel K, Elges S, Grünewald I, Altvater B, Rossig C, Fröhling S, Hafner S, Simmet T, Åman P, Wardelmann E, Huss S, Hartmann W. FUS-DDIT3 fusion protein driven IGF-IR signaling is a therapeutic target in myxoid liposarcoma. Clin Cancer Res. 2017; https://doi.org/10.1158/1078-0432.CCR-17-0130.
206. Trautmann M, Sievers E, Aretz S, Kindler D, Michels S, Friedrichs N, Renner M, Kirfel J, Steiner S, Huss S, Koch A, Penzel R, Larsson O, Kawai A, Tanaka S, Sonobe H, Waha A, Schirmacher P, Mechtersheimer G, Wardelmann E, Büttner R, Hartmann W. SS18-SSX fusion protein-induced Wnt/β-catenin signaling is a therapeutic target in synovial sarcoma. Oncogene. 2014;33(42):5006–16.
207. Tsuneyoshi M, Daimaru Y, Hashimoto H, Enjoji M. Malignant soft tissue neoplasms with the histologic features of renal rhabdoid tumors: an ultrastructural and immunohistochemical study. Hum Pathol. 1985;16:1235–42.
208. Ushijima M, Hashimoto H, Tsuneyoshi M, Enjoji M. Giant cell tumor of the tendon sheath (nodular tenosynovitis). A study of 207 cases to compare the large joint group with the common digit group. Cancer. 1986;57:875–84.
209. van Broekhoven DL, Verhoef C, Grünhagen DJ, van Gorp JM, den Bakker MA, Hinrichs JW, de Voijs CM, van Dalen T. Prognostic value of CTNNB1 gene mutation in primary sporadic aggressive fibromatosis. Ann Surg Oncol. 2015 May;22(5):1464–70.
210. van Gaal JC, Flucke UE, Roeffen MH, de Bont ES, Sleijfer S, Mavinkurve-Groothuis AM, Suurmeijer AJ, van der Graaf WT, Versleijen-Jonkers YM. Anaplastic lymphoma kinase aberrations in rhabdomyosarcoma: clinical and prognostic implications. J Clin Oncol. 2012;30:308–15.
211. Vries RG, Bezrookove V, Zuijderduijn LM, Kia SK, Houweling A, Oruetxebarria I, Raap AK, Verrijzer CP. Cancer-associated mutations in chromatin remodeler hSNF5 promote chromosomal instability by compromising the mitotic checkpoint. Genes Dev. 2005;19:665–70.
212. Walluks K, Chen Y, Woelfel C, Yang L, Cui T, Seliger C, Geier C, Knosel T, Hauke S, Petersen I. Molecular and clinicopathological analysis of dermatofibrosarcoma protuberans. Pathol Res Pract. 2013;209:30–5.
213. Wang W, Evans H, Meis J, Liegl-Atzwanger B, Bovee J, Goldblum J, Billings S, Rubin B, López-Terrada D, Lazar A. FUS rearrangements are rare in 'pure' sclerosing epithelioid fibrosarcoma. Mod Pathol. 2012;25:846–53.
214. Wang Y, Marino-Enriquez A, Bennett RR, et al. Dystrophin is a tumor suppressor in human cancers with myogenic programs. Nat Genet. 2014;46:601–6.
215. Wardelmann E, Biermann K, Merkelbach-Bruse S, Schildhaus H, Thomas N, Buettner R, Pietsch T, Heinicke T, Speidel N, Pink D, et al. Polyclonal resistance in gastrointestinal stromal tumor treated with sequential kinase inhibitors. Clin Cancer Res. 2006;12:6206.
216. Wardelmann E, Hohenberger P, Reichardt P, Merkelbach-Bruse S, Schildhaus H, Büttner R. Gastrointestinal stromal tumors of the stomach. Updates and differences compared to other locations. For Pathol. 2010;31:195–8.
217. Wardelmann E, Hrychyk A, Merkelbach-Bruse S, Pauls K, Goldstein J, Hohenberger P, Losen I, Manegold C, Büttner R, Pietsch T. Association of platelet-derived growth factor receptor alpha mutations with gastric primary site and epithelioid or mixed cell morphology in gastrointestinal stromal tumors. J Mol Diagn. 2004;6:197–204.
218. Wardelmann E, Losen I, Hans V, Neidt I, Speidel N, Bierhoff E, Heinicke T, Pietsch T, Büttner R, Merkelbach-Bruse S. Deletion of Trp-557 and Lys-558 in the juxtamembrane domain of the c-kit protooncogene is associated with metastatic behavior of gastrointestinal stromal tumors. Int J Cancer. 2003;106:887–95.
219. Wardelmann E, Merkelbach-Bruse S, Pauls K, Thomas N, Schildhaus H, Heinicke T, Speidel N, Pietsch T, Buettner R, Pink D, et al. Polyclonal evolution of multiple secondary *KIT* mutations in gastrointestinal stromal tumors under treatment with imatinib mesylate. Clin Cancer Res 2006;12:1743–1749.
220. Wardelmann E, Thomas N, Merkelbach-Bruse S, Pauls K, Speidel N, Büttner R, Bihl H, Leutner C, Heinicke T, Hohenberger P. Acquired resistance to imatinib in gastrointestinal stromal tumors caused by multiple *KIT* mutations. Lancet Oncol. 2005;6:249–51.
221. Weber-Hall S, Anderson J, McManus A, Abe S, Nojima T, Pinkerton R, Pritchard-Jones K, Shipley J. Gains, losses, and amplification of genomic material in rhabdomyosarcoma analyzed by comparative genomic hybridization. Cancer Res. 1996;56:3220–4.
222. West R, Corless C, Chen X, Rubin B, Subramanian S, Montgomery K, Zhu S, Ball C, Nielsen T, Patel R, et al. The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. Am J Pathol. 2004;165:107–13.
223. West RB, Rubin BP, Miller MA, Subramanian S, Kaygusuz G, Montgomery K, Zhu S, Marinelli RJ, De Luca A, Downs-Kelly E, et al. A landscape effect in tenosynovial giant-cell tumor from activation of CSF1 expression by a translocation in a minority of tumor cells. Proc Natl Acad Sci U S A. 2006;103:690–5.
224. Williams A, Bartle G, Sumathi V, Meis J, Mangham D, Grimer R, Kindblom L. Detection of ASPL/TFE3 fusion transcripts and the TFE3 antigen in formalin-fixed, paraffin-embedded tissue in a series of 18 cases of alveolar soft part sarcoma: useful diagnostic tools in cases with unusual histological features. Virchows Arch. 2011;458:291–300.
225. Woelfel C, Liehr T, Weise A, Langrehr J, Kotb W, Pacyna-Gengelbach M, Katenkamp D, Petersen I. Molecular cytogenetic characterization of epithelioid hemangioendothelioma. Cancer Genet. 2011;204:671–6.
226. Xia SJ, Pressey JG, Barr FG. Molecular pathogenesis of rhabdomyosarcoma. Cancer Biol Ther. 2002;1:97–104.
227. Xie Y, Skytting B, Nilsson G, Brodin B, Larsson O. Expression of insulin-like growth factor-1 receptor in synovial sarcoma: association with an aggressive phenotype. Cancer Res. 1999;59:3588–91.
228. Zambrano E, Reyes-Mugica M, Franchi A, Rosai J. An osteoclast-rich tumor of the gastrointestinal tract with features resembling clear cell sarcoma of soft parts: reports of 6 cases of a GIST simulator. Int J Surg Pathol. 2003;11:75–81.
229. Zhong M, De Angelo P, Osborne L, Keane-Tarchichi M, Goldfischer M, Edelmann L, Yang Y, Linehan WM, Merino MJ, Aisner S, Hameed M. Dual-color, break-apart FISH assay on paraffin-embedded tissues as an adjunct to diagnosis of Xp11 translocation renal cell carcinoma and alveolar soft part sarcoma. Am J Surg Pathol. 2010;34:757–66.
230. Zibat A, Missiaglia E, Rosenberger A, Pritchard-Jones K, Shipley J, Hahn H, Fulda S. Activation of the hedgehog pathway confers a poor prognosis in embryonal and fusion gene-negative alveolar rhabdomyosarcoma. Oncogene. 2010;29:6323–30.

Genomic Applications in Melanoma

34

Carlos N. Prieto-Granada, John Van Arnam,
Kabeer K. Shah, Aleodor A. Andea, and Alexander J. Lazar

Introduction

Melanocytes are neural crest-derived cells, which migrate during development primarily to the epidermis and adnexal structures in addition to mucosal surfaces, the meninges, and the eye [1–3]. Melanocytes synthesize melanin pigment in specialized organelles called melanosomes. Each melanocyte distributes melanosomes along dendritic processes to five to eight adjacent keratinocytes. Following ultraviolet (UV) light exposure, the melanosomes localize above the keratinocyte nuclei providing protection from the deleterious effects of UV radiation (UVR) [3].

Overall incidence rates of melanoma appear to be increasing worldwide with 160,000 cases per year and resulting in up to 48,000 deaths per annum. In the USA, melanoma is now the 5th most common cancer among men and 6th among women, with an estimated 91,270 new cases and 9320 deaths projected for the year 2018 [4].

The management of melanoma is currently based on the clinicopathologic characteristics and histology of the primary tumor and metastases including evaluation of many well-established prognostic factors. For primary tumors, these include the features required by the last edition of the American Joint Cancer Commission (AJCC) TNM staging system. The T-stage is mainly dictated by the depth of invasion as determined by Breslow thickness, in addition to the presence or absence of overlying ulceration. Additional features, such as the presence of a vertical growth phase, mitotic activity, and lymphovascular and/or perineural invasion, are predictive of recurrence—both local and distant. The N-stage is obtained by evaluating the draining lymph node basin for metastatic deposits and by detecting satellite lesions and/or in-transit metastasis. Finally, metastatic disease (M-stage) is staged based on tumor involvement beyond the regional lymph node basin including lymph nodes from different regional basins, the skin, soft tissue, visceral sites, and particularly brain, in addition to the detection of elevated serum lactate dehydrogenase (LDH). In the recently adopted AJCC 8th edition, properties of the primary tumor continue to determine risk progression, and thus stage groups offer improved prognostic accuracy and stratification [5].

Melanocytic lesions demonstrate protean clinical and pathological presentations with a wide variety of histological patterns, from bland, low-cellularity lesions such as benign melanocytic nevi to atypical (dysplastic) melanocytic nevi and finally to malignant melanoma. While histologic examination allows accurate classification of melanocytic tumor in the majority of cases, in a subset of lesions, evaluation of histologic parameters cannot reliably distinguish benign (nevi) from malignant (melanoma) proliferations.

Numerous methods, ranging from clinical impression, histologic analysis, immunohistochemical studies, fluorescence in situ hybridization (FISH), to genome-wide techniques including comparative genomic hybridization (CGH), single nucleotide polymorphism (SNP) microarrays, and various sequencing technologies, are used in concert for diagnostic purposes (melanoma versus nevus), risk stratification (likelihood of progression), and the so-called

C. N. Prieto-Granada (✉)
Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA
e-mail: cprieto@uabmc.edu

J. Van Arnam
Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX, USA

K. K. Shah
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

A. A. Andea
Departments of Pathology and Dermatology, University of Michigan, Ann Arbor, MI, USA

A. J. Lazar
Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX, USA

Departments of Genomic Medicine & Dermatology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: alazar@mdanderson.org

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_34

theragnosis (portmanteau of "therapy" and "diagnosis," targeted therapeutic selection). Recent advances in DNA and RNA sequencing have striking diagnostic and therapeutic implications. These same techniques have offered numerous insights into the key molecular pathways that drive these lesions, several of which have been exploited for effective targeted therapy.

This chapter focuses on the underlying mechanisms of the pathogenesis of melanocytic proliferations as a basis for their accurate classification and effective treatment.

Risk Factors

Environmental

The main risk factor predisposing to the development of melanoma is represented by UVR-induced damage from sun exposure, which can be either intermittent (such as in lesions from the trunk and proximal extremities) or chronic (often involving the head and neck and upper distal extremities) (Fig. 34.1). In addition, exposure to artificial sources of UVR has also been linked to increased risk of developing melanoma including both therapeutic (PUVA therapy, nowadays largely replaced by safer narrow-band UV therapy) or aesthetic (tanning beds) purposes. The use of tanning beds in particular has become a genuine public health problem in recent years, and it can even be associated with addictive behavior [6, 7].

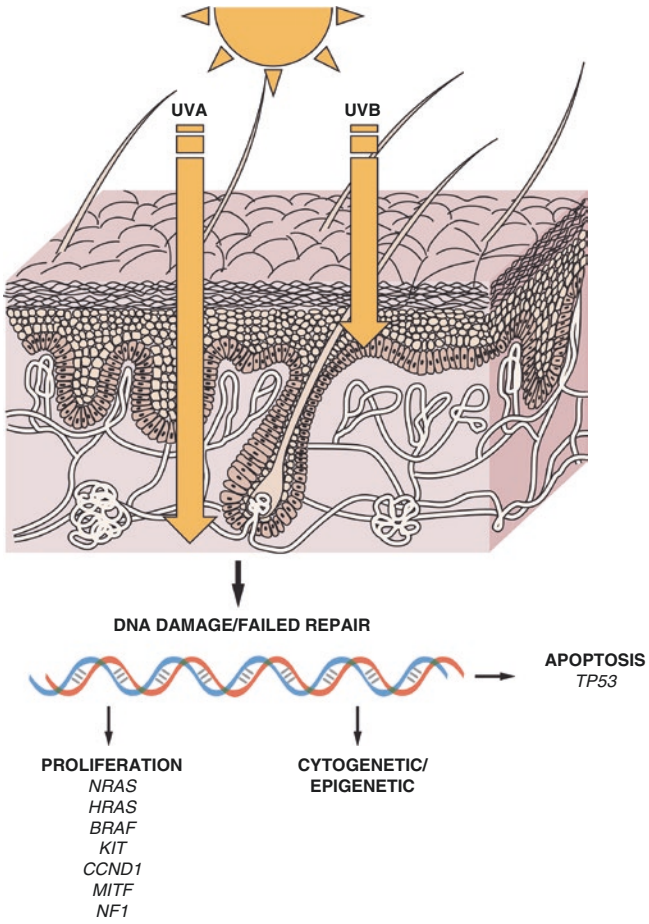


Fig. 34.1 UV radiation is formed by a broad spectrum of different wavelengths (bands) with the most significant categories in human disease being UVA (315–400 nm) and UVB (280–315 nm). UVA causes indirect DNA damage through free radical generation, and UVB directly interacts and alters DNA by forming pyrimidine dimers. Cumulative DNA damage will eventually result in either p53-mediated apoptosis (**right**) or survival of abnormal cells harboring multiple oncogenes and tumor suppressor gene mutations that will drive melanomagenesis (**left**). Many additional bystander mutations are also present. UVR is also likely involved in generation of cytogenetic and epigenetic aberrations (**center**)

Genetic

Numerous genetic syndromes are associated with increased risk for melanoma to various degrees. The following section enumerates select melanoma-associated syndromes associated with germline mutations/variations of the following gene groups:

Oncogenes and tumor suppressor genes:

- *CDKN2A* (9p21.3, OMIM 600160): Dysplastic nevus syndrome/familial atypical multiple mole melanoma (FAMMM) (OMIM 155601) and familial melanoma-pancreatic cancer syndrome (OMIM 606719) [8]
- *CDK4* (12q14.1, OMIM 123829): Susceptibility to cutaneous melanoma syndrome 3 (OMIM 609048) [9, 10]
- *BAP1* (3p21.1, OMIM 603089): *BAP1* tumor predisposition syndrome (OMIM 614327) [11–13]
- *TP53* (17p13.1, OMIM 191170): Li-Fraumeni syndrome (OMIM 151623) [14, 15]
- *PTEN* (10q23.31, OMIM 601728): Cowden syndrome (OMIM 158350) [16, 17]
- *RB* (13q14.2, OMIM 614041): Retinoblastoma syndrome (OMIM 180200) [18–20]
- *MITF* (3p13, OMIM 156845): Melanoma and renal cell carcinoma susceptibility syndrome (OMIM 614456) [21, 22]
- *FLCN* (17p11.2, OMIM 607273): Birt-Hogg-Dubé syndrome (OMIM 135150) [23–25]

Genes affecting DNA damage repair:

- Genes coding nucleotide-excision DNA repair XP proteins (*XPA*, 9q22.33, OMIM 611153; *ERCC3*, 2q14.3, OMIM 133510; *XPC*, 3p25.1, OMIM 6132018; *ERCC2*, 19q13.32, OMIM 126340; *DDB2*, 11p11.2, OMIM 600811; *ERCC4*, 16p13.12, OMIM 133520; *ERCC5*, 13q33.1, OMIM 133530; *POLH*, 6p21.1, OMIM 603968) and xeroderma pigmentosum A-G complementation

groups and variant (OMIM numbers: 278700, 610651, 278720, 278730, 278740, 278760, 278780, 278750) [26–28]
- *RECQL2* (8p12, OMIM 604611): Werner syndrome (OMIM 27770) [29]
- *BRCA2* (13q13.1, OMIM 600185): Familial breast-ovarian cancer syndrome (OMIM 612555) [30]

Genes involved in telomeric maintenance:
- *POT1* (7q31.33, OMIM 606478): Susceptibility to cutaneous melanoma syndrome 10 (OMIM 615848) [31, 32]

Gene alterations/variations resulting in increased photosensitivity:
- *MC1R* (16q24.3, OMIM 155555): Susceptibility to cutaneous melanoma syndrome 5 (OMIM 613099) [33, 34]
- Genes involved in the oculocutaneous albinism group of entities (multiple loci and OMIM numbers)

Molecular Pathways Frequently Affected in Melanoma with Corresponding Morphologic Correlates and Therapeutic Interventions

Clinicopathologically, the current World Health Organization (WHO) classification of cutaneous melanoma is based on examining the morphologic characteristics of morphologic characteristics of the radial growth phase (RGP) the lesion and encompasses four main primary subtypes: superficial spreading melanoma (SSM) (pagetoid and nested RGP) (Fig. 34.2b), lentigo maligna melanoma (LMM) (lentiginous RGP) (Fig. 34.2d), acral lentiginous melanoma (ALM) (lentiginous RGP) (Fig. 34.2e), and the "wastebasket" category that is nodular melanoma (NM) (no RGP and pure vertical growth phase [VGP]) (Fig. 34.2c) [35]. The most common histologic subtype is SSM, primarily affecting fair-skinned young to middle-aged adults and involving intermittently sun-damaged skin of the proximal extremities and trunk. Melanoma variants that arise most commonly in the elderly population include LMM and NM, with LMM being found most commonly associated with chronically sun-damaged skin of the head and neck. Finally, ALM, unlike the aforementioned subtypes, affects darker- and lighter-skinned populations equally, often arising in a field effect of normal-appearing melanocytes of acral sites. Less common melanoma subtypes include the so-called blue nevus-like melanoma, desmoplastic melanoma (pure and mixed) (Fig. 34.2f), spitzoid melanoma (Fig. 34.2l), and nevoid melanoma, among others. Extra-cutaneous sites of primary melanoma are classified separately, and for vulvar and uveal sites, separate staging criteria are employed [36, 37]. Depiction of some of these melanoma subtypes and benign melanocytic lesions is shown in Fig. 34.2.

Although it has been pointed out in the past that this morphology-based classification has limited therapeutic [38] and prognostic relevance [39], the rise of high-throughput sequencing and concomitant development of rational drug design leads to the development of new classification schemes with prognostic and therapeutic importance. Nevertheless, the mutational profile of the lesions sometimes segregates with the aforementioned clinicopathological entities. In the next sections, the main pathways involved in melanoma tumorigenesis will be reviewed. Table 34.1a (epidermal-/epithelial-induced lesions) and Table 34.1b (*non*-epidermal/epithelial-induced lesions) provide a summary of the major mutations involved in melanocytic processes.

MAPK/ERK Pathway

The mitogen-activated protein kinase/extracellular-regulated kinase (MAPK/ERK) pathway (Fig. 34.3) links various extracellular ligands to cellular proliferation, differentiation, and apoptosis pathways through a signal transduction cascade mediated by kinases. Dysregulation of this pathway is a common feature to oncogenesis in a variety of malignancies including melanoma. Investigations by The Cancer Genome Atlas (TCGA) network demonstrated a critical role of the MAPK/ERK pathway with the majority of subtypes exhibiting heavily dysregulated ERK signaling, although less predominantly in acral and mucosal melanoma [40, 41]. Mutations of the MAPK/ERK pathway are almost always mutually exclusive. This is due to the fact that once MAPK/ERK pathway activation is acquired through a mutation, there is no selective pressure for additional mutations in the same pathway. Both benign nevi and melanomas have activation of MAPK/ERK pathway suggesting that dysregulation of additional mutations in other pathways is required for oncogenesis [41, 42, 72].

RAS Family

In the presence of ligand, membrane receptors change conformation and coordinate binding factors and guanine nucleotide exchange factors to arm the Ras family of proteins (H-Ras, K-Ras, and N-Ras) with GTP. Ras-GTP is able to activate the downstream Raf family members until the innate hydrolytic ability produces GDP. In addition, Ras-GTP is able to act on PI3K along with receptor tyrosine kinases (RTKs) (see below), joining two pathways critical for proliferation and survival [73]. Mutations in RAS typically interfere with the ability to switch to the inactive form. In the TCGA study, 28% of cutaneous melanomas had *NRAS* mutations, and nearly all were in the hot spot regions of

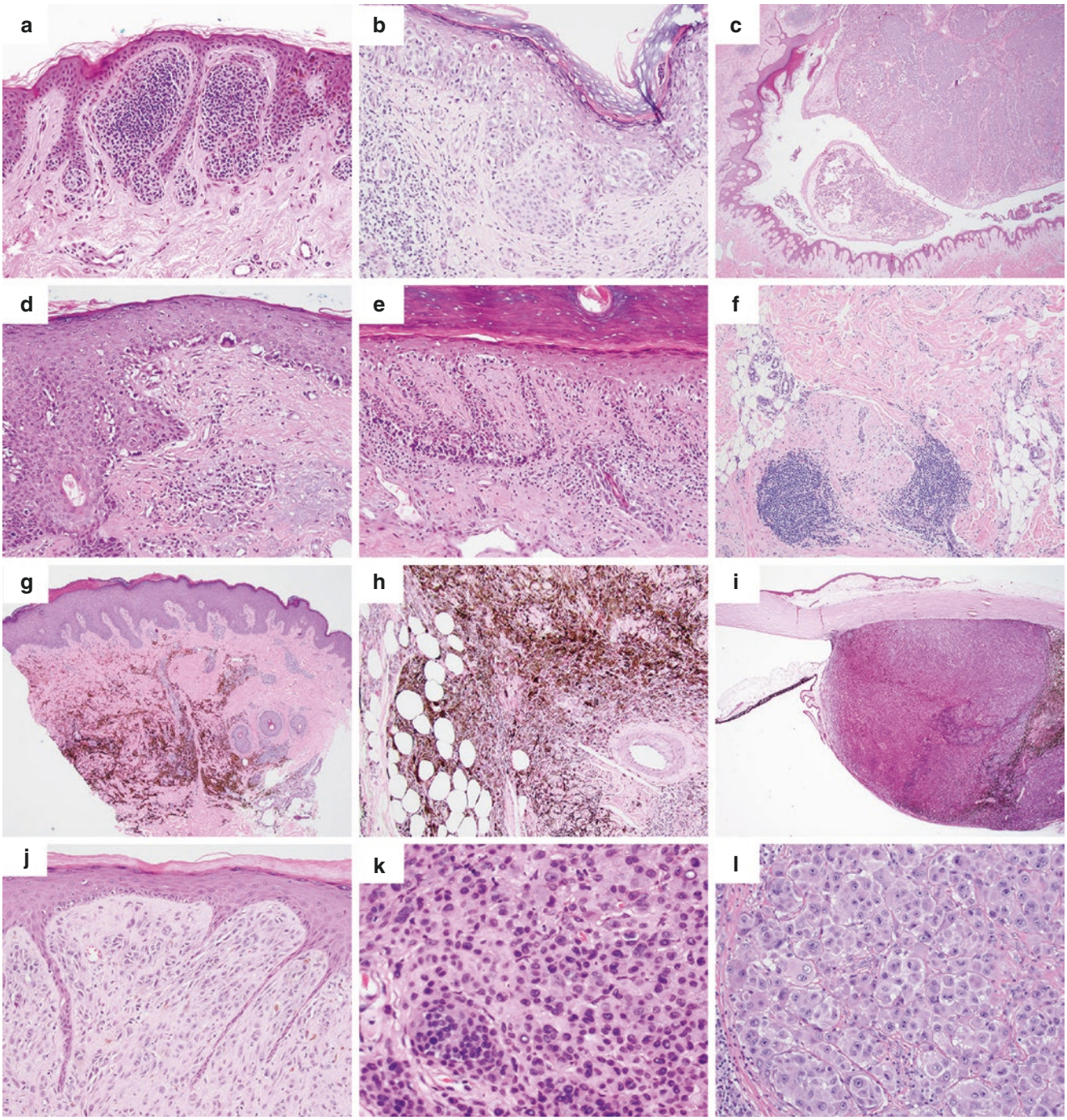


Fig. 34.2 Representative images from H&E-stained sections displaying main morphological patterns of benign and malignant melanocytic lesions. The **first row** (**a**–**c**) exhibits lesions associated with nested and pagetoid pattern and *BRAF/NRAS* mutations: (**a**) common acquired nevus (CAN), (**b**) superficial spreading melanoma (SSM), and (**c**) nodular melanoma (NM). The **second row** (**d**–**f**) shows lesions with predominant lentiginous/spindle cell pattern that are more often associated with *KIT*/*NRAS*/*NF1* aberrations: (**d**) lentigo maligna melanoma (LMM), (**e**) acral lentiginous melanoma (ALM), and (**f**) desmoplastic melanoma (DM) (in this case, the pure variant). The **third row** (**g**–**i**) is composed of lesions with blue nevus-like morphology that mainly harbor *GNAQ*/*GNA11* mutations: (**g**) blue nevus (BN), (**h**) blue nevus-like melanoma (BNLM), and (**i**) uveal melanoma (UM). Finally, the **fourth row** (**j**–**l**) shows representative lesions with spitzoid morphology which are associated with various genetic aberrations including *HRAS* mutations/amplifications (see Fig. 34.7), BAP1 loss as well as *BRAF*, and other kinase-coding gene fusions: (**j**) Spitz nevus, (**k**) melanocytic proliferation with BAP1 loss, and (**l**) spitzoid melanoma

Table 34.1a, b Melanocytic lesions associated (**a**) and not associated (**b**) with epidermal/epithelial structures with their usual anatomical location, microscopic morphologic features, and most common genomic aberrations

(a) Epidermal−/epithelial-associated melanocytic lesions			
Type	**Solar damage—anat. location**	**Main morphologic pattern**	**Main genomic aberrations**
CAN, BLMP, SSM, NM [12, 40–46]	Intermittent—Trunk, proximal extremities	Epithelioid, nested and pagetoid RGP, epithelioid/spindled VGP	*BRAF* p.V600E[a], *NRAS*[a], *TERT*[d], *CDKN2A*[c], *TP53*[c], *PTEN*[c] *BRAF* p.V600E[a]/NRAS[a] and *BAP1*[c] (BLMP)
LMM, NM [40–42, 47]	Chronic—Head and neck, upper trunk	Lentiginous RGP, epithelioid/spindle cell VGP	*NRAS*[a], *KIT*[a], *BRAF* p.V600 K[a], *RAC1*[a], *TERT*[d], *NF1*[c], *CDKN2A*[c], *TP53*[c], *PTEN*[c]
Desmoplastic melanoma [40, 41, 47–49]			*NRAS*[a], *PIK3CA*[a], *PTPN11*[a], *ERBB2*[b], *MAPK1/3 K1*[b], *BRAF*[b], *EGFR*[b], *MET*[b], *TERT*[d], *NFKBIE*[d], *NF1*[c], G691S *RET*[#]
ALN, ALM, NM [40, 41, 50–52]	None—Acral sites	Lentiginous RGP with prominent adnexal extension, epithelioid/spindle cell VGP	*KIT*[a], *NRAS*[a], *HRAS*[a], *KRAS*[a], *BRAF*[a], *CCND1*[b], *GAB2*[b], *TERT*[d], *CDKN2A*[c], *NF1*[c], *NTRK3*, and *ALK* rearrangements
Mucosal melanoma [40, 41, 52–54]	None—Mucosal sites	Lentiginous RGP, epithelioid/spindle cell VGP	*KIT*[a], *NRAS*[a], *KRAS*[a], *BRAF*[a], *NF1*[a], *CCND1*[b], *CDK4*[b], *MDM2*[b], *CDKN2A*[c], *SF3B1*[e]
Spitz nevus and spitzoid melanoma [55–57]	None/unknown—Head and neck, trunk, extremities	Epithelioid and spindle cells, pink cytoplasm, large nuclei, vesicular chromatin, macronucleoli, spindle cells with/without pigment	*HRAS*[a/b], *CDKN2A*[c], *ALK*, *RET*, *MET*, *ROS1*, *BRAF*, and *NTRK1/3* rearrangements
DPN, DPN-like melanoma [58]	Intermittent—Head and neck, upper trunk	Dermal based, inverted triangle, plexiform pattern, epithelioid, and spindle cells with finely pigmented cytoplasm and accompanying macrophages	*NRAS*[a], *BRAF* p.V600[a], *MAP2K1*[a], *CTNNB1*[a], APC[c]
(b) Non-epidermal−/epithelial-associated melanocytic lesions			
Type	**Solar damage—anat. location**	**Main morphologic pattern**	**Main genomic aberrations**
Melanoma arising from congenital nevus [59]	None/UNK-Trunk, head and neck	Large nodules of malignant epithelioid/spindle cells arising in the dermal component of a congenital nevus	*NRAS*[a], *BRAF* p.V600E[a], *BRAF* rearrangements
BN, BN-like melanoma [40, 41, 60–62]	None-Head and neck, trunk, extremities	Dermal based, heavily pigmented or amelanotic tumors composed of spindle cells	*GNAQ*[a], *GNA11*[a], *CYSLTR2*[a], *BAP1*[c], *SF3B1*[c], *EIF1AX*[e] *PLCB4*[a]
Uveal melanoma and CNS melanocytomas [63–66]	None/UNK (CSD?) iris, choroid, and CNS structures	Epithelioid, spindled and mixed tumors usually pigmented	(uveal melanoma)
PEM, PEM-like melanoma [67, 68]	None-Head and neck, trunk, and extremities	Heavily pigmented epithelioid and plump spindle cells, large nuclei with blue nucleoli	*PRKAR1A*[c] and *BRAF*[a], *PRKCA* rearrangements
Clear-cell sarcoma [69–71]	None-Deep soft tissue (tendons, fascia) of extremities	Proliferation of plump spindle cell with large nuclei, prominent nucleoli with occasional giant cells with fibrous bands	*EWSR1-ATF1* and *EWSR1-CREB1* fusions

Abbreviations: *Anat.* anatomical, *CAN* common acquired melanocytic nevus, *SSM* superficial spreading melanoma, *BLMP* BAP1-loss melanocytic proliferation, *NM* nodular melanoma, *ISD* intermittently sun damaged, *RGP* radial growth phase, *VGP* vertical growth phase, *LMM* lentigo maligna melanoma, *DPN* deep penetrating nevus, *BN* blue nevus, *CNS* central nervous system, *CSD* chronic sun damage, *UNK* unknown, *PEM* pigmented epithelioid melanocytoma
Superscripts: [a]gain-of-function mutation, [b]amplification, [c]loss-of-function mutation, [d]promoter methylation, [e]change-of-function mutation

codons 12, 13, and 61 [41]. Similar activating mutations are seen in Rho GTPase RAC1, which are only rarely seen in melanoma [41, 74, 75]. Mutations in GTPase-activating factors (GAFs) like neurofibromin (NF1), which help catalyze Ras-mediated GTP→GDP hydrolysis and decrease the amount of time in the Ras-GTP active state, also lead to prolonged signaling by Ras-GTP [76]. In the TCGA cutaneous melanoma classification, one subtype with *NF1* mutations represented 14% of the total cutaneous melanomas and 39% of non-hot spot RAS/BRAF mutations [41].

Morphologic Correlates of Lesions Harboring RAS Family Mutations Both *NRAS* and *HRAS* can be mutated in nevi, suggesting the need for additional genomic alterations for oncogenesis. *NRAS* mutations are present in approximately 80% of congenital nevi but are rarely seen in acquired

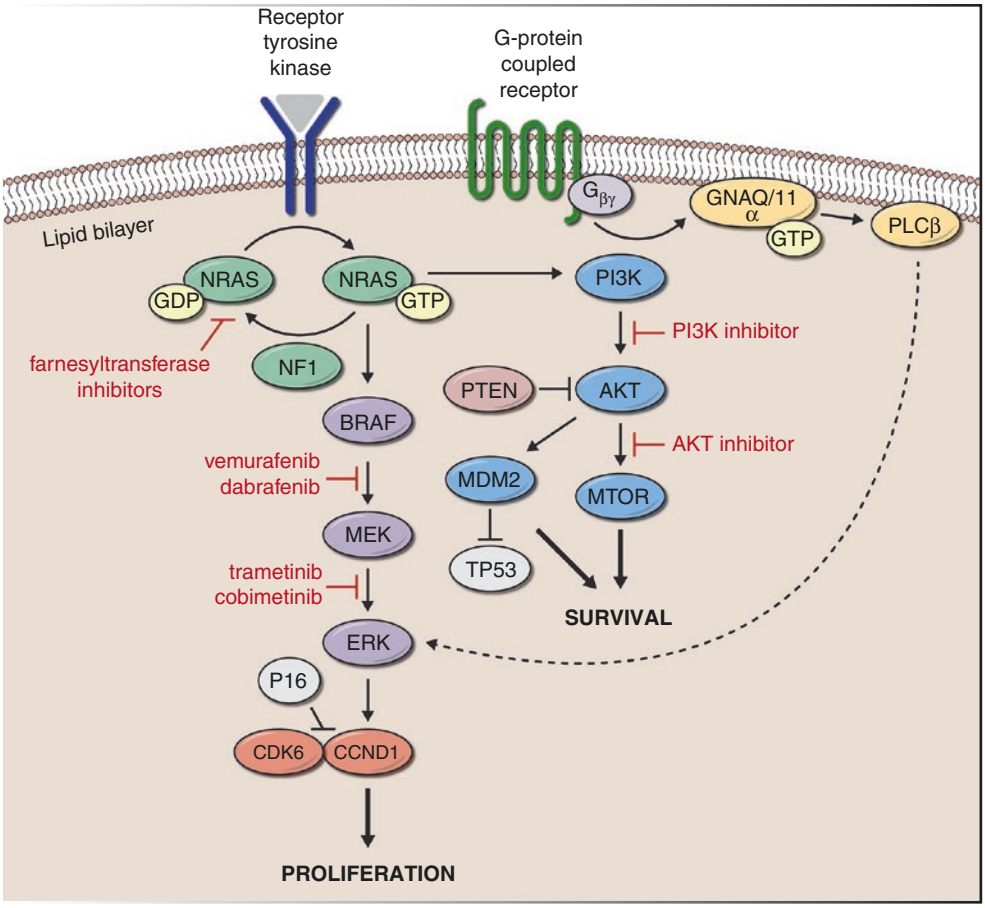


Fig. 34.3 Pathway diagram shows a schematic of two major pathways that initiate and maintain melanoma, the ERK and AKT/PI3K pathways. Also illustrated are some current and investigational strategies for inhibiting these pathways

nevi [43]. Additionally, activating *HRAS* mutations were identified in 14% of Spitz nevi (especially the desmoplastic variant [55]) and in atypical spitzoid tumors with indolent biologic behavior [77, 78]. Interestingly, *HRAS* mutations have not been identified in spitzoid melanoma [79] but have been identified rarely in non-spitzoid malignant lesions [41]. Melanomas with frequent *NF1* mutations include the lentigo maligna (LMM) and desmoplastic melanoma (DM) variants [47, 48], which have been associated with older age of a strong UVR damage signature (chronic sun damage) with a subset of these lesions being related to poor prognosis [80]. In contrast, SSM and NM arising in non-chronically sun-damaged skin tend to show BRAF V600 and/or *NRAS* mutations [42, 81, 82]. Interestingly, while pure-type DM lesions tend to cluster into the *NF1*-mutated group, mixed-type DM lesions (less than 90% classic DM morphology) appear to harbor a mutational profile more akin to SSM and NM [47–49]. Nevertheless, both DM subtypes (pure and mixed) tend to show a somewhat specific polymorphism (G691S) of the *RET* gene [49, 83].

Therapeutic Interventions (RAS Family) Multiple avenue studies that aimed to inhibit the Ras family have been undertaken but without significant therapeutic success, at least as compared to downstream inhibition of the Raf or Mek kinases [84]. Small-molecule inhibition of the Ras family can be accomplished with farnesyl transferase inhibitors, which block the membrane localization of Ras and suppress downstream pathway activation (Fig. 34.3) [85]. A Phase II clinical trial with the farnesyl transferase inhibitor tipifarnib in patients with metastatic melanoma demonstrated inhibition of Ras farnesylation in tumors but without significant clinical response [86]. Based on unique bioinformatic insights, vertical inhibition of *NRAS*-mutated melanoma with simultaneous inhibition of both MEK and CDK4 (which are in the same pathway) has shown promising results [87–89].

RAF Family

The mutational landscape of the Raf family of kinases (ARAF, BRAF, CRAF) in melanoma is dominated by mutations in the *BRAF* gene. The BRAF p.V600E mutation is the most common mutation in melanoma [41, 90, 91]; a single base-pair change (c.1799 T > A) results in the substitution of valine (V) with glutamic acid (E) at the 600 position (Fig. 34.4). This creates a larger residue with a negative charge between regulatory phosphorylation sites at Thr598 and Ser601 that acts as a phosphomimetic. Other activating mutations at this hot spot are V600 K (created by a double substitution), V600D, V600R, and K601. The TCGA sequencing data showed that 87% of

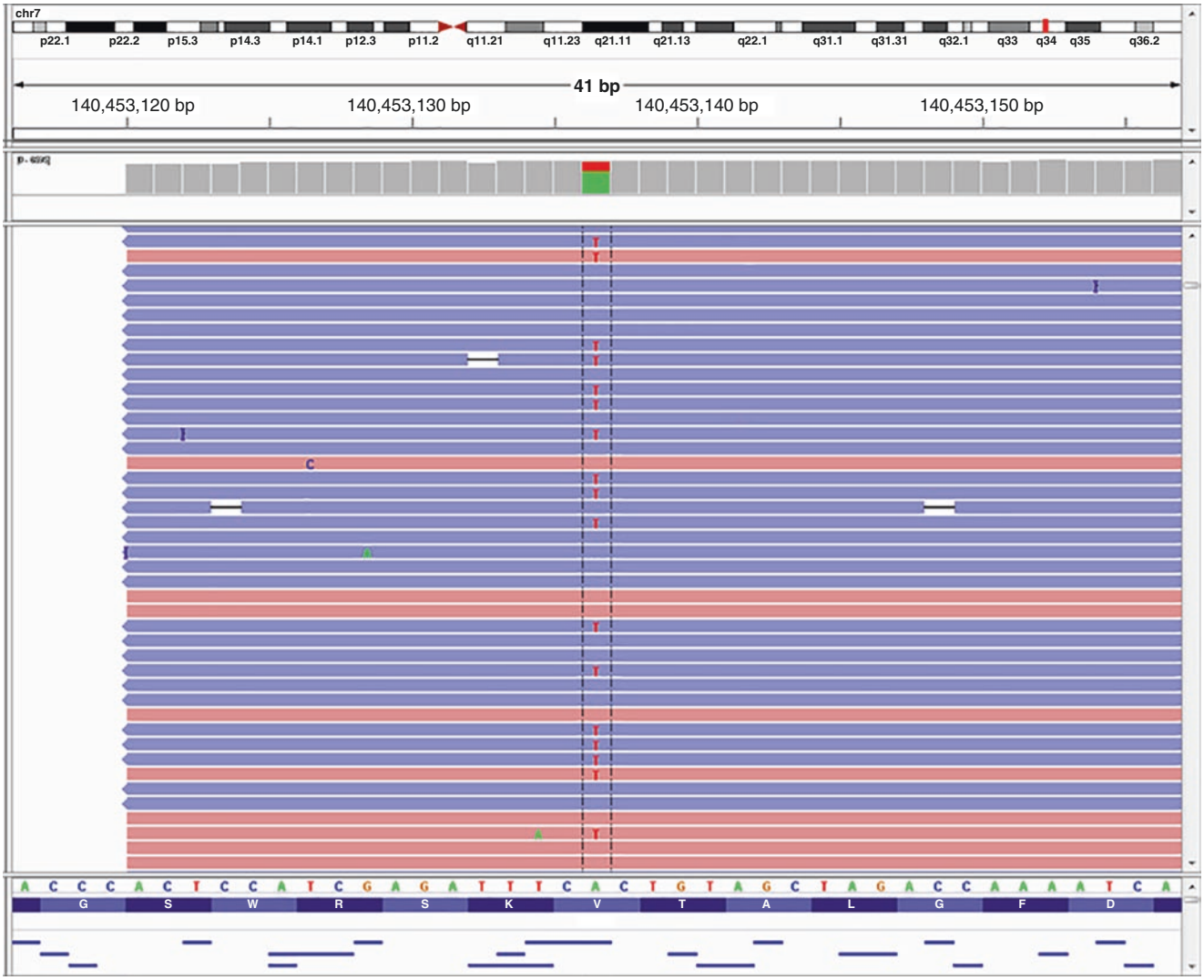


Fig. 34.4 Example of a next-generation sequencing (NGS) graphic demonstrating a V600E mutation in the *BRAF* gene

BRAF mutations in cutaneous melanoma were at the V600 codon with 10% being the V600 K mutation [41, 92]. Mutations within the BRAF kinase phosphate-binding loop also lead to constitutive activity [92, 93].

Melanomas with *BRAF* hot spot mutations are very unlikely to have hot spot *NRAS*, *KIT*, *or NF1* mutations as these alterations tend to be mutually exclusive [41, 44]. The exception is that the weaker activating BRAF mutations in exon 11 will often be coupled with hot spot mutations in the RAS family—*NRAS*, *HRAS*, and *KRAS* [94]. In addition to the abovementioned mutations, recent findings pointed out to oncogenic chromosomal fusions and rearrangements involving *BRAF* as driving mechanisms, particularly in the so-called "triple negative" or "triple wild-type" (triple WT) melanoma tumors [50, 95, 96] with this type of profile being also described in spitzoid lesions [56].

Morphologic Correlates of Lesions Harboring RAF Family Mutations The presence of *BRAF* mutations in melanocytic proliferations is not specific for melanoma, as the majority of acquired nevi (82%) carry BRAF p.V600 mutations, with a similar proportion seen in intradermal and dysplastic nevi [44, 45, 97] but less commonly in congenital nevi [43]. BRAF p.V600 mutations are usually found in BAP1-inactivated nevi (the so-called BAPomas) [46] and in melanocytic lesions associated with Wnt/β-catenin aberrations (deep penetrating nevi [DPN] and DPN-like melanoma; see below in Wnt pathway discussion) [58].

The superficial SSM variant is the prototypical example of a BRAF p.V600-mutated malignant melanocytic tumor [42, 98, 99]. These lesions can be seen arising from pre-existing *BRAF*-mutated melanocytic nevi [100] in young to middle-aged adults. In contrast, a subset of tumors with BRAF p.V600K tend to appear de novo and are associated with older patients showing cumulative chronic actinic damage and are more closely related to the LMM phenotype [101, 102]. Interestingly, both *NRAS* and *BRAF* mutations

are rare in melanomas from black patients [103], and *BRAF* mutations are uncommon in mucosal melanomas [40].

Therapeutic Interventions (RAF Family) Several small-molecule inhibitors of BRAF are FDA-approved for the treatment of metastatic melanoma with *BRAF* mutation. Vemurafenib was the first FDA-approved drug, with a 53% response rate and median progression-free survival of 6.8 months [104, 105]. Dabrafenib newer-generation agent associated with slightly less adverse events was subsequently approved for the treatment of BRAF p.V600E or p.V600K mutant metastatic melanoma [106]. Clinical trials are underway with several other wild-type and mutant RAF inhibitors, including RAF265 [107–109] and XL281 [109, 110]. The development of drug resistance limits the effectiveness of current regimens of BRAF inhibitors, prompting research into overcoming resistance with alternate dosing regimens and combination therapies [111].

MEK Family

MEK1 and MEK2 are both phosphorylation targets for BRAF and kinases for the ERK family. Mutations in this step of the signal transduction cascade are less common than in the Ras or Raf families, occurring in 6–8% of melanomas [40, 112].

Therapeutic Interventions (MEK Family) Inhibitors of the MEK family are attractive given the numerous upstream activators; mutually exclusive mutations of RAS and RAF families, KIT, and NF1 all signal through MEK1/MEK2 (Fig. 34.3) [41, 113, 114]. The small-molecule inhibitor trametinib is a selective MEK1/MEK2 inhibitor and was superior to chemotherapy in BRAF p.V600E-/K-mutated melanoma [115] and in combination with the BRAF inhibitor dabrafenib was superior to dabrafenib alone [116]. Cobimetinib is another MEK1/MEK2 small-molecule inhibitor approved in combination with BRAF inhibitor vemurafenib in BRAF p.V600E-/K-mutated melanomas [117]. Resistance eventually arises, partly mediated by p21-activated kinases (PAKs), which arise as potential targets in resistant cases [118].

ERK Family

The ERK family members, ERK1 and ERK2, are also protein kinases with both cytoplasmic and nuclear targets. Mutations in the genes coding these proteins are relatively rare, but 11% of melanomas demonstrate increases in copy number [40].

GNAQ and GNA11

GNAQ and GNA11 are alpha subunits of heterotrimeric G proteins. Similar to Ras, in the presence of ligand-receptor binding, G protein alpha subunits bind GTP and become activated until intrinsic hydrolase activity converts GTP→GDP + Pi. The activated form of the G protein alpha subunit dissociates from the beta/gamma subunits; each then regulates downstream messengers. The subsequent activation of phospholipase C leads to the production of diacylglycerol (DAG) activating protein kinase C, which in turn activates both the MEK and ERK family and ultimately leads to ERK pathway activity, thus merging into the MAPK/ERK pathway (Fig. 34.3).

GNAQ mutations include exon 5 Q209L/Q209P, which occurs in a RAS-like domain. *GNA11* mutations also involve Q209L in a homologous domain or R183 mutations [40, 41, 63, 64]. Mutations in the Ras-like domain are associated with decreased GTPase hydrolytic activity, stabilizing the active form. Both in vitro and in vivo experiments demonstrate that these mutant forms are associated with activation of the ERK pathway [119]. Although their place in the signaling cascade is thought to be similar, there may be prognostic differences between *GNAQ* and *GNA11* mutations. A mutation in *GNA11* was seen in 32% of primary uveal melanoma and in 57% of the metastases, whereas a *GNAQ* mutation is seen in 45% of uveal melanoma but in 22% of the metastases [64]. However, a direct prognostic role has not been clearly demonstrated in subsequent studies.

Morphologic Correlates of Lesions Harboring *GNAQ* and *GNA11* Mutations Mutations in *GNAQ* and *GNA11* have been identified in the blue nevus, blue nevus-like, and uveal melanocytic proliferations, both benign and malignant including up to 80% of uveal melanoma as well as the central nervous system (CNS) melanocytomas [63–66]. *GNAQ* and *GNA11* mutations are mutually exclusive with hot spot mutations in *NRAS* and *BRAF*. Of note, a subset of blue nevi and uveal melanomas will rarely show activating *CYSLTR2* gene mutations [60] and, similarly to uveal melanomas, a subset of blue nevus-like melanomas can also exhibit *BAP1* and *SFB31* aberrations [61, 62]. In blue nevus-like melanomas, the presence of BAP1 inactivation is associated with poor prognosis [62, 120].

Therapeutic Interventions (GNAQ and GNA11) A Phase III clinical trial comparing selumetinib, a MEK1/MEK2 inhibitor, and dacarbazine in uveal melanoma with *GNAQ*/*GNA11* mutations showed only modest improvement of progression-free survival at the cost of high rate of adverse events [121].

PI3K-AKT Pathway

The PI3K/AKT/mTOR pathway (Fig. 34.3) is a common signaling pathway which influences proliferation, cell

migration, and survival. Although mutations in the main pathway components are uncommon relative to those of the ERK pathway, the PI3K pathway is constitutively active in the majority of melanomas [40, 122], due to the cross talk at multiple levels in the signaling cascade. Inhibition of the PI3K-AKT pathway is therapeutically relevant and made more so by its role in mediating mechanisms of ERK pathway resistance. The PI3K family is activated by membrane receptor tyrosine kinases (RTKs, see RTK section below) and can also be activated by Ras family members; this activation leads to the phosphorylation of phosphatidylinositol 4,5-bisphosphate (PIP_2)-- > phosphatidylinositol 3,4,5-trisphosphate (PIP_3). The PIP_3 molecule then binds to the kinase AKT, which leads to the recruitment of the protein-lipid complex to the plasma membrane for phosphorylation by PDK1 and the mTORC2 complex [122]. The different isoforms of the PI3K catalytic subunit include p110α, p110β, and p110δ (i.e., PI3KCA, PI3KCB, and PI3KCD).

Targets of phospho-AKT include NF-kB, BAD, FKHR-L1, and GSK3B, among others, with diffuse roles in proliferation, survival, and migration. In addition, PIP3 activates other proteins with PH domains and leads to the activation of Rac and Tec families [122]. Two thirds of melanomas have PI3K activity, which can be documented by the expression of phospho-AKT, while inhibition of AKT in vitro leads to apoptosis [40, 123, 124]. Despite the prevalence of an overactive PI3K pathway, mutations in the PI3K family are relatively rare in melanoma, occurring in 3–5% of lesions [40, 41, 125]. Similarly, mutations in *AKT1* and *AKT3* occur in only approximately 1% of melanomas [40, 41]. However, in tumors with acquired resistance to BRAF inhibitor therapy, whole-exome sequencing has demonstrated that 22% of them had PI3K-AKT pathway upregulating mutations [126], suggesting that such mutations promote resistance to these agents.

PTEN The tumor suppressor PTEN is a lipid phosphatase that hydrolyzes the 3′ phosphate group of PIP_3 converting it to PIP_2, preventing the activation and membrane localization of AKT. Loss of PTEN activity is an important mechanism in PI3K-AKT pathway activation and is relatively common in melanoma, occurring via several mechanisms. In one series, 46% of melanoma had *PTEN* gene aberrancies, including 31% with homozygous deletion, 8% with substitutions or insertions/deletions (indels), and 7% with structural variants [40]. Alterations in *PTEN* are mutually exclusive with *NRAS* mutations, which are sufficient for both ERK and PI3K-AKT pathway mutations. *PTEN* alterations often coexist with *BRAF* mutations, which only affect the ERK pathway [40]. The combination of *BRAF* mutation and PTEN inactivation is predictive of metastasis to the brain [127–129]. Furthermore, PTEN loss is also associated with reduced tumor-infiltrating lymphocytes and lower response rates to immune checkpoint inhibitors (see below in the immunomodulatory therapy section) [130].

Therapeutic Interventions of PI3K-AKT Pathway The downstream effectors of the PI3K-AKT pathway are an attractive target for melanoma therapy due to the prevalence of pathway activation whether it stems from *KIT* or *NRAS* mutation or *PTEN* loss of function. In addition, there is significant evidence that the PI3K-AKT pathway is involved in BRAF or combined with BRAF/MEK inhibitor resistance [126, 131–134]. Antagonistic agents against this pathway can be classified into single target such as anti-PI3K, AKT, or mTORC1 or dual-target inhibitors aimed at PI3K/AKT and at mTORC1/mTORC2. PI3K inhibitors include pan-isoform agents and isoform-specific agents with the ones targeting the p110β isoform being a more attractive target for melanoma that shows loss of PTEN [135]. Similarly, targeting the AKT3 isoform of AKT appears as an attractive prospective target for melanomas with PTEN loss [123, 124, 136]. The so-called rapalogs, inhibitors of mTOR such as rapamycin and rapamycin analogs, have shown minimal activity in melanoma [137], which is likely due to their selectivity to mTORC1 triggering a paradoxical negative feedback favoring activation of PI3K, AKT, and ERK [138]. Thus, dual mTORC1/mTORC2 antagonism inhibits cancer cell growth and survival more effectively when compared to inhibition of mTORC1 only [139]. Similarly, melanoma cell lines appear much more affected by dual PI3K/mTOR inhibition in comparison with agents that target single components [140, 141]. Finally, although simultaneous targeting of ERK and PI3K has resulted in significant toxicity [142], work continues, as this is an attractive therapeutic coupling.

Membrane-Associated Receptor Tyrosine Kinases (RTKs)

Binding of ligand to extracellular receptor tyrosine kinases (RTKs) leads to activation of the lipid kinase PI3K as well as activation of Ras family members.

Epithelial Growth Factor Receptor (EGFR)

Although epithelial growth factor receptor (EGFR) and other growth factor mutations feature prominently in the mechanisms of oncogenesis in other malignancies, relatively low expression of EGFR is noted in the majority of melanoma lesions [143], and the therapeutic efficacy of targeted inhibitors is limited [144]. In fact, in advanced melanoma tumors, it appears that the *EGFR* gene is indirectly

downregulated as a downstream effect from activation/hypomethylation of the *TBC1D16* gene (see below in the epigenetics section) [145].

KIT (CD117)

KIT is the most frequently mutated RTK in melanoma. Copy number amplification containing the *KIT* locus (4q12) is often coexistent with a mutation. The most common point mutations, L576P and V559A of exon 11 and K642E of exon 13, disable the inhibitory portion of KIT [40, 41]. In the TCGA cutaneous melanoma classification, *KIT* mutants represented a significant portion of the so-called "triple wild-type" (triple WT) subtype, that is, without *BRAF*, *NRAS*, or *NF1* mutations. This is likely due to KIT's ability to activate the ERK pathway at the level of the Ras family [41].

Morphologic Correlates of Lesions Harboring *KIT* Mutations Lesions with *KIT* mutations tend to show a lentiginous in situ component and RGP and are seen in 30–39% of mucosal melanomas, 15–36% of ALM, 20% of melanomas of the nail apparatus (related to the ALM variant), and up to 28% of LMM and melanomas arising from chronically sun-damaged skin [42, 51–53, 146]. On the other hand, these mutations are rarely seen in melanomas arising from intermittently sun-damaged skin, which often show a nested/pagetoid RPG (SSM variant) [147]. *KIT* mutations are rare in nevi [97, 148] even in examples from acral and mucosal sites with these lesions mainly showing BRAF p.V600E mutations [51, 54]. Thus, it would seem that ALM lesions do not arise from pre-existing nevi but de novo and are often multifocal, apparently taking origin from a field effect of morphologically normal-appearing melanocytes harboring amplifications of multiple genes [149].

Therapeutic Interventions (*KIT* Mutations) A wide range of small-molecule RTK inhibitors is available owing to the important role of this pathway in carcinogenesis. Early clinical trials for KIT inhibitor imatinib in melanoma did not require *KIT* mutation or amplification testing on the basis for enrollment; only a few responders were identified [150–153]. High KIT protein levels by immunohistochemistry were not predictive of response [151, 154]. More recent clinical trials have had several partial responders, with nearly all having lesions bearing activating point mutations (mainly L576P and V559A of exon 11 and K642E of exon 13) instead of *KIT* amplifications [155–158]. Thus, responses tend to be seen primarily in melanomas with activating *KIT* mutations known to be responsive to KIT inhibitors, including imatinib, in other neoplasms such as gastrointestinal stromal tumor (GIST). Association of a *KIT*-activating mutation with *KIT* copy number gains increases the likelihood of a response to therapy [155, 156, 158, 159]. Unfortunately, compared to GIST [160], responses to KIT inhibition in melanoma are less uniform and less predictable.

WNT Pathway

This pathway is crucial in melanocyte development from migration of neural crest precursors to survival, differentiation, and maintenance of melanoblasts through MITF and other factors [161, 162]. There are three main Wnt signaling pathways: (1) the canonical pathway that involves the Frizzled (FZD) receptor and β-catenin; (2) the noncanonical pathway, which is mediated by PKC/Ca^{+2}; and (3) the planar cell polarity pathway that is associated with Jun kinase (JNK). Only the canonical and noncanonical pathways have been seen to be implicated in melanoma tumorigenesis. While the former is usually associated with the Wnt1 and Wnt3A ligands and with increased differentiation, transformation, and proliferation (mainly RGP attributes), the latter is associated with Wnt5A and with decreased immunogenicity, metastasis, and mesenchymal invasion (hallmarks of the VGP) [161, 163]. Interestingly, recent findings suggest that the so-called deep penetrating growth pattern in both nevi and melanomas is correlated with the presence of activating mutations/fusions involving both MAPK components (*BRAF*, *HRAS*, and *MAP2K1*) and β-catenin (mostly *CTNNB1*) [58].

Activation of Wnt, particularly Wnt5A, has been correlated with resistance to BRAF inhibitor agents [164, 165]. Furthermore, activation of the Wnt canonical pathway in melanoma tumors appears to prevent antitumoral immunity [166] and it is one of the mechanisms of resistance to the newly developed immune-checkpoint blockage therapies (see below in the immunomodulatory therapy section) [166, 167]. Thus, targeting the Wnt pathway appears as an attractive way to increase the efficacy of these two therapeutic interventions. However, due to its complexity, plasticity, and the active involvement in a wide variety of homeostatic processes, modulating this pathway remains quite a challenge [163].

TERT Promoter Mutations

The reverse transcriptase component of telomerase (TERT, hTERT) is a large multicomponent ribonucleoprotein polymerase coded by the *TERT* gene (5p15.33) that is in charge of telomere maintenance. While *TERT* is silenced in the majority of normal cells in which progressive shortening telomeres result in the so-called replicative senescence state, TERT is very active in highly replicating cell populations such as epithelial, hematopoietic, and others [168]. Malignant cells exploit this mechanism to reach immortalization [169].

Two types of mutually exclusive *TERT* promoter mutations (TPMs) were characterized in around 71% of melanomas, which activate *TERT* promoter and *TERT* gene transcription via generation of a de novo binding site for the

ETS family of transcription factors [170, 171]. Following this discovery, TPMs were detected in over 50 different malignant tumor types [168].

In melanoma tumorigenesis, TPMs appear relatively early in the tumor progression [100] and act in a two-step fashion by enabling both cancer cell immortalizations and also by promoting genomic instability [100, 172]. Although TPMs are seen relatively more frequent in *BRAF*-mutated melanomas (SSM type) [171, 173], ALMs, mucosal melanomas (MucM) [174, 175], and possibly spitzoid melanomas [176] also show TPMs. Melanomas and spitzoid lesions with TPMs are associated with poor prognosis [175, 176]. Preliminary studies are ongoing in search for an effective pharmacologic *TERT* repressor [177].

Melanoma Epigenetics

Epigenetic alterations constitute all events that alter DNA expression without inherent changes in its sequence and may include DNA methylation/demethylation, posttranslational modifications (PTMs) of histone tails/remodeling of nucleosome complexes, as well as changes in the chromatin-remodeling complexes, Polycomb Repressive Complex 2 (PRC2), and long and short noncoding RNAs (lcRNA and miRNA).

DNA Methylation/Demethylation DNA methylation is carried out in the cytosine-guanine (CpG) dinucleotides, and several genes have been characterized as being hypermethylated in melanoma, including *MGMT*, *CDKN2A*, *PTEN*, *RAR-β2*, and *RASSF1A*, which are related with poor prognosis [178]. In fact, it appears that depletion of the DNA methylation maintenance methyltransferase (DNMT1) appears to be related in diffuse activation of the so-called "cancer-germline" gene signature [179]. In addition, the CpG island methylator phenotype (CIMP) in melanoma tends to be associated with somatic mutations of chromatin-remodeling genes (*ARID2* and *IDH1*), among others [41, 178]. On the other hand, both DNA demethylation and hypomethylation also occur in melanoma tumors. While the DNA demethylation involves the 5-hydroxymethyl-cytosine (5hmC) intermediary and is associated with *TET* and *IDH2* downregulation [180] (5hmC can be labeled by immunohistochemistry and correlates with tumor progression, see section on epigenetic markers identified by IHC), hypomethylation and activation affect mainly melanoma metastases involving the *TBC1D16* gene and are associated with poor prognosis [145].

Histone PTMs Histone "marks"/modifications found in the tails of histones are involved in nucleosome stability and are ruled by dialectical interactions between lysine acetyltransferases (KATs) and histone deacetylases (HDACs) [181]. HDAC activity may have relevance since the use of HDAC inhibitors results in the upregulation of PD-L1 and PD-L2 [182]. Animal models of melanoma initiation also describe the interaction between the acetylated at lysine 27 form of H3 (H3K27ac) mark with the Sox10/MITF and AP-1/TEAD transcription factors as a molecular switch from a proliferative-predominant to a migratory-predominant status with the latter showing increased resistance to MAPK therapeutics [183].

Chromatin-Remodeling Complexes and PRC2 The components of the SWI/SNF chromatin-remodeling complexes are composed of the BAF (BRM/BRG1 and ARID1A/ARID1B) and pBAF (BRAG1, ARID2, and BRD7) cores with one of the targets of these complexes being the Polycomb Repressive Complexes (PRC1 and PRC2). Genes coding these complexes are mutated in up to 30% of melanoma tumors, usually when they reach the invasive phase [41, 74, 100]. PRC2 is formed by the complex of EZH1 or EZH2 complemented by SUZ12 and EED components with the main PRC2 intermediary being the trimethylated at lysine 27 form of H3 (H3K27me3) histone mark. *EZH2* shows gain-of-function mutations and amplifications in a subset of melanomas with aggressive and poor prognostic features [184, 185]. Increased EZH2 activity leading to E-cadherin (*CDH1*) gene silencing is likely one of the mechanisms of PD-1 therapy resistance [186]. Increased EZH2 activity appears to be prevalent in a multitude of entities; multiple antagonistic compounds are being developed and tested [187]. Interestingly, the other components of PRC2, namely, EED and SUZ12, appear to be disabled in a fraction of tumors [41] which also appears to be a feature of malignant peripheral nerve sheath tumors (MPNSTs) [188], further extending these entities overlapping features. This finding opens a new therapeutic avenue by targeting the bromodomain-containing protein 4 (BRD4), which results in SUZ12 loss and subsequent epigenetic switching from H3K27Me3 to H3K27Ac [189].

Noncoding RNAs These can be divided into long-coding (lcRNA) and micro (miRNA)-RNAs, with the most relevant component of this group being the so-called survival-associated melanoma-specific oncogenic IcRNA (SAMMSON) which lies in close proximity to *MITF* locus (3p13–14) and thus amplified in up to 10% of melanomas. Knocking down SAMMSON reduces viability of melanoma cells independently from *TP53* and *RAF* mutations, and combined therapy of an anti-SAMMSON and BRAF p.V600 (dabrafenib) inhibitors resulted in tumor growth arrest and apoptosis [190, 191]. On the other hand, profiling miRNAs in melanomas has some prognostic value in metastatic cases, particularly when involving CNS [192, 193]. In addition,

miRNAs could potentially be used in the so-called liquid biopsies as they can be found in the plasma of melanoma patients as exosomes [194, 195].

Proteomic and Genomic Applications for the Diagnosis of Melanocytic Lesions

The next section describes diagnostic tools based on genomic and subsequent proteomic changes present in melanoma and other melanocytic lesions, ranging from the simplest proteomic method, immunohistochemistry (IHC), to more complex genomic analytic methodologies.

Proteomics

Immunohistochemistry

BRAF V600E (VE1) and NRAS Q61L/Q61R

Detection of the BRAF p.V600E mutation using the VE1 mouse monoclonal antibody via immunohistochemistry is a sensitive and specific marker of mutation status, particularly for tumors with epithelioid morphology [196–198]. IHC can be helpful in cases for which DNA sequencing is not possible (decalcified or very scant specimens) (Fig. 34.5c, f) [199]. However, 13% of BRAF p.V600 mutant cutaneous melanomas have mutations other than V600E (chiefly V600K) (see above in the Raf pathway). Since immunohistochemical detection is blind to these alterations and mutations at other sites [200, 201], sequencing should be considered in metastatic melanoma, particularly in the BRAF p.V600E IHC-negative tumors which may still be eligible for targeted therapy. Immunohistochemistry for the protein products from *NRAS* mutations Q61L and Q61R appears to be specific and sensitive and could potentially provide with additional information in small specimens [202, 203].

CDKN2A ($p16^{INK4A}$)

IHC directed to the cyclin-dependent kinase (CDK) protein $p16^{INK4a}$ is mainly used as surrogate marker for *CDKN2A* loss (Figs. 34.5b, e, and 34.6c). Loss of this marker is a screening tool to stratify atypical spitzoid proliferations and determine which lesion deserves further evaluation with fluorescence in situ hybridization (FISH) and other methods [204] (see CGH and FISH sections below). Loss of $p16^{INK4a}$ expression in some cases could be helpful in the context of morphology and other factors to distinguish Spitz nevi from spitzoid melanomas [205, 206]. Of note, only the diffuse loss of $p16^{INK4A}$ expression should be interpreted as abnormal. Cell-to-cell variation in immunoreactivity creating a so-called "checkerboard" pattern is commonly seen in nevi and should not be counted as loss of expression. Some studies also suggest $p16^{INK4A}$ as a prognostic tool, particularly when combined with Ki-67 [207–209] and also as a marker of tumor progression [210].

BAP1 and Kinase Fusions in Spitzoid Tumors

Following the original description [12], melanocytic proliferations with BAP1 loss, colloquially known as "BAPomas" or "Wiesner's nevus," have been increasingly recognized not only because of their distinct morphologic features but also by loss of expression of BAP1, usually in a subset of lesional cells (Fig. 34.5a, d) [46]. Recognition of these lesions is important because it might indicate the presence of a syndromic germline *BAP1* mutation that is associated with multiple cutaneous and internal malignancies [11, 211, 212]. In addition, other melanocytic lesions, mostly blue nevus-like melanomas, have shown loss of BAP1 expression with corresponding somatic mutations [61, 62, 213]. A multitude of activating chromosomal fusions involving genes coding for kinases such as ALK, RET, ROS1, MET, NTRK1/NTRK3, and others were recently described occurring in atypical Spitz tumors, some of which correlate with morphologic features (especially ALK- and NTRK1-rearranged lesions). The majority of these kinases will be detectable by IHC (overexpression) (see below in FISH section) [56, 57].

PTEN and β-Catenin

Disruptions of the Wnt/β-catenin and AKT pathways are well-known events in melanoma. Loss of PTEN and increased expression of β-catenin detected in melanoma tumors by IHC often correlate with tumor progression, poor prognosis [128, 163, 214], and resistance to anti-PD-1 therapeutic strategies [130, 166, 167]. In addition, β-catenin and cyclin D1 nuclear expression have been recently described in melanocytic lesions with the deep penetrating growth pattern (both benign and malignant). These lesions are characterized by activation of MAPK pathway associated with activating mutations of *CTNNB1* gene [58].

Epigenetic Markers: 5-Hydroxymethylcytosine (5hmC) and PRC2

Both 5-hydroxymethylcytosine (5hmC) and the trimethylated at lysine 27 form of H3 (H3K27me3) are mediators of underlying epigenetic processes (see above in epigenetics section), both detectable by IHC. Loss of expression 5hmC correlates with morphologic atypia/dysplasia in melanocytic lesions [215] with potential uses for microstaging of nevoid/small-cell melanomas [216] as well as accurate discrimination between metastatic deposits and nodal capsular nevic

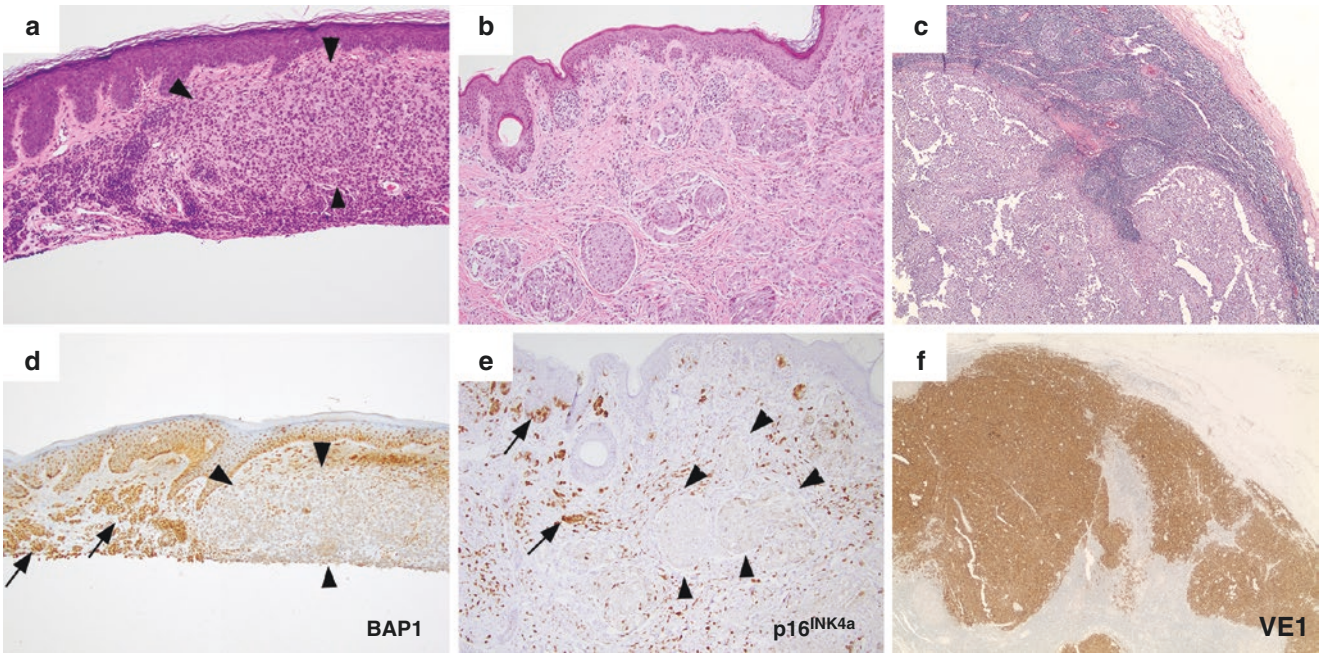


Fig. 34.5 Images of commonly used immunohistochemical stains in melanocytic lesions with their corresponding H&E-stained images. (**a**) and (**d**) show a BAP1-loss melanocytic lesion (same as the lesion depicted in Fig. 34.2k) with the classic combined nevus morphology present as an admixed banal melanocytic nevus (**a, arrows**) and a spitzoid epithelioid proliferation composed of medium-sized cells with abundant pink cytoplasm and nuclei with vesicular chromatin (**a, arrowheads**). The corresponding BAP1 IHC exhibits loss of BAP1 expression in the epithelioid component (**d, arrowheads**) with good internal control in the banal nevus component (**d, arrows**) and the epidermis. (**b**) and (**e**) show an atypical spitzoid lesion arising in a background of a congenital nevus; the corresponding immunohistochemistry for $p16^{INK4a}$ exhibits preservation of staining in the background congenital nevus (**e, arrows**) and complete loss of staining in the atypical spitzoid tumor (**e, arrowheads**). Finally, (**c**) and (**f**) show a lymph node melanoma metastasis exhibiting diffuse and strong immunoreactivity for VE1 (BRAF p.V600E)(**f**)

nests [217]. Furthermore, immunohistochemistry for 5hmC and H3K27me3 can assist in the challenging differentiation between proliferative nodules and melanoma arising in a congenital nevus with melanoma nodules exhibiting both complete loss of 5hmC [218] and reduction of H3K27me3 expression [219]. While increased expression of EZH2, HK27me3, and HK4me2 was detected both by IHC (particularly at the invasive front) [220] and by mass spectrometry in primary and metastatic melanomas, the opposite finding, particularly with H3K27me3 expression, was found in a considerable subset of melanomas [221].

Mass Spectrometry

Matrix-assisted laser desorption ionization (MALDI)-imaging mass spectrometry (IMS) analysis is capable of comprehensive evaluation of the nature and spatial distribution of peptides and proteins, in addition to numerous other biological compounds including DNA segments, lipids, and other metabolites. This method can reveal a unique combination of 5–20 proteins, the so-called molecular signature, of a given entity [222]. One group has investigated the utility of this method as an ancillary test in the classification of histologically ambiguous spitzoid neoplasms [223–225] and in differentiating congenital nevi from melanoma in pediatric patients [226, 227]. The results are promising; however, there are technological barriers associated with this test that prevent its current wider implementation.

Assessment of Chromosomal Copy Number Aberrations with Comparative Genomic Hybridization (CGH) and Fluorescence In Situ Hybridization (FISH)

Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) allows the simultaneous detection of DNA gains and losses across the entire genome. Two versions of this method are currently used. In one, tumor DNA and reference (normal) DNA are labeled with two different fluorochromes (red and green), mixed and co-hybridized on a microarray. Each dot on the microarray contains DNA from a specific genomic locus, and therefore the resolution of the array is proportional with the number of dots. The intensity ratio of the fluorochromes labeling the tumor and normal DNA co-hybridized to each dot of the

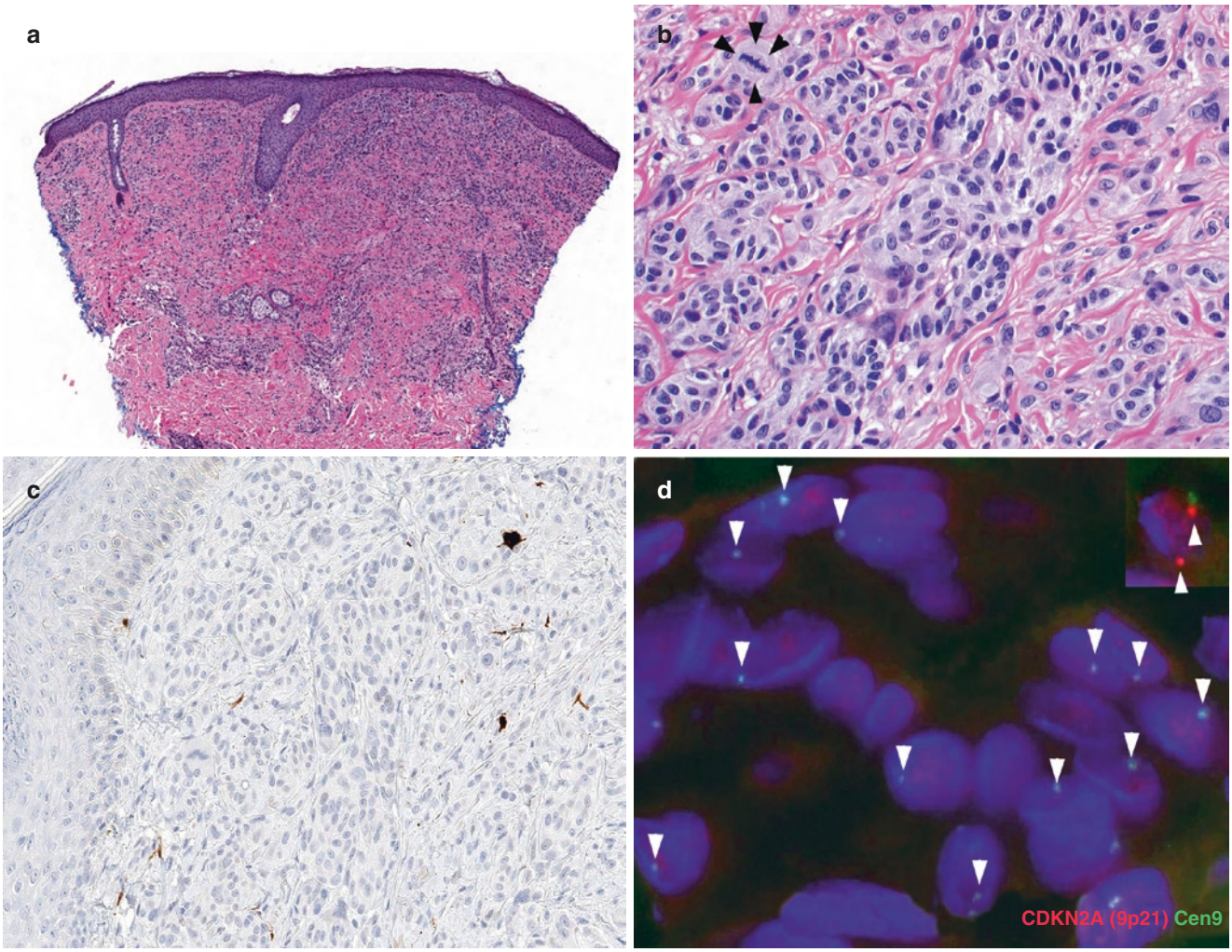


Fig. 34.6 Atypical Spitz tumor with homozygous loss of *CDKN2A*. (**a**) Dermal melanocytic neoplasm with an infiltrative pattern**.** (**b**) The lesion is composed of epithelioid melanocytes with large nuclei with prominent nucleoli and occasional hyperchromasia. Mitotic figures are easily noted (**arrowheads**). (**c**) Diffuse loss of $p16^{INK4a}$ expression by immunohistochemistry. (**d**) FISH shows complete absence of 9p21 signals in the nuclei of the lesional cells with one copy of centromere 9 (green-colored probe, arrowheads). Inset shows a nonneoplastic cell with normal complement of 9p21 (red-colored probe, **arrowheads**). In this case, the findings suggest an increased risk for adverse outcome

microarray is used to estimate the tumor DNA copy number at that locus. The other version of CGH uses only tumor DNA with no normal DNA. The intensity of the signal at each dot on the array is compared with a reference from normal tissue, which allows estimation of the DNA copy number at that locus. Single nucleotide polymorphism (SNP) microarrays are a variant of CGH arrays that interrogate specifically genomic loci containing SNPs. In addition to copy number data, these SNP platforms also provide allele frequency data, allow detection of copy neutral loss of heterozygosity (LOH) events, and can provide data on selected point mutations.

The notion that melanomas harbor multiple and somewhat consistent chromosomal copy number aberrations was already suggested decades ago by traditional cytogenetic analysis [228, 229]. Using genome-wide CGH, it was found initially that melanoma exhibits abundant chromosomal numerical abnormalities [230, 231]. Since these early publications, several studies have established that most melanomas have an unstable genome with numerous clonal DNA abnormalities including segmental gains of chromosomal regions 1q, 6p, 7, 8q, 11q13, 17q, and 20q; amplification of chromosomal regions 1q31, 4q12, 5p13, 5p15, 11q13, and 12q14; and reduced copy numbers of chromosomal regions 3q, 4q, 6q, 7, 8p, 8q, 9p, 10, 11p, 11q, 13, and 21q. In contrast, the vast majority of benign melanocytic nevi have no chromosomal aberrations or may show specific isolated abnormalities (such as 11p gains in desmoplastic Spitz nevi), which are not found in melanomas [42, 120, 232–238]. This nonoverlapping pattern of genomic aberrations provided an opportunity for diagnostic strategies based on tests evaluating DNA copy number alterations such as CGH or SNP microarrays. An initial study investigating the performance

of CGH in differentiating melanomas from nevi found that the assay would work well with a sensitivity of 96% and specificity of 98% [234]. A more recent study has documented that the number of chromosomal abnormalities correlates with disease progression [100]. Other investigators found that a high number of focal chromosomal abnormalities and the presence of chromothripsis in the primary melanoma correlate with aggressive behavior [239].

Recent advances in technology have improved the ability to effectively analyze degraded DNA from formalin-fixed paraffin-embedded (FFPE) tissue. One of these technologies is based on the use of molecular inversion probes (MIP) complementary to selected SNPs [237, 238]. This variant of SNP array requires only 80 ng of input DNA, and the probes have a footprint of only 40 bp which allows evaluation of highly degraded DNA. In addition, because the MIPs (instead of tumor DNA) are hybridized to the microarray, the signal-to-noise ratio is greatly improved over conventional CGH or SNP microarrays. A typical SNP microarray plot consists of two tracks (Figs. 34.7d, 34.8c, and 34.9d). The upper track depicts copy number changes (log ratio on the vertical and chromosome locus on the horizontal). Gains and losses are reflected by deflections of the average line above or below 0, respectively (Figs. 34.8c and 34.9d, arrows and arrowheads). The lower track shows allelic ratio for each SNP on the array (B-allele frequency on the vertical and chromosome locus on the horizontal). In the normal state, this track consists of three lines, a middle heterozygous line and two outer lines which are homozygous for the A and B alleles. A LOH event is represented by a split in the middle heterozygous line (Figs. 34.8c and 34.9d). This is usually associated with corresponding losses or gains at that locus; however, occasionally, an LOH without numerical abnormalities can be encountered representing a copy neutral LOH which is a form of acquired uniparental disomy.

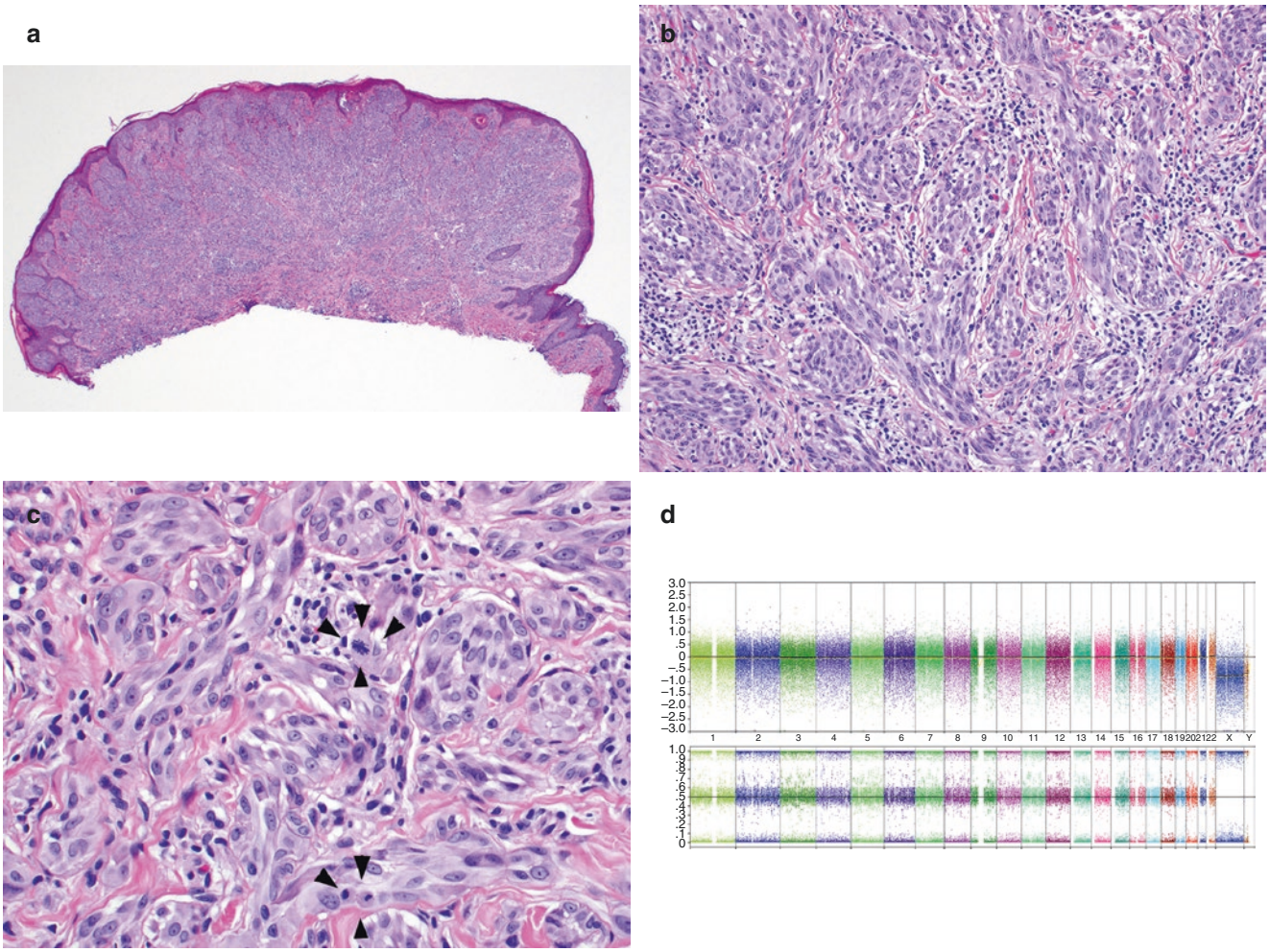


Fig. 34.7 Atypical Spitz tumor with features suggestive of ALK rearrangements. (**a**) Predominantly dermal lesion with exophytic profile. (**b**) At higher magnification it demonstrates a plexiform pattern with intersecting fascicles of fusiform melanocytes with large nuclei, prominent nucleoli, and abundant eosinophilic cytoplasm imparting a spitzoid morphology. An associated lymphocytic host response is present. (**c**) Mitotic figures are noted (**arrowheads**). (**d**) SNP microarray results show no significant numerical abnormalities or LOH events. In the context of this atypical Spitz tumor with borderline morphology, the results of the SNP array favor an indolent biologic behavior with a very low risk for progression similar with that of an atypical nevus

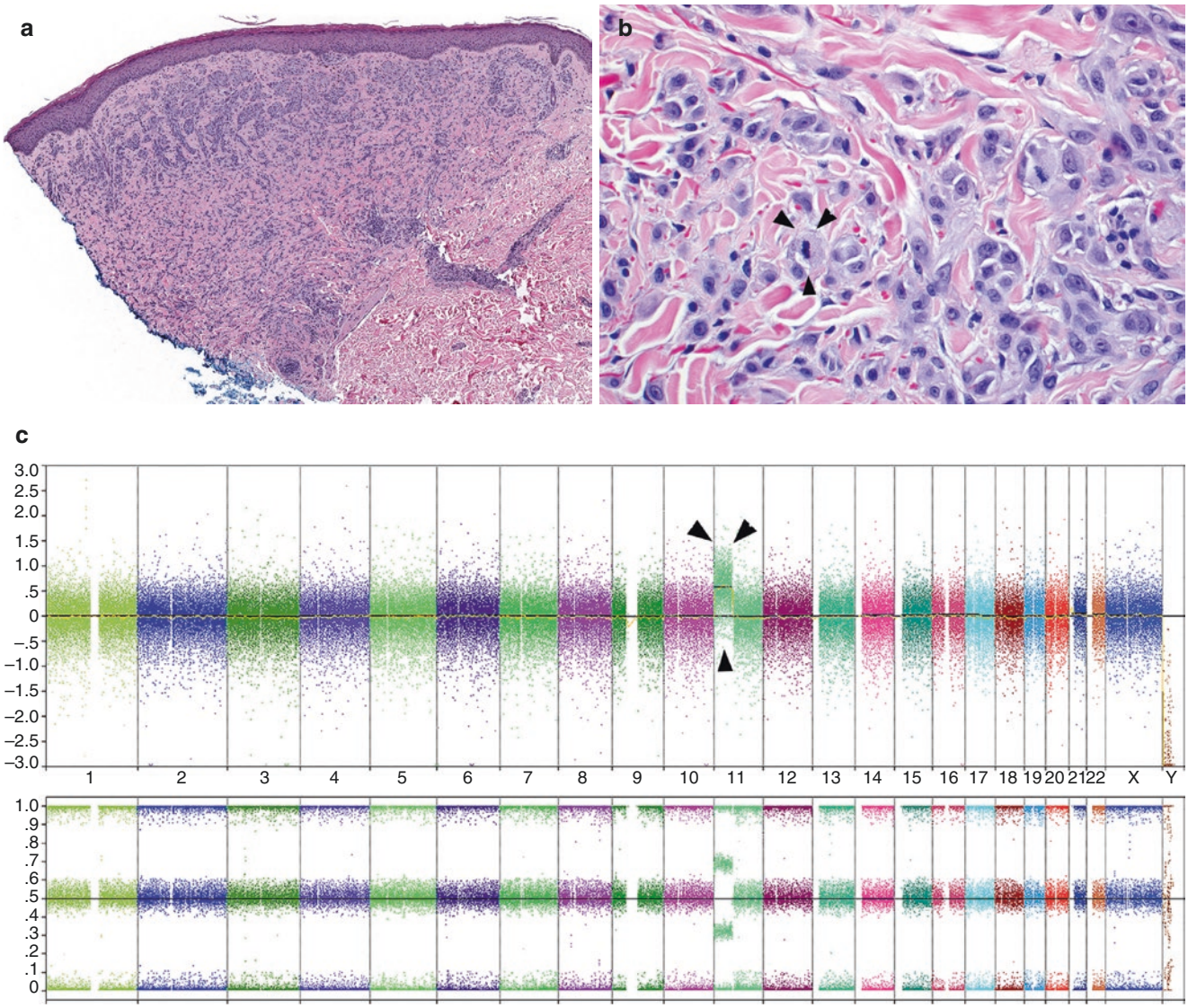


Fig. 34.8 Atypical Spitz tumor with gain of 11p. (**a**) Dermal melanocytic proliferation infiltrating through a desmoplastic stroma. (**b**) Atypical epithelioid and spindle melanocytes with spitzoid morphology are associated with thickened collagen fibers; a mitotic figure is noted in this field (**arrowheads**). (**c**) SNP microarray results show an isolated gain of 11p (**arrowheads**) with no other abnormalities. While the histology is worrisome and places this lesion in a borderline category, the results of the SNP array favor an indolent biologic behavior with a very low risk for progression and help establish a diagnosis of desmoplastic Spitz nevus

Several variants of CGH or SNP microarrays are being increasingly used in clinical setting as an ancillary test for histologically ambiguous melanocytic lesion. In this context, the absence of chromosomal abnormalities (or the presence of abnormalities known to be associated with nevi) favors an indolent lesion or a nevus (see composite Figs. 34.7 and 34.8). In contrast, the presence of multiple abnormalities favors a diagnosis of melanoma (composite Fig. 34.9).

Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) is another commonly used diagnostic test to aid in the classification of ambiguous melanocytic lesions that has evolved as an alternative to CGH/SNP microarrays. FISH evaluates interphase nuclei for copy number changes at selected loci. The selection of most optimal FISH probes was based on the analysis of data accumulated from CGH experiments performed on a wide variety of melanocytic lesions. An initial study identified three gene loci that are most often affected in malignant melanocytic tumors, either by losses (*MYB*) or gains (*RREB1* and *CCND1*) [240]. An initial panel included these three probes along with centromere 6 (to help quantify the two chromosome 6 loci). Due to the relative poor performance in the subsets of spitzoid and spindle melanomas, a second-generation probe set was developed that included probes for 9p21 (*CDKN2A*) and 8q24 (*MYC*) (composite Fig. 34.10) [57, 240, 241]. Currently, the following loci are included in FISH analysis of melanocytic lesions:

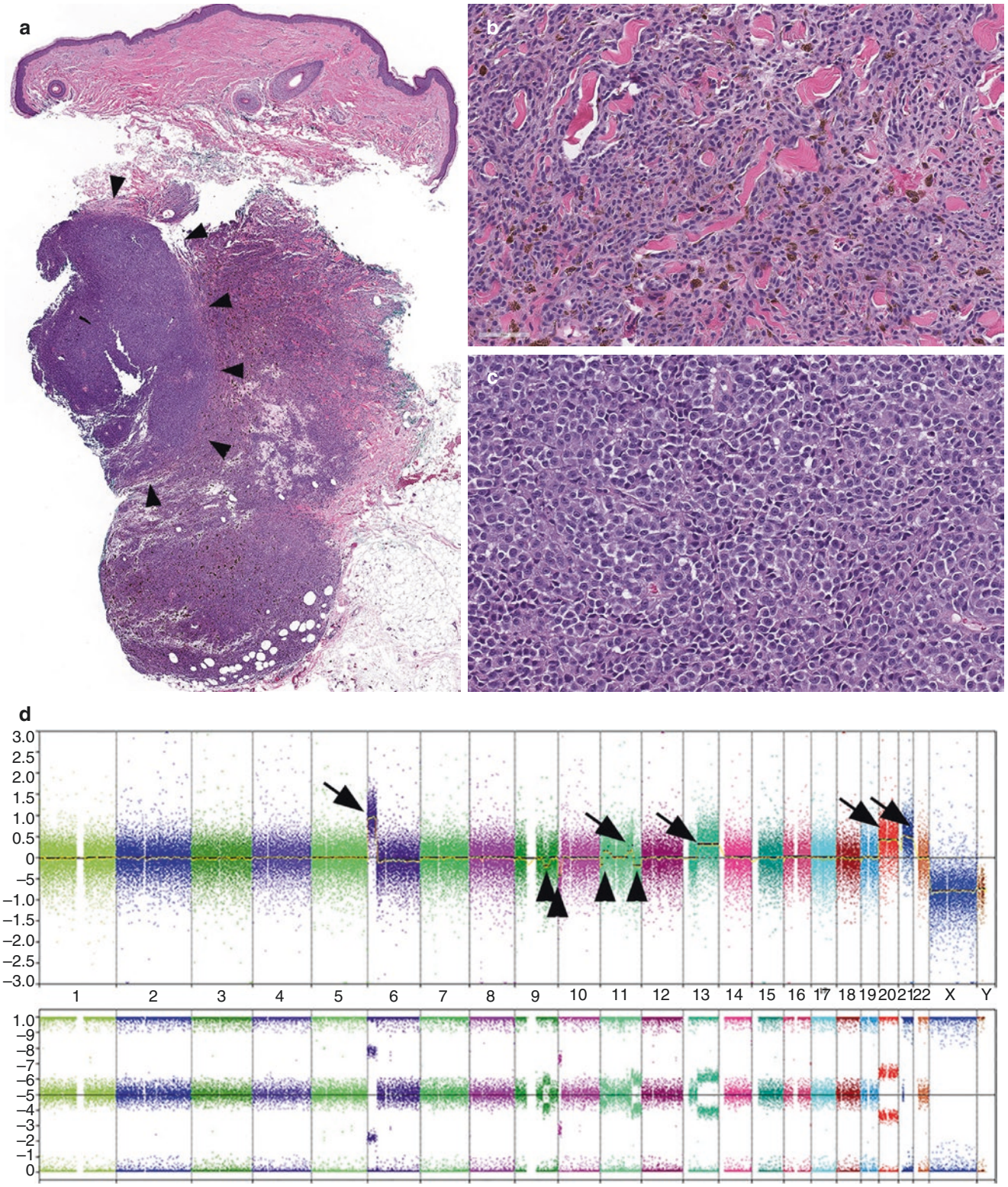


Fig. 34.9 Melanoma arising in a cellular blue nevus. (**a**). Intradermal cellular lesion bulging in the subcutis in a pattern suggesting a cellular blue nevus. An atypical expansile nodular proliferation is noted in the upper right side of this lesion (**arrowheads**). (**b**) Most of the lesion is composed of small spindle cells with ovoid uniform nuclei and pale cytoplasm with melanin pigment, arranged in connecting trabeculae and nests consistent with a cellular blue nevus. (**c**) The atypical nodule is composed of cytologically atypical epithelioid cells with conspicuous nucleoli concerning for malignant transformation. (**d**) SNP microarray results show multiple abnormalities including gains of 6p, 11p, 11q, 13q, 20, and 21 and losses of 9q, 10p, 11p, and 11q. In the context of this case, the findings favor a diagnosis of melanoma arising in a cellular blue nevus. (**e**) FISH with 6p25 (*RREB1*) probe (red) showing more than two signals in the majority of nuclei (80% of enumerated nuclei) consistent with a positive FISH result

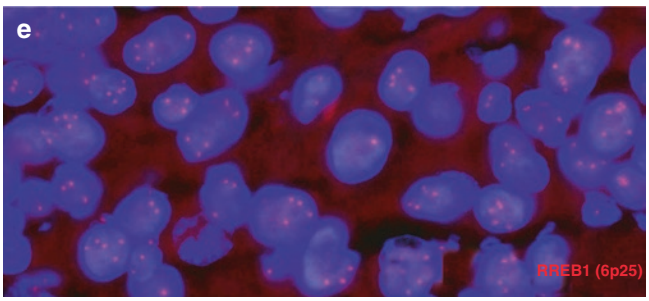


Fig. 34.9 (continued)

- *RREB1* (6p24.3, Ras-responsive element-binding protein 1 and Raf-responsive zinc-finger protein) codes for a transcription factor which increases Ras/Raf response via specific binding to the calcitonin gene promotor.
- *MYB* (6p23.3, avian myeloblastosis viral oncogene homolog) is an oncogene that encodes the MYB proto-oncogene protein, which has a critical role in hematopoiesis and tumorigenesis when aberrantly expressed.
- *MYC* (8q24.21, avian myelocytomatosis viral gene homolog) is a proto-oncogene and transcription factor, which activates multiple genes and encodes for proteins that regulate cell cycle and apoptosis.
- *CDKN2A* (9p21.3, cyclin-dependent kinase 2A) is a tumor suppressor gene, which encodes p16 (p16INK4a) and p14 (p14arf) proteins, regulators of the cell cycle. p16 inhibits cyclin-dependent kinase CDK4 and CDK6, dysregulating the cell cycle progression from G1 to S phase via retinoblastoma (*RB1*) gene products. Complementary gene, p14, is normally protective of p53 degradation and, when dysregulated, allows for inhibited apoptosis and tumorigenesis.
- *CCND1* (11q13.3, cyclin-dependent kinase 1) encodes a proto-oncogene, which alters cell-cycle regulation through CDK4/CDK6.

According to the published criteria, FISH test should be considered positive if *either one* of the following cutoffs is met:

- >55% of nuclei contain more signals for *RREB1* (6p25) than centromere 6 (CEP6) signals (Fig. 34.10a).
- >29% of nuclei contained more than two *RREB1* (6p25) signals (Fig. 34.9e).
- >40% of nuclei contain fewer *MYB* (6q23) than CEP6 signals (Fig. 34.10a).
- >29% of nuclei demonstrate homozygous loss of *CDKN2A* (9p21) (Figs. 34.6d and 34.10c).
- >38% of nuclei contain more than two signals for *CCND1* (11q13) (Fig. 34.10a).

All studies recommend a total of 30 cells from 3 different areas of the tumor to be enumerated. [240–250] While this approach is complex, it can be performed reliably (see, e.g., Figs. 34.5 and 34.7). This combination of FISH probes, particularly following the addition of probes for *CDKN2A* and *MYB*, has a reported sensitivity and specificity for unambiguous melanocytic lesions ranging from 83–90% to 90–95.4%, respectively [240, 244, 251, 252].

The use of FISH in the evaluation of spitzoid lesions has been helpful overall. Combining two large published series [252, 253] totaling 892 ambiguous lesions, almost half (47%) of the lesions showed Spitz-like features. Common FISH abnormalities seen in spitzoid melanomas include *MYB* (6q23) loss (as high as 72% of positive cases in one series [253]) and 11q gain (as opposed to 11p gain in desmoplastic Spitz nevus). Recent studies using either the expanded probe set, adding *CDKN2A* and *MYC* to the original three-probe set, or an alternate four-probe set containing *CDKN2A*, *CCND1*, *MYB*, and *MYC*, have improved the sensitivity from 70% to 85% and specificity to approaching 100% [241, 250, 253–255]. It appears that spitzoid lesions that show homozygous loss of *CDKN2A* or with one of the following aberrations: *TERT* promoter mutations, loss of *PTEN* as well as *NRAS*, and/or *BRAF* mutations demonstrate a more aggressive behavior (composite Fig. 34.6) [57, 176]. In contrast, lesions with isolated loss of *MYB* were found to have a less aggressive clinical course, and therefore isolated loss of *MYB* should not be used to upgrade an atypical Spitz tumor to spitzoid melanoma [57, 255, 256].

Morphologic Correlates of Copy Number Aberrations

FISH has been successfully employed to diagnose melanocytic entities including melanoma arising in a nevus (microstaging) [257], lentiginous melanocytic proliferations from chronically sun-damaged skin [258], acral lentiginous melanoma [259], blue nevus/deep penetrating nevus-like melanoma (composite Fig. 34.9) [243, 246], nevoid melanoma [242], and conjunctival melanocytic lesions [260], among others. CGH was found useful in evaluating pediatric melanocytic lesions and attempting to distinguish proliferative nodules from melanoma arising from giant congenital nevi. The proliferative nodules demonstrate either no numerical abnormalities or gains and/or losses of whole chromosomes, which in some cases may be accompanied by a rare partial chromosomal aberration, while melanomas arising in this context typically exhibit numerous segmental losses and gains [59, 261].

Several associations between histomorphological features and copy number changes have emerged. Lesions of chronic sun-damaged skin frequently show gain of *CCND1* [244, 251–253, 262], while nevoid melanoma, a morphologic mimic of banal nevi, shows significant gains of *MYC* [242, 249]. A subset of Spitz nevi, particularly the desmoplastic variant, demonstrate gains of 11p [55, 233], unlike conven-

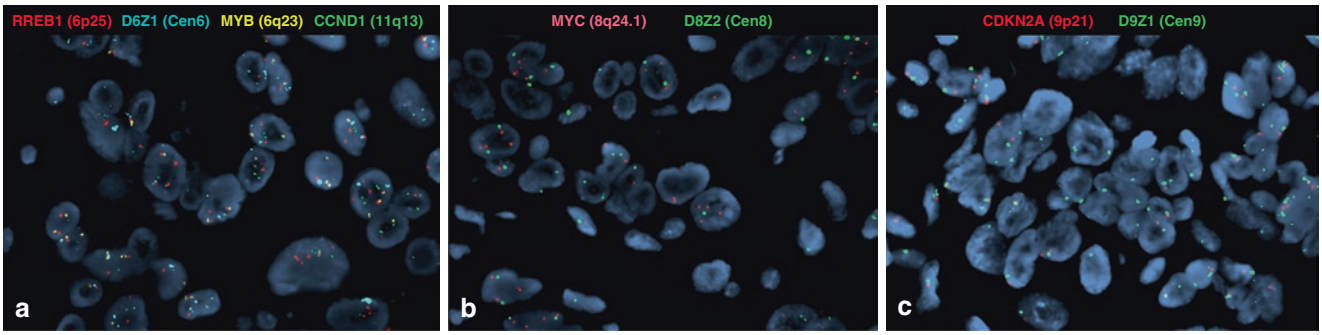


Fig. 34.10 The expanded eight-probe melanoma FISH panel includes probes at the following loci: *RREB1* (6p25, SpectrumRed™), *MYB* (6q23, SpectrumGold™), *CCND1* (11q13, SpectrumGreen™), *MYC* (8q24.1, SpectrumOrange™), *CDKN2A* (9p21, SpectrumOrange™), and centromeric probes at chromosome 6 (D6Z1, SpectrumAqua™), 8 (D8Z2, SpectrumGreen™), 9 (D9Z1, SpectrumGreen™). Inset **a** evaluates RREB1, MYB, CCND1, DZ61 which shows increased copies of RREB1 and CCND1. Inset **b** shows an increased MYC:D8Z2 ratio. Inset **c** shows heterozygous loss of CDKN2A, relative to D9Z1

tional nevi which do not harbor any copy number changes (composite Fig. 34.8). Isolated gain of 11p is a reassuring feature in atypical spitzoid proliferation and favors a nevus, as melanomas are not known to demonstrate this abnormality. Spitz nevi have a wide morphologic spectrum and considerable overlap with spitzoid melanoma, challenging even expert dermatopathologists [263]. As mentioned above in the proteomics and other sections, it has been shown that Spitz tumors harbor chromosomal fusions involving loci coding kinases such as NTRK1, NTRK3, ROS1, ALK, RET, and BRAF with correlation between the presence of certain given fusions and the morphology of the lesion, particularly in *ALK*- and *NTRK1*-rearranged tumors (composite Fig. 34.7) [56, 57]. The BAP1-inactivated melanocytic tumors (Fig. 34.5a, d) are characterized by losses involving chromosome 3p21 (or larger areas of chromosome 3) spanning the *BAP1* gene locus [12, 46]. In a similar fashion to the 11p gain, an isolated loss at 3p21 is reassuring and, in the absence of other abnormalities favors a BAP1-inactivated nevus over melanoma. Finally, although polyploidy can be seen in malignant melanocytic lesions, banal Spitz nevi have been shown to harbor tetraploidy [245, 253], which is characterized by four copies of each probe in nine or more cells [245, 253], and can represent a pitfall when interpreting FISH by increasing the likelihood of a false-positive result.

Pros and Cons of FISH and CGH

The advantage of CGH/SNP array methodologies over FISH is that they provide a complete genomic overview of copy number abnormalities and LOH. The disadvantages include tissue requirements (usually five to ten sections at ten microns), the need for 20–30% tumor purity in the sample, higher cost, and longer turnaround time. In addition, since this is an evolving field, lesions may harbor genomic aberrations of unknown significance. Furthermore, correlation with morphologic features (i.e., for microstaging) is not practical. Finally, the majority of the laboratories will have serious limitations when attempting to implement CGH/SNP arrays either related to infrastructure or reimbursement. Advantages to utilizing FISH over CGH include the ability to evaluate specimens with a small tumor burden and/or small specimens (theoretically only two to three slides at four microns are needed) and a shorter turnaround time. In addition, correlation with morphologic features is possible using FISH. Also for most laboratories, there are fewer barriers to implement FISH compared to CGH. However, FISH interpretation can pose a significant challenge, requiring specialized skills. When compared to CGH in recent small studies, FISH is noted to underdiagnose melanoma, commonly due to the current probe strategy, cutoff values, and errors in interpretation [120, 236–238]. There is also an increased risk for false-positive results, mainly due to tetraploidy (see above).

In summary, CGH/SNP array and FISH evaluation of melanocytic lesions are helpful diagnostic tools in histologically ambiguous lesions and can provide prognostic information. However, one must be aware of their limitations and always correlate the results with a thorough morphological evaluation. Multiple excellent reviews on the subject are available in the literature [57, 263, 264].

Gene Expression Analysis

In recent years, several groups have identified multiple genes purported to have diagnostic and prognostic relevance in melanoma [265–283]. These genes are suspected to be particularly informative for low-stage (I/II) melanomas that subsequently progress to metastasis. This is akin to the recently adopted change to the latest AJCC staging for breast cancers (8th Ed.) in which biomarker (ER/PR/HER2) status has played an importantly role in prognostication [284]. Current technol-

ogy allows focused interrogation of the transcriptome on FFPE and quantitative-reverse transcriptase-polymerase chain reaction (qRT-PCR). Other technologies used in clinical practice include gene expression microarray [274, 276, 278] and next-generation sequencing platforms (RNA-Seq) [277, 281]. The genes of interest are related to tissue development, epithelial differentiation, cell junction and adhesion, wound healing, immune response, and cell cycle progression [265–283].

In the USA, there are presently two clinically utilized commercial tests. The former is used predominantly for diagnostic purposes [274, 280, 282] and the latter for prognostic information [275, 276]. Studies using the diagnostic test, a 23-gene signature, have shown a receiver operating characteristic (ROC) curve with an area under the curve (AUC) for predicting melanoma versus nevi of 0.96, sensitivity of 90%, and specificity of 91% when compared to expert opinion [274, 282]. The prognostic assay is most frequently used in uveal melanomas [285]. Using this 31-gene signature that classifies lesions into high risk and low risk, this test has been shown to be independently predictive of outcome, namely, disease-free survival, distant metastasis-free survival, and overall survival [275, 276]. One study showed a positive predictive value of the test for distant metastasis to be 50% while the negative predictive value to be 82% [276]. Additionally, when used in combination with sentinel lymph node (SLN) biopsy status, improved prognostication was achieved [275].

Another group using chromatin immunoprecipitation-based quantitative PCR (ChIP-qPCR) for cell adhesion and cell-cycle genotyping showed that increased expression of *ITGB3*, *LAMB1*, *PLAT*, and *TP53* genes was associated with significant increase in SLN metastasis. This gene signature was characterized as the cell adhesion phenotype by quantitative PCR (CAP-QPCR) [277]. Another study, which compared the CAP-QPCR gene signature to FISH, revealed a significant correlation of prognostic pathologic markers (Breslow's depth, mitoses, SLN metastasis) with this gene signature [283].

It should be noted that these are still relatively small studies with limited follow-up and survival data, and further investigation of histologically ambiguous lesions is warranted.

Immunomodulatory Therapy and the Role of Genomic Analysis

The role of the immune system's antitumoral surveillance has long been recognized in melanoma [286]. Immunosuppressed patients have a higher incidence of melanoma and carry a worse prognosis [287]. On the opposing end of the spectrum, spontaneous regression of melanocytic nevi and melanoma is well described [288, 289]. There are varying numbers of CD8+ T-cell-infiltrating melanoma [290, 291], and tumors with high levels of CD8+ T cells tend to have longer overall survival than those with few CD8+ lymphocytes [286]. While T cells are the main mediators and effectors of the antitumoral immune response, the tumoral inflammatory infiltrate also contains immunosuppressive elements including tumor-associated macrophages (TAMs), regulatory T cells (Tregs), and myeloid-derived suppressor cells. Although these immunoregulatory cells normally act as quenchers of a cellular immune response, in the context of a malignancy, they are mainly summoned by factors secreted by the tumor, like TGF-β. Although this complex interaction challenges a simplistic interpretation of the immune infiltrate's antitumoral role, it opens a window of opportunity to harness the host's own immune system as a therapeutic weapon which is facilitated by the highly immunogenic nature of melanocytic lesions [291–293].

Interferon Alpha-2b and IL-2

Historically, interferon alpha-2b (IFN) therapy has been a mainstay for patients with clinically resected melanoma but high risk of disease relapse [294, 295]. More evidence exists for improved relapse-free survival than improved overall survival with IFN therapy [296, 297]. High-dose interleukin-2 (IL-2) was approved by the FDA for the treatment of metastatic melanoma in 1998, with an overall response rate of 16% in a meta-analysis of trials but with a severe toxicity profile significantly limiting its use [298]. Furthermore, improvements in survival are limited [299].

Anti-CTL4 and Anti-PD-L1 Blockage

More recently, therapies targeting immune checkpoint mechanisms between T cells and antigen-presenting cells (APCs) have come to the forefront of oncology (Fig. 34.11). The main interaction between the T cell and the APC is the binding of the T-cell receptor to the MHC-antigen complex, but numerous pathways can temper a successful interaction. Without a costimulatory signal mediated by protein B7 on the APC binding to CD28 on the T cell, the T cell enters a state of anergy. Activated T cells produce cytotoxic T-lymphocyte-associated protein-4 (CTLA-4, CD152), which competes with B7 and prevents the necessary costimulatory response. The monoclonal antibody ipilimumab blocks CTLA-4 and allows increased B7-CD28 costimulation. Ipilimumab therapy demonstrates significant improvement in overall survival in patients with metastatic melanoma [300–302] and in relapse-free survival in the adjuvant setting [303]. Since only a subset of patients show a dramatic response and toxicity can be significant, the identification of biomarkers predictive of CTLA-4 monoclonal antibody therapy response is an area of active research.

Another important negative regulator of activated lymphocytes is programmed death-1 (PD-1), which inhibits T-cell

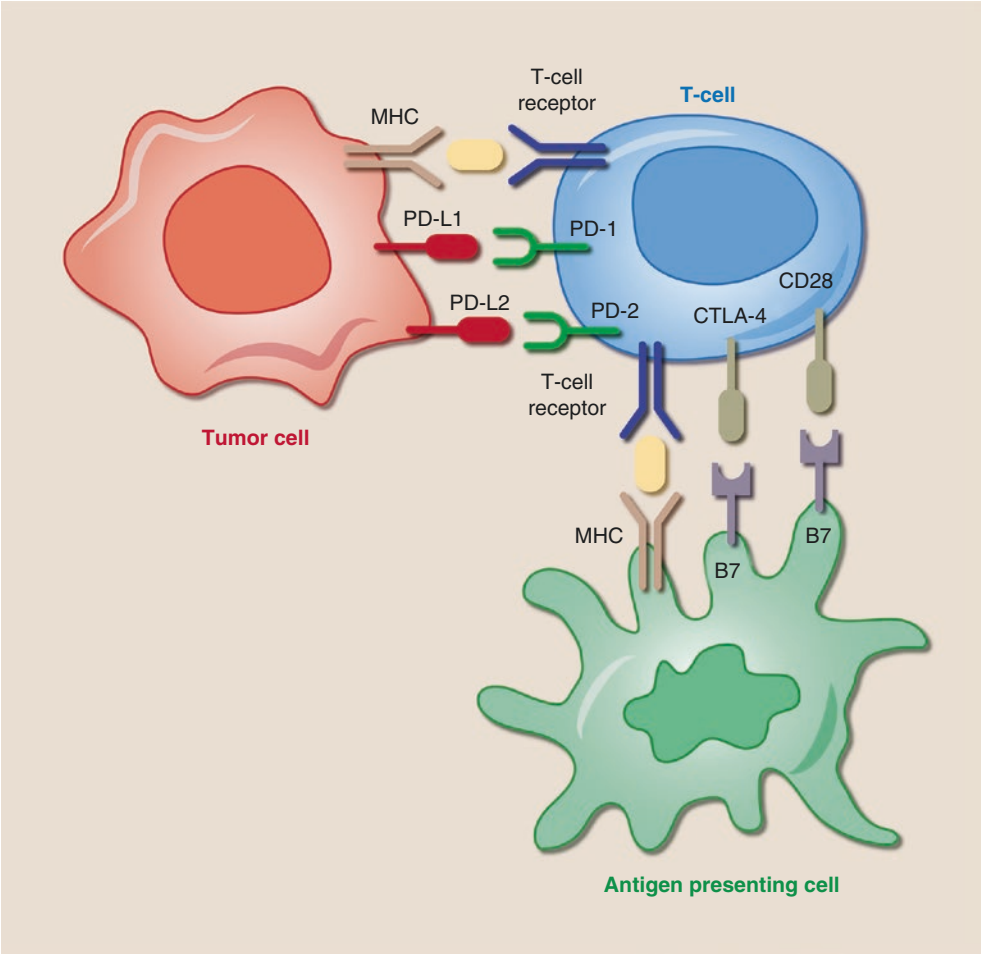


Fig. 34.11 Schematic overview of the tolerogenic mechanisms that limit antitumor cellular immunity. The programmed death-1/programmed death ligand 1 (PD-1/PD-L1) axis puts a break on the cytotoxic activities of lymphocytes that recognize melanoma. Similarly, the interaction between cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and B7 promotes immune tolerance via peripheral generation of T regulatory lymphocytes (Tregs). Breaking tumor immune tolerance through antagonists can induce a clinically effective host immune response to melanoma

function when bound to its ligands PD-L1 (expressed on many normal tissues) and PD-L2 (expressed in myeloid cells) (Fig. 34.11) [304, 305]. A wide variety of malignant cells express PD-L1, which aids in immune escape. Nivolumab and pembrolizumab are monoclonal antibodies against PD-1 and are approved for therapeutic use in patients with advanced melanoma either alone or in combination with other therapies such as ipilimumab or targeted molecular therapies [306–310]. Immunohistochemical staining for PD-L1 on tumor cells can be used as a complementary diagnostic test for PD-1 inhibitor therapy eligibility, but it is not required for use since cases negative for this marker can still respond, particularly to combined anti-PD-L1 + ipilimumab therapy [311].

Genomic Profiling as Predictors of Response with Immunomodulatory Therapy

Recent genomic profiling of pre-therapy metastatic melanoma responsive and unresponsive to anti-PD-1 therapy identified an increased mutational load in the responders versus the nonresponders, along with increased likelihood of BRCA2 mutation [312]. Higher burdens of unbalanced copy number loss also predict lower likelihood of response [313]. In addition, an expression signature seen in anti-PD-1 nonresponders significantly overlapped with an expression signature seen in tumors resistant to ERK inhibition [312, 314]. Signaling through the Wnt and PI3K/AKT pathways is also associated with immune cell exclusion and lower likelihood of response [130, 166]. Jak1/Jak2 mutations have also been found in melanoma with acquired resistance to PD-1 blockade [315, 316]. Although genomic profiling is not required for anti-PD-1 therapy, its widespread adoption for choosing targeted therapy makes it well suited to identify tumors in which immune checkpoint inhibition is more likely to be effective as this area evolves.

Moving Forward: Immunotherapy, Melanoma Neoantigens, and the Microbiota

Adoptive T-Cell Therapy, Antitumor Vaccines, BRAF Inhibition, and Melanoma Neoantigens

First developed in 1988, adoptive T-cell therapy (ACT) [317], using expanded autologous ex vivo tumor-infiltrating lymphocytes (TILs), achieved durable clinical responses, particularly when using gene-modified T cells with either tumor-associ-

ated antigens (TAA) or chimeric antigen receptors (CARs) [318]. This approach has been successfully applied in combination with a BRAF inhibitor to a limited number of patients [319] since it is known that BRAF inhibition favors an antitumoral lymphocytic response [320, 321]. On the other hand, antitumor vaccination for advanced melanoma originally developed in the late 1990s usually targeting TAAs (most commonly MAGE-1/MAGE-3, NY-ESO-1, Mart-1, gp100, and tyrosinase) has not been successful. [322] However, an interesting and novel approach is directing these immunotherapy strategies to tumoral neoantigens, which are not present in normal tissues and are known to occur in melanoma [323, 324]. Advances in recent technology in sequencing have allowed to investigate the presence of these neoantigens via whole-exome sequencing and RNA-Seq techniques to single out tumoral neoantigens to which ACT [325–327] and newly developed antitumor vaccines [322, 328, 329] could be directed to, improving and optimizing results.

Immunotherapy and the Microbiota/Microbiome

We are constantly accompanied by trillions of microorganisms including viruses, bacteria and fungi which coat several of our bodily surfaces and actually sum up to an "organ" with many functions that have been only recently began to be understood, including regulation of the immune system [330]. Thus, the relationship of the gut microbiota, or microbiome (combined genetic representation of the microbiota), and immunotherapy was the natural next step with initial studies performed in mice and with relationship to CTLA-4 blockage [331]. Further studies performed in humans in the context of PD-1 therapy for melanoma [332, 333] and epithelial malignancies [334] provided additional information about the immune modulation caused by these organisms along with the most beneficial microbial profiles that will result in better responses to this therapy. These new developments open exciting new avenues of intervention by modification of the patient's gut flora via fecal transplants to increase their responsiveness to immunotherapy.

Conclusions

The clinical utility of molecular/genomic testing in melanoma has grown substantially in recent years due to two primary factors: (1) our increased understanding of the genomic features of melanoma and (2) the clinical impact of such tests in directing patient management. The applicability of testing will likely continue to grow as targeted therapies are increasingly tested and employed in the adjuvant setting in earlier-stage disease. This chapter represents a brief snapshot of our current understanding of melanoma, and some of the approaches are currently used or may be deployed in the near future. This should provide a firm understanding of melanoma molecular testing and genomics that will inform the reader's understanding of this field as it continues to evolve over the coming months and years.

Acknowledgments The authors would like to acknowledge Katheryn Pearce[3] and the Division of Cytogenetics for their help with procuring FISH images.

References

1. Goldgeier MH, Klein LE, Klein-Angerer S, Moellmann G, Nordlund JJ. The distribution of melanocytes in the leptomeninges of the human brain. J Invest Dermatol. 1984;82(3):235–8.
2. Clemmensen OJ. Fenger C melanocytes in the anal canal epithelium. Histopathology. 1991;18(3):237–41.
3. Lin JY. Fisher DE melanocyte biology and skin pigmentation. Nature. 2007;445(7130):843–50.
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30.
5. Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross MI, Lazar AJ, Faries MB, et al. Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(6):472–92.
6. Zhang M, Qureshi AA, Geller AC, Frazier L, Hunter DJ, Han J. Use of tanning beds and incidence of skin cancer. J Clin Oncol. 2012;30(14):1588–93.
7. Harrington CR, Beswick TC, Leitenberger J, Minhajuddin A, Jacobe HT, Adinoff B. Addictive-like behaviours to ultraviolet light among frequent indoor tanners. Clin Exp Dermatol. 2011;36(1):33–8.
8. Soura E, Eliades PJ, Shannon K, Stratigos AJ, Tsao H. Hereditary melanoma: update on syndromes and management: genetics of familial atypical multiple mole melanoma syndrome. J Am Acad Dermatol. 2016;74(3):395–407. quiz 408-310.
9. Soufir N, Avril MF, Chompret A, Demenais F, Bombled J, Spatz A, Stoppa-Lyonnet D, Benard J, et al. Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group. Hum Mol Genet. 1998;7(2):209–16.
10. Molven A, Grimstvedt MB, Steine SJ, Harland M, Avril MF, Hayward NK, Akslen LA. A large Norwegian family with inherited malignant melanoma, multiple atypical nevi, and CDK4 mutation. Genes Chromosomes Cancer. 2005;44(1):10–8.
11. Testa JR, Cheung M, Pei J, Below JE, Tan Y, Sementino E, Cox NJ, Dogan AU, et al. Germline BAP1 mutations predispose to malignant mesothelioma. Nat Genet. 2011;43(10):1022–5.
12. Wiesner T, Obenauf AC, Murali R, Fried I, Griewank KG, Ulz P, Windpassinger C, Wackernagel W, et al. Germline mutations in BAP1 predispose to melanocytic tumors. Nat Genet. 2011;43(10):1018–21.
13. Carbone M, Yang H, Pass HI, Krausz T, Testa JR, Gaudino G. BAP1 and cancer. Nat Rev Cancer. 2013;13(3):153–9.
14. Curiel-Lewandrowski C, Speetzen LS, Cranmer L, Warneke JA, Loescher LJ. Multiple primary cutaneous melanomas in Li-Fraumeni syndrome. Arch Dermatol. 2011;147(2):248–50.
15. Giavedoni P, Ririe M, Carrera C, Puig S, Malvehy J. Familial melanoma associated with Li-Fraumeni syndrome and atypical

mole syndrome: total-body digital photography, dermoscopy and confocal microscopy. Acta Derm Venereol. 2017;97(6):720–3.
16. Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res. 2012;18(2):400–7.
17. Birck A, Ahrenkiel V, Zeuthen J, Hou-Jensen K, Guldberg P. Mutation and allelic loss of the PTEN/MMAC1 gene in primary and metastatic melanoma biopsies. J Invest Dermatol. 2000;114(2):277–80.
18. Eng C, Li FP, Abramson DH, Ellsworth RM, Wong FL, Goldman MB, Seddon J, Tarbell N, et al. Mortality from second tumors among long-term survivors of retinoblastoma. J Natl Cancer Inst. 1993;85(14):1121–8.
19. Bataille V, Hiles R, Bishop JA. Retinoblastoma, melanoma and the atypical mole syndrome. Br J Dermatol. 1995;132(1):134–8.
20. Kleinerman RA, Tucker MA, Tarone RE, Abramson DH, Seddon JM, Stovall M, Li FP, Fraumeni JF Jr. Risk of new cancers after radiotherapy in long-term survivors of retinoblastoma: an extended follow-up. J Clin Oncol. 2005;23(10):2272–9.
21. Bertolotto C, Lesueur F, Giuliano S, Strub T, de Lichy M, Bille K, Dessen P, d'Hayer B, et al. A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. Nature. 2011;480(7375):94–8.
22. Yokoyama S, Woods SL, Boyle GM, Aoude LG, MacGregor S, Zismann V, Gartside M, Cust AE, et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. Nature. 2011;480(7375):99–103.
23. Cocciolone RA, Crotty KA, Andrews L, Haass NK, Moloney FJ. Multiple desmoplastic melanomas in Birt-Hogg-Dube syndrome and a proposed signaling link between folliculin, the mTOR pathway, and melanoma susceptibility. Arch Dermatol. 2010;146(11):1316–8.
24. Fontcuberta IC, Salomao DR, Quiram PA, Pulido JS. Choroidal melanoma and lid fibrofolliculomas in Birt-Hogg-Dube syndrome. Ophthalmic Genet. 2011;32(3):143–6.
25. Mota-Burgos A, Acosta EH, Marquez FV, Mendiola M, Herrera-Ceballos E. Birt-Hogg-Dube syndrome in a patient with melanoma and a novel mutation in the FCLN gene. Int J Dermatol. 2013;52(3):323–6.
26. Kraemer KH, Lee MM, Andrews AD, Lambert WC. The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm. Arch Dermatol. 1994;130(8):1018–21.
27. Wang Y, Digiovanna JJ, Stern JB, Hornyak TJ, Raffeld M, Khan SG, Oh KS, Hollander MC, et al. Evidence of ultraviolet type mutations in xeroderma pigmentosum melanomas. Proc Natl Acad Sci U S A. 2009;106(15):6279–84.
28. Wang Y, Tan XH, DiGiovanna JJ, Lee CC, Stern JB, Raffeld M, Jaffe ES, Kraemer KH. Genetic diversity in melanoma metastases from a patient with xeroderma pigmentosum. J Invest Dermatol. 2010;130(4):1188–91.
29. Shibuya H, Kato A, Kai N, Fujiwara S, Goto M. A case of Werner syndrome with three primary lesions of malignant melanoma. J Dermatol. 2005;32(9):737–44.
30. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst. 1999;91(15):1310–6.
31. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, et al. POT1 loss-of-function variants predispose to familial melanoma. Nat Genet. 2014;46(5):478–81.
32. Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, Ghiorzo P, Bressac-de Paillerets B, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat Genet. 2014;46(5):482–6.
33. Landi MT, Bauer J, Pfeiffer RM, Elder DE, Hulley B, Minghetti P, Calista D, Kanetsky PA, et al. MC1R germline variants confer risk for BRAF-mutant melanoma. Science. 2006;313(5786):521–2.
34. Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, Randerson-Moor J, Aitken JF, et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat Genet. 2009;41(8):920–5.
35. LeBoit P. WHO classification of tumours of skin. Lyon: IARC Press; 2006.
36. Chung AF, Woodruff JM, Lewis JL Jr. Malignant melanoma of the vulva: a report of 44 cases. Obstet Gynecol. 1975;45(6):638–46.
37. Chattopadhyay C, Kim DW, Gombos DS, Oba J, Qin Y, Williams MD, Esmaeli B, Grimm EA, et al. Uveal melanoma: from diagnosis to treatment and the science in between. Cancer. 2016;122(15):2299–312.
38. Cummins DL, Cummins JM, Pantle H, Silverman MA, Leonard AL, Chanmugam A. Cutaneous malignant melanoma. Mayo Clin Proc. 2006;81(4):500–7.
39. Barnhill RL, Fine JA, Roush GC, Berwick M. Predicting five-year outcome for patients with cutaneous melanoma in a population-based study. Cancer. 1996;78(3):427–32.
40. Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, Patch AM, Kakavand H, et al. Whole-genome landscapes of major melanoma subtypes. Nature. 2017;545(7653):175–80.
41. Akbani RAK, Aksoy BA, Albert M, et al. Genomic classification of cutaneous melanoma. Cell. 2015;161(7):1681–96.
42. Curtin JA, Fridlyand J, Kageshita T, Patel HN, Busam KJ, Kutzner H, Cho KH, Aiba S, et al. Distinct sets of genetic alterations in melanoma. N Engl J Med. 2005;353(20):2135–47.
43. Bauer J, Curtin JA, Pinkel D, Bastian BC. Congenital melanocytic nevi frequently harbor NRAS mutations but no BRAF mutations. J Invest Dermatol. 2007;127(1):179–82.
44. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, Moses TY, Hostetter G, et al. High frequency of BRAF mutations in nevi. Nat Genet. 2003;33(1):19–20.
45. Yeh I, von Deimling A, Bastian BC. Clonal BRAF mutations in melanocytic nevi and initiating role of BRAF in melanocytic neoplasia. J Natl Cancer Inst. 2013;105(12):917–9.
46. Wiesner T, Murali R, Fried I, Cerroni L, Busam K, Kutzner H, Bastian BC. A distinct subset of atypical Spitz tumors is characterized by BRAF mutation and loss of BAP1 expression. Am J Surg Pathol. 2012;36(6):818–30.
47. Krauthammer M, Kong Y, Bacchiocchi A, Evans P, Pornputtapong N, Wu C, McCusker JP, Ma S, et al. Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. Nat Genet. 2015;47(9):996–1002.
48. Wiesner T, Kiuru M, Scott SN, Arcila M, Halpern AC, Hollmann T, Berger MF, Busam KJ NF1 Mutations Are Common in Desmoplastic Melanoma. Am J Surg Pathol. 2015;39(10):1357–62.
49. Jahn SW, Kashofer K, Halbwedl I, Winter G, El-Shabrawi-Caelen L, Mentzel T, Hoefler G, Liegl-Atzwanger B. Mutational dichotomy in desmoplastic malignant melanoma corroborated by multigene panel analysis. Mod Pathol. 2015;28(7):895–903.
50. Turner J, Couts K, Sheren J, Saichaemchan S, Ariyawutyakorn W, Avolio I, Cabral E, Glogowska M, et al. Kinase gene fusions in defined subsets of melanoma. Pigment Cell Melanoma Res. 2017;30(1):53–62.
51. Moon KR, Choi YD, Kim JM, Jin S, Shin MH, Shim HJ, Lee JB, Yun SJ. Genetic alterations in primary acral melanoma and acral melanocytic nevus in Korea: common mutated genes show distinct cytomorphological features. J Invest Dermatol. 2018;138(4):933–45.
52. Hintzsche JD, Gorden NT, Amato CM, Kim J, Wuensch KE, Robinson SE, Applegate AJ, Couts KL, et al. Whole-exome sequencing identifies recurrent SF3B1 R625 mutation and comutation of NF1 and KIT in mucosal melanoma. Melanoma Res. 2017;27(3):189–99.

53. Hou JY, Baptiste C, Hombalegowda RB, Tergas AI, Feldman R, Jones NL, Chatterjee-Paer S, Bus-Kwolfski A, et al. Vulvar and vaginal melanoma: a unique subclass of mucosal melanoma based on a comprehensive molecular analysis of 51 cases compared with 2253 cases of nongynecologic melanoma. Cancer. 2017;123(8):1333–44.
54. Tseng D, Kim J, Warrick A, Nelson D, Pukay M, Beadling C, Heinrich M, Selim MA, et al. Oncogenic mutations in melanomas and benign melanocytic nevi of the female genital tract. J Am Acad Dermatol. 2014;71(2):229–36.
55. Bastian BC, LeBoit PE, Pinkel D. Mutations and copy number increase of HRAS in Spitz nevi with distinctive histopathological features. Am J Pathol. 2000;157(3):967–72.
56. Wiesner T, He J, Yelensky R, Esteve-Puig R, Botton T, Yeh I, Lipson D, Otto G, et al. Kinase fusions are frequent in Spitz tumours and spitzoid melanomas. Nat Commun. 2014;5:3116.
57. Tetzlaff MT, Reuben A, Billings SD, Prieto VG, Curry JL. Toward a molecular-genetic classification of Spitzoid neoplasms. Clin Lab Med. 2017;37(3):431–48.
58. Yeh I, Lang UE, Durieux E, Tee MK, Jorapur A, Shain AH, Haddad V, Pissaloux D, et al. Combined activation of MAP kinase pathway and beta-catenin signaling cause deep penetrating nevi. Nat Commun. 2017;8(1):644.
59. Yelamos O, Arva NC, Obregon R, Yazdan P, Wagner A, Guitart J, Gerami P. A comparative study of proliferative nodules and lethal melanomas in congenital nevi from children. Am J Surg Pathol. 2015;39(3):405–15.
60. Moller I, Murali R, Muller H, Wiesner T, Jackett LA, Scholz SL, Cosgarea I, van de Nes JA, et al. Activating cysteinyl leukotriene receptor 2 (CYSLTR2) mutations in blue nevi. Mod Pathol. 2017;30(3):350–6.
61. Griewank KG, Muller H, Jackett LA, Emberger M, Moller I, van de Nes JA, Zimmer L, Livingstone E, et al. SF3B1 and BAP1 mutations in blue nevus-like melanoma. Mod Pathol. 2017;30(7):928–39.
62. Costa S, Byrne M, Pissaloux D, Haddad V, Paindavoine S, Thomas L, Aubin F, Lesimple T, et al. Melanomas associated with blue nevi or mimicking cellular blue nevi: clinical, pathologic, and molecular study of 11 cases displaying a high frequency of GNA11 mutations, BAP1 expression loss, and a predilection for the scalp. Am J Surg Pathol. 2016;40(3):368–77.
63. Van Raamsdonk CD, Bezrookove V, Green G, Bauer J, Gaugler L, O'Brien JM, Simpson EM, Barsh GS, et al. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. Nature. 2009;457(7229):599–602.
64. Van Raamsdonk CD, Griewank KG, Crosby MB, Garrido MC, Vemula S, Wiesner T, Obenauf AC, Wackernagel W, et al. Mutations in GNA11 in uveal melanoma. N Engl J Med. 2010;363(23):2191–9.
65. Murali R, Wiesner T, Rosenblum MK, Bastian BC. GNAQ and GNA11 mutations in melanocytomas of the central nervous system. Acta Neuropathol. 2012;123(3):457–9.
66. Robertson AG, Shih J, Yau C, Gibb EA, Oba J, Mungall KL, Hess JM, Uzunangelov V, et al. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. Cancer Cell. 2017;32(2):204–20. e215.
67. Cohen JN, Joseph NM, North JP, Onodera C, Zembowicz A, LeBoit PE. Genomic analysis of pigmented epithelioid Melanocytomas reveals recurrent alterations in PRKAR1A, and PRKCA genes. Am J Surg Pathol. 2017;41(10):1333–46.
68. Zembowicz A, Knoepp SM, Bei T, Stergiopoulos S, Eng C, Mihm MC, Stratakis CA. Loss of expression of protein kinase a regulatory subunit 1alpha in pigmented epithelioid melanocytoma but not in melanoma or other melanocytic lesions. Am J Surg Pathol. 2007;31(11):1764–75.
69. Antonescu CR, Nafa K, Segal NH, Dal Cin P, Ladanyi M. EWS-CREB1: a recurrent variant fusion in clear cell sarcoma--association with gastrointestinal location and absence of melanocytic differentiation. Clin Cancer Res. 2006;12(18):5356–62.
70. Hisaoka M, Ishida T, Kuo TT, Matsuyama A, Imamura T, Nishida K, Kuroda H, Inayama Y, et al. Clear cell sarcoma of soft tissue: a clinicopathologic, immunohistochemical, and molecular analysis of 33 cases. Am J Surg Pathol. 2008;32(3):452–60.
71. Wang WL, Mayordomo E, Zhang W, Hernandez VS, Tuvin D, Garcia L, Lev DC, Lazar AJ, et al. Detection and characterization of EWSR1/ATF1 and EWSR1/CREB1 chimeric transcripts in clear cell sarcoma (melanoma of soft parts). Mod Pathol. 2009;22(9):1201–9.
72. Lewis TB, Robison JE, Bastien R, Milash B, Boucher K, Samlowski WE, Leachman SA, Dirk Noyes R, et al. Molecular classification of melanoma using real-time quantitative reverse transcriptase-polymerase chain reaction. Cancer. 2005;104(8):1678–86.
73. Von Hoff DD, LoRusso PM, Rudin CM, Reddy JC, Yauch RL, Tibes R, Weiss GJ, Borad MJ, et al. Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. N Engl J Med. 2009;361(12):1164–72.
74. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, et al. A landscape of driver mutations in melanoma. Cell. 2012;150(2):251–63.
75. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nat Genet. 2012;44(9):1006–14.
76. Mahalingam M. NF1 and neurofibromin: emerging players in the genetic landscape of desmoplastic melanoma. Adv Anat Pathol. 2017;24(1):1–14.
77. van Engen-van Grunsven AC, van Dijk MC, Ruiter DJ, Klaasen A, Mooi WJ, Blokx WA. HRAS-mutated Spitz tumors: a subtype of Spitz tumors with distinct features. Am J Surg Pathol. 2010;34(10):1436–41.
78. Da Forno PD, Pringle JH, Fletcher A, Bamford M, Su L, Potter L, Saldanha G. BRAF, NRAS and HRAS mutations in spitzoid tumours and their possible pathogenetic significance. Br J Dermatol. 2009;161(2):364–72.
79. van Dijk MC, Bernsen MR, Ruiter DJ. Analysis of mutations in B-RAF, N-RAS, and H-RAS genes in the differential diagnosis of Spitz nevus and spitzoid melanoma. Am J Surg Pathol. 2005;29(9):1145–51.
80. Cirenajwis H, Lauss M, Ekedahl H, Torngren T, Kvist A, Saal LH, Olsson H, Staaf J, et al. NF1-mutated melanoma tumors harbor distinct clinical and biological characteristics. Mol Oncol. 2017;11(4):438–51.
81. Bastian BC. The molecular pathology of melanoma: an integrated taxonomy of melanocytic neoplasia. Annu Rev Pathol. 2014;9:239–71.
82. Viros A, Fridlyand J, Bauer J, Lasithiotakis K, Garbe C, Pinkel D, Bastian BC. Improving melanoma classification by integrating genetic and morphologic features. PLoS Med. 2008;5(6):e120.
83. Narita N, Tanemura A, Murali R, Scolyer RA, Huang S, Arigami T, Yanagita S, Chong KK, et al. Functional RET G691S polymorphism in cutaneous malignant melanoma. Oncogene. 2009;28(34):3058–68.
84. Ostrem JM, Shokat KM. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. Nat Rev Drug Discov. 2016;15(11):771–85.
85. Fedorenko IV, Gibney GT, Smalley KS. NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. Oncogene. 2013;32(25):3009–18.
86. Gajewski TF, Salama AK, Niedzwiecki D, Johnson J, Linette G, Bucher C, Blaskovich MA, Sebti SM, et al. Phase II study of the

farnesyltransferase inhibitor R115777 in advanced melanoma (CALGB 500104). J Transl Med. 2012;10:246.
87. Kwong LN, Costello JC, Liu H, Jiang S, Helms TL, Langsdorf AE, Jakubosky D, Genovese G, et al. Oncogenic NRAS signaling differentially regulates survival and proliferation in melanoma. Nat Med. 2012;18(10):1503–10.
88. Li J, Xu M, Yang Z, Li A, Dong J. Simultaneous inhibition of MEK and CDK4 leads to potent apoptosis in human melanoma cells. Cancer Investig. 2010;28(4):350–6.
89. Johnpulle RA, Johnson DB, Sosman JA. Molecular targeted therapy approaches for BRAF wild-type melanoma. Curr Oncol Rep. 2016;18(1):6.
90. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, et al. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949–54.
91. Dhomen N, Reis-Filho JS, da Rocha Dias S, Hayward R, Savage K, Delmas V, Larue L, Pritchard C, et al. Oncogenic Braf induces melanocyte senescence and melanoma in mice. Cancer Cell. 2009;15(4):294–303.
92. Kiel C, Benisty H, Llorens-Rico V, Serrano L. The yin-yang of kinase activation and unfolding explains the peculiarity of Val600 in the activation segment of BRAF. elife. 2016;5:e12814.
93. Holderfield M, Merritt H, Chan J, Wallroth M, Tandeske L, Zhai H, Tellew J, Hardy S, et al. RAF inhibitors activate the MAPK pathway by relieving inhibitory autophosphorylation. Cancer Cell. 2013;23(5):594–602.
94. Siroy AE, Boland GM, Milton DR, Roszik J, Frankian S, Malke J, Haydu L, Prieto VG, et al. Beyond BRAF(V600): clinical mutation panel testing by next-generation sequencing in advanced melanoma. J Invest Dermatol. 2015;135(2):508–15.
95. Hutchinson KE, Lipson D, Stephens PJ, Otto G, Lehmann BD, Lyle PL, Vnencak-Jones CL, Ross JS, et al. BRAF fusions define a distinct molecular subset of melanomas with potential sensitivity to MEK inhibition. Clin Cancer Res. 2013;19(24):6696–702.
96. Kim HS, Jung M, Kang HN, Kim H, Park CW, Kim SM, Shin SJ, Kim SH, et al. Oncogenic BRAF fusions in mucosal melanomas activate the MAPK pathway and are sensitive to MEK/PI3K inhibition or MEK/CDK4/6 inhibition. Oncogene. 2017;36(23):3334–45.
97. Shitara D, Tell-Marti G, Badenas C, Enokihara MM, Alos L, Larque AB, Michalany N, Puig-Butille JA, et al. Mutational status of naevus-associated melanomas. Br J Dermatol. 2015;173(3):671–80.
98. Poynter JN, Elder JT, Fullen DR, Nair RP, Soengas MS, Johnson TM, Redman B, Thomas NE, et al. BRAF and NRAS mutations in melanoma and melanocytic nevi. Melanoma Res. 2006;16(4):267–73.
99. Blokx WA, van Dijk MC, Ruiter DJ. Molecular cytogenetics of cutaneous melanocytic lesions – diagnostic, prognostic and therapeutic aspects. Histopathology. 2010;56(1):121–32.
100. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, Dummer R, North J, et al. The genetic evolution of melanoma from precursor lesions. N Engl J Med. 2015;373(20):1926–36.
101. Menzies AM, Haydu LE, Visintin L, Carlino MS, Howle JR, Thompson JF, Kefford RF, Scolyer RA, et al. Distinguishing clinicopathologic features of patients with V600E and V600K BRAF-mutant metastatic melanoma. Clin Cancer Res. 2012;18(12):3242–9.
102. Stadelmeyer E, Heitzer E, Resel M, Cerroni L, Wolf P, Dandachi N. The BRAF V600K mutation is more frequent than the BRAF V600E mutation in melanoma in situ of lentigo maligna type. J Invest Dermatol. 2014;134(2):548–50.
103. Ladstein RG, Bachmann IM, Straume O, Akslen LA. Tumor necrosis is a prognostic factor in thick cutaneous melanoma. Am J Surg Pathol. 2012;36(10):1477–82.
104. Sosman JA, Kim KB, Schuchter L, Gonzalez R, Pavlick AC, Weber JS, McArthur GA, Hutson TE, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N Engl J Med. 2012;366(8):707–14.
105. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, O'Dwyer PJ, Lee RJ, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med. 2010;363(9):809–19.
106. Falchook GS, Long GV, Kurzrock R, Kim KB, Arkenau TH, Brown MP, Hamid O, Infante JR, et al. Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. Lancet. 2012;379(9829):1893–901.
107. Williams TE, Subramanian S, Verhagen J, McBride CM, Costales A, Sung L, Antonios-McCrea W, McKenna M, et al. Discovery of RAF265: a potent mut-B-RAF inhibitor for the treatment of metastatic melanoma. ACS Med Chem Lett. 2015;6(9):961–5.
108. Su Y, Vilgelm AE, Kelley MC, Hawkins OE, Liu Y, Boyd KL, Kantrow S, Splittgerber RC, et al. RAF265 inhibits the growth of advanced human melanoma tumors. Clin Cancer Res. 2012;18(8):2184–98.
109. Arkenau HT, Kefford R, Long GV. Targeting BRAF for patients with melanoma. Br J Cancer. 2011;104(3):392–8.
110. Dickson MA, Gordon MS, Edelman G, Bendell JC, Kudchadkar RR, LoRusso PM, Johnston SH, Clary DO, et al. Phase I study of XL281 (BMS-908662), a potent oral RAF kinase inhibitor, in patients with advanced solid tumors. Investig New Drugs. 2015;33(2):349–56.
111. Das Thakur M, Salangsang F, Landman AS, Sellers WR, Pryer NK, Levesque MP, Dummer R, McMahon M, et al. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. Nature. 2013;494(7436):251–5.
112. Nikolaev SI, Rimoldi D, Iseli C, Valsesia A, Robyr D, Gehrig C, Harshman K, Guipponi M, et al. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. Nat Genet. 2011;44(2):133–9.
113. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, Ye Q, Lobo JM, et al. BRAF mutation predicts sensitivity to MEK inhibition. Nature. 2006;439(7074):358–62.
114. Hatzivassiliou G, Haling JR, Chen H, Song K, Price S, Heald R, Hewitt JF, Zak M, et al. Mechanism of MEK inhibition determines efficacy in mutant KRAS- versus BRAF-driven cancers. Nature. 2013;501(7466):232–6.
115. Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, Demidov LV, Hassel JC, et al. Improved survival with MEK inhibition in BRAF-mutated melanoma. N Engl J Med. 2012;367(2):107–14.
116. Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. N Engl J Med. 2015;372(1):30–9.
117. Larkin J, Ascierto PA, Dreno B, Atkinson V, Liszkay G, Maio M, Mandala M, Demidov L, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. N Engl J Med. 2014;371(20):1867–76.
118. Lu H, Liu S, Zhang G, Bin W, Zhu Y, Frederick DT, Hu Y, Zhong W, et al. PAK signalling drives acquired drug resistance to MAPK inhibitors in BRAF-mutant melanomas. Nature. 2017;550(7674):133–6.
119. Ambrosini G, Pratilas CA, Qin LX, Tadi M, Surriga O, Carvajal RD, Schwartz GK. Identification of unique MEK-dependent genes in GNAQ mutant uveal melanoma involved in cell growth, tumor cell invasion, and MEK resistance. Clin Cancer Res. 2012;18(13):3552–61.
120. Chan MP, Andea AA, Harms PW, Durham AB, Patel RM, Wang M, Robichaud P, Fisher GJ, et al. Genomic copy number analysis of a spectrum of blue nevi identifies recurrent aberrations of entire

chromosomal arms in melanoma ex blue nevus. Mod Pathol. 2016;29(3):227–39.
121. Carvajal RD, Sosman JA, Quevedo JF, Milhem MM, Joshua AM, Kudchadkar RR, Linette GP, Gajewski TF, et al. Effect of selumetinib vs chemotherapy on progression-free survival in uveal melanoma: a randomized clinical trial. JAMA. 2014;311(23):2397–405.
122. Russo AE, Torrisi E, Bevelacqua Y, Perrotta R, Libra M, McCubrey JA, Spandidos DA, Stivala F, et al. Melanoma: molecular pathogenesis and emerging target therapies (review). Int J Oncol. 2009;34(6):1481–9.
123. Robertson GP. Functional and therapeutic significance of Akt deregulation in malignant melanoma. Cancer Metastasis Rev. 2005;24(2):273–85.
124. Stahl JM, Sharma A, Cheung M, Zimmerman M, Cheng JQ, Bosenberg MW, Kester M, Sandirasegarane L, et al. Deregulated Akt3 activity promotes development of malignant melanoma. Cancer Res. 2004;64(19):7002–10.
125. Omholt K, Krockel D, Ringborg U, Hansson J. Mutations of PIK3CA are rare in cutaneous melanoma. Melanoma Res. 2006;16(2):197–200.
126. Shi H, Hugo W, Kong X, Hong A, Koya RC, Moriceau G, Chodon T, Guo R, et al. Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. Cancer Discov. 2014;4(1):80–93.
127. Chen G, Chakravarti N, Aardalen K, Lazar AJ, Tetzlaff MT, Wubbenhorst B, Kim SB, Kopetz S, et al. Molecular profiling of patient-matched brain and extracranial melanoma metastases implicates the PI3K pathway as a therapeutic target. Clin Cancer Res. 2014;20(21):5537–46.
128. Bucheit AD, Chen G, Siroy A, Tetzlaff M, Broaddus R, Milton D, Fox P, Bassett R, et al. Complete loss of PTEN protein expression correlates with shorter time to brain metastasis and survival in stage IIIB/C melanoma patients with BRAFV600 mutations. Clin Cancer Res. 2014;20(21):5527–36.
129. Davies MA, Stemke-Hale K, Lin E, Tellez C, Deng W, Gopal YN, Woodman SE, Calderone TC, et al. Integrated molecular and clinical analysis of AKT activation in metastatic melanoma. Clin Cancer Res. 2009;15(24):7538–46.
130. Peng W, Chen JQ, Liu C, Malu S, Creasy C, Tetzlaff MT, Xu C, McKenzie JA, et al. Loss of PTEN promotes resistance to T cell-mediated immunotherapy. Cancer Discov. 2016;6(2):202–16.
131. Trunzer K, Pavlick AC, Schuchter L, Gonzalez R, McArthur GA, Hutson TE, Moschos SJ, Flaherty KT, et al. Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma. J Clin Oncol. 2013;31(14):1767–74.
132. Nathanson KL, Martin AM, Wubbenhorst B, Greshock J, Letrero R, D'Andrea K, O'Day S, Infante JR, et al. Tumor genetic analyses of patients with metastatic melanoma treated with the BRAF inhibitor dabrafenib (GSK2118436). Clin Cancer Res. 2013;19(17):4868–78.
133. Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, Place CS, Taylor-Weiner A, et al. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. Cancer Discov. 2014;4(1):94–109.
134. Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, Chen Z, Lee MK, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature. 2010;468(7326):973–7.
135. Lin J, Sampath D, Nannini MA, Lee BB, Degtyarev M, Oeh J, Savage H, Guan Z, et al. Targeting activated Akt with GDC-0068, a novel selective Akt inhibitor that is efficacious in multiple tumor models. Clin Cancer Res. 2013;19(7):1760–72.
136. Vasudevan KM, Barbie DA, Davies MA, Rabinovsky R, McNear CJ, Kim JJ, Hennessy BT, Tseng H, et al. AKT-independent signaling downstream of oncogenic PIK3CA mutations in human cancer. Cancer Cell. 2009;16(1):21–32.
137. Margolin K, Longmate J, Baratta T, Synold T, Christensen S, Weber J, Gajewski T, Quirt I, et al. CCI-779 in metastatic melanoma: a phase II trial of the California Cancer Consortium. Cancer. 2005;104(5):1045–8.
138. Carracedo A, Ma L, Teruya-Feldstein J, Rojo F, Salmena L, Alimonti A, Egia A, Sasaki AT, et al. Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer. J Clin Invest. 2008;118(9):3065–74.
139. Chresta CM, Davies BR, Hickson I, Harding T, Cosulich S, Critchlow SE, Vincent JP, Ellston R, et al. AZD8055 is a potent, selective, and orally bioavailable ATP-competitive mammalian target of rapamycin kinase inhibitor with in vitro and in vivo antitumor activity. Cancer Res. 2010;70(1):288–98.
140. Deng W, Gopal YN, Scott A, Chen G, Woodman SE, Davies MA. Role and therapeutic potential of PI3K-mTOR signaling in de novo resistance to BRAF inhibition. Pigment Cell Melanoma Res. 2012;25(2):248–58.
141. Aziz SA, Jilaveanu LB, Zito C, Camp RL, Rimm DL, Conrad P, Kluger HM. Vertical targeting of the phosphatidylinositol-3 kinase pathway as a strategy for treating melanoma. Clin Cancer Res. 2010;16(24):6029–39.
142. Tolcher AW, Patnaik A, Papadopoulos KP, Rasco DW, Becerra CR, Allred AJ, Orford K, Aktan G, et al. Phase I study of the MEK inhibitor trametinib in combination with the AKT inhibitor afuresertib in patients with solid tumors and multiple myeloma. Cancer Chemother Pharmacol. 2015;75(1):183–9.
143. Rakosy Z, Vizkeleti L, Ecsedi S, Voko Z, Begany A, Barok M, Krekk Z, Gallai M, et al. EGFR gene copy number alterations in primary cutaneous malignant melanomas are associated with poor prognosis. Int J Cancer. 2007;121(8):1729–37.
144. Wu XC, Eide MJ, King J, Saraiya M, Huang Y, Wiggins C, Barnholtz-Sloan JS, Martin N, et al. Racial and ethnic variations in incidence and survival of cutaneous melanoma in the United States, 1999-2006. J Am Acad Dermatol. 2011;65(5 Suppl 1):S26–37.
145. Vizoso M, Ferreira HJ, Lopez-Serra P, Carmona FJ, Martinez-Cardus A, Girotti MR, Villanueva A, Guil S, et al. Epigenetic activation of a cryptic TBC1D16 transcript enhances melanoma progression by targeting EGFR. Nat Med. 2015;21(7):741–50.
146. Curtin JA, Busam K, Pinkel D, Bastian BC. Somatic activation of KIT in distinct subtypes of melanoma. J Clin Oncol. 2006;24(26):4340–6.
147. Handolias D, Salemi R, Murray W, Tan A, Liu W, Viros A, Dobrovic A, Kelly J, et al. Mutations in KIT occur at low frequency in melanomas arising from anatomical sites associated with chronic and intermittent sun exposure. Pigment Cell Melanoma Res. 2010;23(2):210–5.
148. Park E, Yang S, Emley A, DeCarlo K, Richards J, Mahalingam M. Lack of correlation between immunohistochemical expression of CKIT and KIT mutations in atypical acral nevi. Am J Dermatopathol. 2012;34(1):41–6.
149. Bastian BC, Kashani-Sabet M, Hamm H, Godfrey T, Moore DH 2nd, Brocker EB, LeBoit PE. Pinkel D gene amplifications characterize acral melanoma and permit the detection of occult tumor cells in the surrounding skin. Cancer Res. 2000;60(7):1968–73.
150. Ugurel S, Hildenbrand R, Zimpfer A, La Rosee P, Paschka P, Sucker A, Keikavoussi P, Becker JC, et al. Lack of clinical efficacy of imatinib in metastatic melanoma. Br J Cancer. 2005;92(8):1398–405.
151. Wyman K, Atkins MB, Prieto V, Eton O, McDermott DF, Hubbard F, Byrnes C, Sanders K, et al. Multicenter phase II trial of high-dose imatinib mesylate in metastatic melanoma: significant toxicity with no clinical efficacy. Cancer. 2006;106(9):2005–11.

152. Kim KB, Eton O, Davis DW, Frazier ML, McConkey DJ, Diwan AH, Papadopoulos NE, Bedikian AY, et al. Phase II trial of imatinib mesylate in patients with metastatic melanoma. Br J Cancer. 2008;99(5):734–40.
153. Kluger HM, Dudek AZ, McCann C, Ritacco J, Southard N, Jilaveanu LB, Molinaro A, Sznol M. A phase 2 trial of dasatinib in advanced melanoma. Cancer. 2011;117(10):2202–8.
154. Hofmann UB, Kauczok-Vetter CS, Houben R, Becker JC. Overexpression of the KIT/SCF in uveal melanoma does not translate into clinical efficacy of imatinib mesylate. Clin Cancer Res. 2009;15(1):324–9.
155. Carvajal RD, Antonescu CR, Wolchok JD, Chapman PB, Roman RA, Teitcher J, Panageas KS, Busam KJ, et al. KIT as a therapeutic target in metastatic melanoma. JAMA. 2011;305(22):2327–34.
156. Guo J, Si L, Kong Y, Flaherty KT, Xu X, Zhu Y, Corless CL, Li L, et al. Phase II, open-label, single-arm trial of imatinib mesylate in patients with metastatic melanoma harboring c-Kit mutation or amplification. J Clin Oncol. 2011;29(21):2904–9.
157. Lee SJ, Kim TM, Kim YJ, Jang KT, Lee HJ, Lee SN, Ahn MS, Hwang IG, et al. Phase II trial of nilotinib in patients with metastatic malignant melanoma harboring KIT gene aberration: a multicenter trial of Korean Cancer Study Group (UN10-06). Oncologist. 2015;20(11):1312–9.
158. Hodi FS, Corless CL, Giobbie-Hurder A, Fletcher JA, Zhu M, Marino-Enriquez A, Friedlander P, Gonzalez R, et al. Imatinib for melanomas harboring mutationally activated or amplified KIT arising on mucosal, acral, and chronically sun-damaged skin. J Clin Oncol. 2013;31(26):3182–90.
159. Carvajal RD, Lawrence DP, Weber JS, Gajewski TF, Gonzalez R, Lutzky J, O'Day SJ, Hamid O, et al. Phase II study of nilotinib in melanoma harboring KIT alterations following progression to prior KIT inhibition. Clin Cancer Res. 2015;21(10):2289–96.
160. Lee CK, Goldstein D, Gibbs E, Joensuu H, Zalcberg J, Verweij J, Casali PG, Maki RG, et al. Development and validation of prognostic nomograms for metastatic gastrointestinal stromal tumour treated with imatinib. Eur J Cancer. 2015;51(7):852–60.
161. O'Connell MP, Weeraratna AT. Hear the Wnt Ror: how melanoma cells adjust to changes in Wnt. Pigment Cell Melanoma Res. 2009;22(6):724–39.
162. Vibert L, Aquino G, Gehring I, Subkankulova T, Schilling TF, Rocco A, Kelsh RN. An ongoing role for Wnt signaling in differentiating melanocytes in vivo. Pigment Cell Melanoma Res. 2017;30(2):219–32.
163. Xue G, Romano E, Massi D, Mandala M. Wnt/beta-catenin signaling in melanoma: preclinical rationale and novel therapeutic insights. Cancer Treat Rev. 2016;49:1–12.
164. Anastas JN, Kulikauskas RM, Tamir T, Rizos H, Long GV, von Euw EM, Yang PT, Chen HW, et al. WNT5A enhances resistance of melanoma cells to targeted BRAF inhibitors. J Clin Invest. 2014;124(7):2877–90.
165. Chien AJ, Haydu LE, Biechele TL, Kulikauskas RM, Rizos H, Kefford RF, Scolyer RA, Moon RT, et al. Targeted BRAF inhibition impacts survival in melanoma patients with high levels of Wnt/beta-catenin signaling. PLoS One. 2014;9(4):e94748.
166. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. Nature. 2015;523(7559):231–5.
167. Spranger S, Gajewski TF. Impact of oncogenic pathways on evasion of antitumour immune responses. Nat Rev Cancer. 2018;18(3):139–47.
168. Bell RJ, Rube HT, Xavier-Magalhaes A, Costa BM, Mancini A, Song JS, Costello JF. Understanding TERT promoter mutations: a common path to immortality. Mol Cancer Res. 2016;14(4):315–23.
169. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, et al. Specific association of human telomerase activity with immortal cells and cancer. Science. 1994;266(5193):2011–5.
170. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013;339(6122):957–9.
171. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013;339(6122):959–61.
172. Chiba K, Lorbeer FK, Shain AH, McSwiggen DT, Schruf E, Oh A, Ryu J, Darzacq X, et al. Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. Science. 2017;357(6358):1416–20.
173. Heidenreich B, Nagore E, Rachakonda PS, Garcia-Casado Z, Requena C, Traves V, Becker J, Soufir N, et al. Telomerase reverse transcriptase promoter mutations in primary cutaneous melanoma. Nat Commun. 2014;5:3401.
174. Liang WS, Hendricks W, Kiefer J, Schmidt J, Sekar S, Carpten J, Craig DW, Adkins J, et al. Integrated genomic analyses reveal frequent TERT aberrations in acral melanoma. Genome Res. 2017;27(4):524–32.
175. Griewank KG, Murali R, Puig-Butille JA, Schilling B, Livingstone E, Potrony M, Carrera C, Schimming T et al. TERT promoter mutation status as an independent prognostic factor in cutaneous melanoma. J Natl Cancer Inst. 2014;106(9).
176. Lee S, Barnhill RL, Dummer R, Dalton J, Wu J, Pappo A, Bahrami A. TERT promoter mutations are predictive of aggressive clinical behavior in patients with spitzoid melanocytic neoplasms. Sci Rep. 2015;5:11200.
177. Kang HJ, Cui Y, Yin H, Scheid A, Hendricks WP, Schmidt J, Sekulic A, Kong D et al. A pharmacological chaperone molecule induces Cancer cell death by restoring tertiary DNA structures in mutant hTERT promoters. J. Am. Chem. Soc. 2016;138(41):13673–92.
178. Moran B, Silva R, Perry AS, Gallagher WM Epigenetics of malignant melanoma. Semin Cancer Biol. 2018;51:80–88.
179. Cannuyer J, Van Tongelen A, Loriot A, De Smet C. A gene expression signature identifying transient DNMT1 depletion as a causal factor of cancer-germline gene activation in melanoma. Clin Epigenetics. 2015;7:114.
180. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 2012;150(6):1135–46.
181. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12–27.
182. Woods DM, Sodre AL, Villagra A, Sarnaik A, Sotomayor EM, Weber J. HDAC inhibition upregulates PD-1 ligands in melanoma and augments immunotherapy with PD-1 blockade. Cancer Immunol Res. 2015;3(12):1375–85.
183. Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, et al. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. Nat Commun. 2015;6:6683.
184. Sengupta D, Byrum SD, Avaritt NL, Davis L, Shields B, Mahmoud F, Reynolds M, Orr LM, et al. Quantitative histone mass spectrometry identifies elevated histone H3 lysine 27 (Lys27) trimethylation in melanoma. Mol Cell Proteomics. 2016;15(3):765–75.
185. Souroullas GP, Jeck WR, Parker JS, Simon JM, Liu JY, Paulk J, Xiong J, Clark KS, et al. An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. Nat Med. 2016;22(6):632–40.
186. Shields BD, Mahmoud F, Taylor EM, Byrum SD, Sengupta D, Koss B, Baldini G, Ransom S, et al. Indicators of responsiveness to immune checkpoint inhibitors. Sci Rep. 2017;7(1):807.
187. Tiffen J, Gallagher SJ, Hersey P. EZH2: an emerging role in melanoma biology and strategies for targeted therapy. Pigment Cell Melanoma Res. 2015;28(1):21–30.

188. Lee W, Teckie S, Wiesner T, Ran L, Prieto Granada CN, Lin M, Zhu S, Cao Z, et al. PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nat Genet. 2014;46(11):1227–32.
189. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, Helin K, Hornick JL, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. Nature. 2014;514(7521):247–51.
190. Goding CR. Targeting the lncRNA SAMMSON reveals metabolic vulnerability in melanoma. Cancer Cell. 2016;29(5):619–21.
191. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eyckerman S, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature. 2016;531(7595):518–22.
192. Hanniford D, Zhong J, Koetz L, Gaziel-Sovran A, Lackaye DJ, Shang S, Pavlick A, Shapiro R, et al. A miRNA-based signature detected in primary melanoma tissue predicts development of brain metastasis. Clin Cancer Res. 2015;21(21):4903–12.
193. Tembe V, Schramm SJ, Stark MS, Patrick E, Jayaswal V, Tang YH, Barbour A, Hayward NK, et al. MicroRNA and mRNA expression profiling in metastatic melanoma reveal associations with BRAF mutation and patient prognosis. Pigment Cell Melanoma Res. 2015;28(3):254–66.
194. Pfeffer SR, Grossmann KF, Cassidy PB, Yang CH, Fan M, Kopelovich L, Leachman SA, Pfeffer LM. Detection of exosomal miRNAs in the plasma of melanoma patients. J Clin Med. 2015;4(12):2012–27.
195. Alegre E, Sanmamed MF, Rodriguez C, Carranza O, Martin-Algarra S, Gonzalez A. Study of circulating microRNA-125b levels in serum exosomes in advanced melanoma. Arch Pathol Lab Med. 2014;138(6):828–32.
196. Busam KJ, Hedvat C, Pulitzer M, von Deimling A, Jungbluth AA. Immunohistochemical analysis of BRAF(V600E) expression of primary and metastatic melanoma and comparison with mutation status and melanocyte differentiation antigens of metastatic lesions. Am J Surg Pathol. 2013;37(3):413–20.
197. Long GV, Wilmott JS, Capper D, Preusser M, Zhang YE, Thompson JF, Kefford RF, von Deimling A, et al. Immunohistochemistry is highly sensitive and specific for the detection of V600E BRAF mutation in melanoma. Am J Surg Pathol. 2013;37(1):61–5.
198. Tetzlaff MT, Pattanaprichakul P, Wargo J, Fox PS, Patel KP, Estrella JS, Broaddus RR, Williams MD, et al. Utility of BRAF V600E immunohistochemistry expression pattern as a surrogate of BRAF mutation status in 154 patients with advanced melanoma. Hum Pathol. 2015;46(8):1101–10.
199. Rapisuwon S, Busam KJ, Parks K, Chapman PB, Lee E, Atkins MB. Discordance between Cobas BRAF V600 testing and VE1 immunohistochemistry in a melanoma patient with bone marrow metastases. Am J Dermatopathol. 2016;38(9):687–9.
200. Ponti G, Tomasi A, Maiorana A, Ruini C, Maccaferri M, Cesinaro AM, Depenni R, Manni P, et al. BRAFp.V600E, p.V600K, and p.V600R mutations in malignant melanoma: do they also differ in Immunohistochemical assessment and clinical features? Appl Immunohistochem Mol Morphol. 2016;24(1):30–4.
201. Heinzerling L, Kuhnapfel S, Meckbach D, Baiter M, Kaempgen E, Keikavoussi P, Schuler G, Agaimy A, et al. Rare BRAF mutations in melanoma patients: implications for molecular testing in clinical practice. Br J Cancer. 2013;108(10):2164–71.
202. Kakavand H, Walker E, Lum T, Wilmott JS, Selinger CI, Smith E, Saw RP, Yu B, et al. BRAF(V600E) and NRAS(Q61L/Q61R) mutation analysis in metastatic melanoma using immunohistochemistry: a study of 754 cases highlighting potential pitfalls and guidelines for interpretation and reporting. Histopathology. 2016;69(4):680–6.
203. Massi D, Simi L, Sensi E, Baroni G, Xue G, Scatena C, Caldarella A, Pinzani P, et al. Immunohistochemistry is highly sensitive and specific for the detection of NRASQ61R mutation in melanoma. Mod Pathol. 2015;28(4):487–97.
204. Harms PW, Hocker TL, Zhao L, Chan MP, Andea AA, Wang M, Harms KL, Wang ML, et al. Loss of p16 expression and copy number changes of CDKN2A in a spectrum of spitzoid melanocytic lesions. Hum Pathol. 2016;58:152–60.
205. Al Dhaybi R, Agoumi M, Gagne I, McCuaig C, Powell J, Kokta V. p16 expression: a marker of differentiation between childhood malignant melanomas and spitz nevi. J Am Acad Dermatol. 2011;65(2):357–63.
206. Hilliard NJ, Krahl D, Sellheyer K. p16 expression differentiates between desmoplastic spitz nevus and desmoplastic melanoma. J Cutan Pathol. 2009;36(7):753–9.
207. Lade-Keller J, Riber-Hansen R, Guldberg P, Schmidt H, Hamilton-Dutoit SJ, Steiniche T. Immunohistochemical analysis of molecular drivers in melanoma identifies p16 as an independent prognostic biomarker. J Clin Pathol. 2014;67(6):520–8.
208. Rowe CJ, Tang F, Hughes MC, Rodero MP, Malt M, Lambie D, Barbour A, Hayward NK, et al. Molecular markers to complement sentinel node status in predicting survival in patients with high-risk locally invasive melanoma. Int J Cancer. 2016;139(3):664–72.
209. Uguen A, Uguen M, Guibourg B, Talagas M, Marcorelles P, De Braekeleer M The p16-Ki-67-HMB45 Immunohistochemistry Scoring System is Highly Concordant With the Fluorescent In Situ Hybridization Test to Differentiate Between Melanocytic Nevi and Melanomas. Appl Immunohistochem Mol Morphol. 2018;26 (6):361–7.
210. Strickler AG, Schaefer JT, Slingluff CL Jr, Wick MR. Immunolabeling for p16, WT1, and Fli-1 in the assignment of growth phase for cutaneous melanomas. Am J Dermatopathol. 2014;36(9):718–22.
211. de la Fouchardiere A, Cabaret O, Savin L, Combemale P, Schvartz H, Penet C, Bonadona V, Soufir N et al. Germline BAP1 mutations predispose also to multiple basal cell carcinomas. Clin Genet. 2015; 88 (3):273–77.
212. Murali R, Wiesner T, Scolyer RA. Tumours associated with BAP1 mutations. Pathology. 2013;45(2):116–26.
213. Murali R, Wilmott JS, Jakrot V, Al-Ahmadie HA, Wiesner T, McCarthy SW, Thompson JF, Scolyer RA. BAP1 expression in cutaneous melanoma: a pilot study. Pathology. 2013;45(6):606–9.
214. Massi D, Romano E, Rulli E, Merelli B, Nassini R, De Logu F, Bieche I, Baroni G, et al. Baseline beta-catenin, programmed death-ligand 1 expression and tumour-infiltrating lymphocytes predict response and poor prognosis in BRAF inhibitor-treated melanoma patients. Eur J Cancer. 2017;78:70–81.
215. Larson AR, Dresser KA, Zhan Q, Lezcano C, Woda BA, Yosufi B, Thompson JF, Scolyer RA, et al. Loss of 5-hydroxymethylcytosine correlates with increasing morphologic dysplasia in melanocytic tumors. Mod Pathol. 2014;27(7):936–44.
216. Lee JJ, Cook M, Mihm MC, Xu S, Zhan Q, Wang TJ, Murphy GF, Lian CG. Loss of the epigenetic mark, 5-Hydroxymethylcytosine, correlates with small cell/nevoid subpopulations and assists in microstaging of human melanoma. Oncotarget. 2015;6(35):37995–8004.
217. Lee JJ, Granter SR, Laga AC, Saavedra AP, Zhan Q, Guo W, Xu S, Murphy GF, et al. 5-Hydroxymethylcytosine expression in metastatic melanoma versus nodal nevus in sentinel lymph node biopsies. Mod Pathol. 2015;28(2):218–29.
218. Pavlova O, Fraitag S, Hohl D 5-Hydroxymethylcytosine Expression in Proliferative Nodules Arising within Congenital Nevi Allows Differentiation from Malignant Melanoma. J Invest Dermatol. 2016;136(12):2453–61.
219. Busam KJ, Shah KN, Gerami P, Sitzman T, Jungbluth AA, Kinsler V. Reduced H3K27me3 expression is common in nodular melanomas of childhood associated with congenital melanocytic nevi but not in proliferative nodules. Am J Surg Pathol. 2017;41(3):396–404.
220. Kampilafkos P, Melachrinou M, Kefalopoulou Z, Lakoumentas J, Sotiropoulou-Bonikou G. Epigenetic modifications in cutaneous

malignant melanoma: EZH2, H3K4me2, and H3K27me3 immunohistochemical expression is enhanced at the invasion front of the tumor. Am J Dermatopathol. 2015;37(2):138–44.
221. Le Guellec S, Macagno N, Velasco V, Lamant L, Lae M, Filleron T, Malissen N, Cassagnau E, et al. Loss of H3K27 trimethylation is not suitable for distinguishing malignant peripheral nerve sheath tumor from melanoma: a study of 387 cases including mimicking lesions. Mod Pathol. 2017;30(12):1677–87.
222. Lazova R, Seeley EH. Proteomic mass spectrometry imaging for skin cancer diagnosis. Dermatol Clin. 2017;35(4):513–9.
223. Alomari AK, Klump V, Neumeister V, Ariyan S, Narayan D, Lazova R. Comparison of the expression of vimentin and actin in spitz nevi and spitzoid malignant melanomas. Am J Dermatopathol. 2015;37(1):46–51.
224. Lazova R, Seeley EH, Keenan M, Gueorguieva R, Caprioli RM. Imaging mass spectrometry – a new and promising method to differentiate Spitz nevi from Spitzoid malignant melanomas. Am J Dermatopathol. 2012;34(1):82–90.
225. Lazova R, Seeley EH, Kutzner H, Scolyer RA, Scott G, Cerroni L, Fried I, Kozovska ME, et al. Imaging mass spectrometry assists in the classification of diagnostically challenging atypical spitzoid neoplasms. J Am Acad Dermatol. 2016;75(6):1176–86. e1174.
226. Alomari AK, Glusac EJ, Choi J, Hui P, Seeley EH, Caprioli RM, Watsky KL, Urban J, et al. Congenital nevi versus metastatic melanoma in a newborn to a mother with malignant melanoma – diagnosis supported by sex chromosome analysis and imaging mass spectrometry. J Cutan Pathol. 2015;42(10):757–64.
227. Lazova R, Yang Z, El Habr C, Lim Y, Choate KA, Seeley EH, Caprioli RM, Yangqun L. Mass spectrometry imaging can distinguish on a proteomic level between proliferative nodules within a benign congenital nevus and malignant melanoma. Am J Dermatopathol. 2017;39(9):689–95.
228. Balaban G, Herlyn M, Guerry D IV, Bartolo R, Koprowski H, Clark WH, Nowell PC. Cytogenetics of human malignant melanoma and premalignant lesions. Cancer Genet Cytogenet. 1984;11(4):429–39.
229. Sisley K, Cottam DW, Rennie IG, Parsons MA, Potter AM, Potter CW, Rees RC. Non-random abnormalities of chromosomes 3, 6, and 8 associated with posterior uveal melanoma. Genes Chromosomes Cancer. 1992;5(3):197–200.
230. Mertens F, Johansson B, Hoglund M, Mitelman F. Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms. Cancer Res. 1997;57(13):2765–80.
231. Speicher MR, Prescher G, du Manoir S, Jauch A, Horsthemke B, Bornfeld N, Becher R, Cremer T. Chromosomal gains and losses in uveal melanomas detected by comparative genomic hybridization. Cancer Res. 1994;54(14):3817–23.
232. Bastian BC, LeBoit PE, Hamm H, Brocker EB, Pinkel D. Chromosomal gains and losses in primary cutaneous melanomas detected by comparative genomic hybridization. Cancer Res. 1998;58(10):2170–5.
233. Bastian BC, Wesselmann U, Pinkel D, Leboit PE. Molecular cytogenetic analysis of Spitz nevi shows clear differences to melanoma. J Invest Dermatol. 1999;113(6):1065–9.
234. Bastian BC, Olshen AB, LeBoit PE, Pinkel D. Classifying melanocytic tumors based on DNA copy number changes. Am J Pathol. 2003;163(5):1765–70.
235. Maize JC Jr, McCalmont TH, Carlson JA, Busam KJ, Kutzner H, Bastian BC. Genomic analysis of blue nevi and related dermal melanocytic proliferations. Am J Surg Pathol. 2005;29(9):1214–20.
236. Raskin L, Ludgate M, Iyer RK, Ackley TE, Bradford CR, Johnson TM, Fullen DR. Copy number variations and clinical outcome in atypical spitz tumors. Am J Surg Pathol. 2011;35(2):243–52.
237. Chandler WM, Rowe LR, Florell SR, Jahromi MS, Schiffman JD, South ST. Differentiation of malignant melanoma from benign nevus using a novel genomic microarray with low specimen requirements. Arch Pathol Lab Med. 2012;136(8):947–55.
238. Wang L, Rao M, Fang Y, Hameed M, Viale A, Busam K, Jhanwar SC. A genome-wide high-resolution array-CGH analysis of cutaneous melanoma and comparison of array-CGH to FISH in diagnostic evaluation. J Mol Diagn. 2013;15(5):581–91.
239. Hirsch D, Kemmerling R, Davis S, Camps J, Meltzer PS, Ried T, Gaiser T. Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. Cancer Res. 2013;73(5):1454–60.
240. Gerami P, Jewell SS, Morrison LE, Blondin B, Schulz J, Ruffalo T, Matushek P IV, Legator M, et al. Fluorescence in situ hybridization (FISH) as an ancillary diagnostic tool in the diagnosis of melanoma. Am J Surg Pathol. 2009;33(8):1146–56.
241. Gammon B, Beilfuss B, Guitart J, Gerami P. Enhanced detection of spitzoid melanomas using fluorescence in situ hybridization with 9p21 as an adjunctive probe. Am J Surg Pathol. 2012;36(1):81–8.
242. Gerami P, Wass A, Mafee M, Fang Y, Pulitzer MP, Busam KJ. Fluorescence in situ hybridization for distinguishing nevoid melanomas from mitotically active nevi. Am J Surg Pathol. 2009;33(12):1783–8.
243. Pouryazdanparast P, Newman M, Mafee M, Haghighat Z, Guitart J, Gerami P. Distinguishing epithelioid blue nevus from blue nevus-like cutaneous melanoma metastasis using fluorescence in situ hybridization. Am J Surg Pathol. 2009;33(9):1396–400.
244. Gerami P, Mafee M, Lurtsbarapa T, Guitart J, Haghighat Z, Newman M. Sensitivity of fluorescence in situ hybridization for melanoma diagnosis using RREB1, MYB, Cep6, and 11q13 probes in melanoma subtypes. Arch Dermatol. 2010;146(3):273–8.
245. Isaac AK, Lertsburapa T, Pathria Mundi J, Martini M, Guitart J, Gerami P. Polyploidy in spitz nevi: a not uncommon karyotypic abnormality identifiable by fluorescence in situ hybridization. Am J Dermatopathol. 2010;32(2):144–8.
246. Gammon B, Beilfuss B, Guitart J, Busam KJ, Gerami P. Fluorescence in situ hybridization for distinguishing cellular blue nevi from blue nevus-like melanoma. J Cutan Pathol. 2011;38(4):335–41.
247. Gerami P, Beilfuss B, Haghighat Z, Fang Y, Jhanwar S, Busam KJ. Fluorescence in situ hybridization as an ancillary method for the distinction of desmoplastic melanomas from sclerosing melanocytic nevi. J Cutan Pathol. 2011;38(4):329–34.
248. Pouryazdanparast P, Haghighat Z, Beilfuss BA, Guitart J, Gerami P. Melanocytic nevi with an atypical epithelioid cell component: clinical, histopathologic, and fluorescence in situ hybridization findings. Am J Surg Pathol. 2011;35(9):1405–12.
249. Yelamos O, Busam KJ, Lee C, Meldi Sholl L, Amin SM, Merkel EA, Obregon R, Guitart J, et al. Morphologic clues and utility of fluorescence in situ hybridization for the diagnosis of nevoid melanoma. J Cutan Pathol. 2015;42(11):796–806.
250. Gerami P, Li G, Pouryazdanparast P, Blondin B, Beilfuss B, Slenk C, Du J, Guitart J, et al. A highly specific and discriminatory FISH assay for distinguishing between benign and malignant melanocytic neoplasms. Am J Surg Pathol. 2012;36(6):808–17.
251. Morey AL, Murali R, McCarthy SW, Mann GJ, Scolyer RA. Diagnosis of cutaneous melanocytic tumours by four-colour fluorescence in situ hybridisation. Pathology. 2009;41(4):383–7.
252. Vergier B, Prochazkova-Carlotti M, de la Fouchardiere A, Cerroni L, Massi D, De Giorgi V, Bailly C, Wesselmann U, et al. Fluorescence in situ hybridization, a diagnostic aid in ambiguous melanocytic tumors: European study of 113 cases. Mod Pathol. 2011;24(5):613–23.
253. North JP, Garrido MC, Kolaitis NA, LeBoit PE, McCalmont TH, Bastian BC. Fluorescence in situ hybridization as an ancillary tool in the diagnosis of ambiguous melanocytic neoplasms: a review of 804 cases. Am J Surg Pathol. 2014;38(6):824–31.

254. Requena C, Rubio L, Traves V, Sanmartin O, Nagore E, Llombart B, Serra C, Fernandez-Serra A, et al. Fluorescence in situ hybridization for the differential diagnosis between Spitz naevus and spitzoid melanoma. Histopathology. 2012;61(5):899–909.
255. Gerami P, Scolyer RA, Xu X, Elder DE, Abraham RM, Fullen D, Prieto VG, Leboit PE, et al. Risk assessment for atypical spitzoid melanocytic neoplasms using FISH to identify chromosomal copy number aberrations. Am J Surg Pathol. 2013;37(5):676–84.
256. Shen L, Cooper C, Bajaj S, Liu P, Pestova E, Guitart J, Gerami P. Atypical spitz tumors with 6q23 deletions: a clinical, histological, and molecular study. Am J Dermatopathol. 2013;35(8):804–12.
257. Newman MD, Lertsburapa T, Mirzabeigi M, Mafee M, Guitart J, Gerami P. Fluorescence in situ hybridization as a tool for microstaging in malignant melanoma. Mod Pathol. 2009;22(8):989–95.
258. Newman MD, Mirzabeigi M, Gerami P. Chromosomal copy number changes supporting the classification of lentiginous junctional melanoma of the elderly as a subtype of melanoma. Mod Pathol. 2009;22(9):1258–62.
259. Su J, Yu W, Liu J, Zheng J, Huang S, Wang Y, Qi S, Ma X, et al. Fluorescence in situ hybridisation as an ancillary tool in the diagnosis of acral melanoma: a review of 44 cases. Pathology. 2017;49(7):740–9.
260. Busam KJ, Fang Y, Jhanwar SC, Pulitzer MP, Marr B, Abramson DH. Distinction of conjunctival melanocytic nevi from melanomas by fluorescence in situ hybridization. J Cutan Pathol. 2010;37(2):196–203.
261. Bastian BC, Xiong J, Frieden IJ, Williams ML, Chou P, Busam K, Pinkel D, LeBoit PE. Genetic changes in neoplasms arising in congenital melanocytic nevi: differences between nodular proliferations and melanomas. Am J Pathol. 2002;161(4):1163–9.
262. Boi S, Leonardi E, Fasanella S, Cantaloni C, Micciolo R. The four-color FISH probe in the diagnosis of melanocytic lesions. J Eur Acad Dermatol Venereol: JEADV. 2010;24(10):1235–6.
263. Ferrara G, Misciali C, Brenn T, Cerroni L, Kazakov DW, Perasole A, Russo R, Ricci R, et al. The impact of molecular morphology techniques on the expert diagnosis in melanocytic skin neoplasms. Int J Surg Pathol. 2013;21(5):483–92.
264. Ferrara G, De Vanna AC. Fluorescence in situ hybridization for melanoma diagnosis: a review and a reappraisal. Am J Dermatopathol. 2016;38(4):253–69.
265. Onken MD, Worley LA, Ehlers JP, Harbour JW. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. Cancer Res. 2004;64(20):7205–9.
266. Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, Federman S, Miller JR 3rd, et al. The gene expression signatures of melanoma progression. Proc Natl Acad Sci U S A. 2005;102(17):6092–7.
267. Smith AP, Hoek K, Becker D. Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. Cancer Biol Ther. 2005;4(9):1018–29.
268. Jaeger J, Koczan D, Thiesen HJ, Ibrahim SM, Gross G, Spang R, Kunz M. Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. Clin Cancer Res. 2007;13(3):806–15.
269. Koh SS, Opel ML, Wei JP, Yau K, Shah R, Gorre ME, Whitman E, Shitabata PK, et al. Molecular classification of melanomas and nevi using gene expression microarray signatures and formalin-fixed and paraffin-embedded tissue. Mod Pathol. 2009;22(4):538–46.
270. Jonsson G, Busch C, Knappskog S, Geisler J, Miletic H, Ringner M, Lillehaug JR, Borg A, et al. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. Clin Cancer Res. 2010;16(13):3356–67.
271. Scatolini M, Grand MM, Grosso E, Venesio T, Pisacane A, Balsamo A, Sirovich R, Risio M, et al. Altered molecular pathways in melanocytic lesions. Int J Cancer. 2010;126(8):1869–81.
272. Mauerer A, Roesch A, Hafner C, Stempfl T, Wild P, Meyer S, Landthaler M, Vogt T. Identification of new genes associated with melanoma. Exp Dermatol. 2011;20(6):502–7.
273. Harbst K, Staaf J, Lauss M, Karlsson A, Masback A, Johansson I, Bendahl PO, Vallon-Christersson J, et al. Molecular profiling reveals low- and high-grade forms of primary melanoma. Clin Cancer Res. 2012;18(15):4026–36.
274. Clarke LE, Warf MB, Flake DD 2nd, Hartman AR, Tahan S, Shea CR, Gerami P, Messina J, et al. Clinical validation of a gene expression signature that differentiates benign nevi from malignant melanoma. J Cutan Pathol. 2015;42(4):244–52.
275. Gerami P, Cook RW, Russell MC, Wilkinson J, Amaria RN, Gonzalez R, Lyle S, Jackson GL, et al. Gene expression profiling for molecular staging of cutaneous melanoma in patients undergoing sentinel lymph node biopsy. J Am Acad Dermatol. 2015;72(5):780–5. e783.
276. Gerami P, Cook RW, Wilkinson J, Russell MC, Dhillon N, Amaria RN, Gonzalez R, Lyle S, et al. Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. Clin Cancer Res. 2015;21(1):175–83.
277. Meves A, Nikolova E, Heim JB, Squirewell EJ, Cappel MA, Pittelkow MR, Otley CC, Behrendt N, et al. Tumor cell adhesion as a risk factor for sentinel lymph node metastasis in primary cutaneous melanoma. J Clin Oncol. 2015;33(23):2509–15.
278. Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, Wolter P, Patel P, et al. Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. Oncotarget. 2015;6(13):11683–93.
279. Warf MB, Flake DD 2nd, Adams D, Gutin A, Kolquist KA, Wenstrup RJ, Roa BB. Analytical validation of a melanoma diagnostic gene signature using formalin-fixed paraffin-embedded melanocytic lesions. Biomark Med. 2015;9(5):407–16.
280. Cockerell CJ, Tschen J, Evans B, Bess E, Kidd J, Kolquist KA, Rock C, Clarke LE. The influence of a gene expression signature on the diagnosis and recommended treatment of melanocytic tumors by dermatopathologists. Medicine (Baltimore). 2016;95(40):e4887.
281. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189–96.
282. Clarke LE, Flake DD 2nd, Busam K, Cockerell C, Helm K, McNiff J, Reed J, Tschen J, et al. An independent validation of a gene expression signature to differentiate malignant melanoma from benign melanocytic nevi. Cancer. 2017;123(4):617–28.
283. Sominidi-Damodaran S, Guo R, Meves A, Bridges AG. Expanded traditional melanoma FISH testing versus CAP-QPCR to identify high-risk melanocytic lesions. Int J Dermatol. 2017;56(9):e182–4.
284. Cserni G, Chmielik E, Cserni B, Tot T The new TNM-based staging of breast cancer. Virchows Arch. 2018;472(5):697–703.
285. Plasseraud KM, Wilkinson JK, Oelschlager KM, Poteet TM, Cook RW, Stone JF, Monzon FA. Gene expression profiling in uveal melanoma: technical reliability and correlation of molecular class with pathologic characteristics. Diagn Pathol. 2017;12(1):59.
286. Clemente CG, Mihm MC Jr, Bufalino R, Zurrida S, Collini P, Cascinelli N. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. Cancer. 1996;77(7):1303–10.
287. Kubica AW, Brewer JD. Melanoma in immunosuppressed patients. Mayo Clin Proc. 2012;87(10):991–1003.
288. Kalialis LV, Drzewiecki KT, Klyver H. Spontaneous regression of metastases from melanoma: review of the literature. Melanoma Res. 2009;19(5):275–82.
289. Menzies SW, McCarthy WH. Complete regression of primary cutaneous malignant melanoma. Arch Surg. 1997;132(5):553–6.

290. Harlin H, Meng Y, Peterson AC, Zha Y, Tretiakova M, Slingluff C, McKee M, Gajewski TF. Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment. Cancer Res. 2009;69(7):3077–85.
291. Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, Gajewski TF. Up-regulation of PD-L1, IDO, and T(regs) in the melanoma tumor microenvironment is driven by CD8(+) T cells. Sci Transl Med. 2013;5(200):200ra116.
292. Kerkar SP, Restifo NP. Cellular constituents of immune escape within the tumor microenvironment. Cancer Res. 2012;72(13):3125–30.
293. Jacobs JF, Nierkens S, Figdor CG, de Vries IJ, Adema GJ. Regulatory T cells in melanoma: the final hurdle towards effective immunotherapy? Lancet Oncol. 2012;13(1):e32–42.
294. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial EST 1684. J Clin Oncol. 1996;14(1):7–17.
295. Kirkwood JM, Ibrahim JG, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, Smith TJ, Rao U, et al. High- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. J Clin Oncol. 2000;18(12):2444–58.
296. Wheatley K, Ives N, Hancock B, Gore M, Eggermont A, Suciu S. Does adjuvant interferon-alpha for high-risk melanoma provide a worthwhile benefit? A meta-analysis of the randomised trials. Cancer Treat Rev. 2003;29(4):241–52.
297. Mocellin S, Pasquali S, Rossi CR, Nitti D. Interferon alpha adjuvant therapy in patients with high-risk melanoma: a systematic review and meta-analysis. J Natl Cancer Inst. 2010;102(7):493–501.
298. Rosenberg SA. IL-2: the first effective immunotherapy for human cancer. J Immunol. 2014;192(12):5451–8.
299. Atkins MB, Lotze MT, Dutcher JP, Fisher RI, Weiss G, Margolin K, Abrams J, Sznol M, et al. High-dose recombinant interleukin 2 therapy for patients with metastatic melanoma: analysis of 270 patients treated between 1985 and 1993. J Clin Oncol. 1999;17(7):2105–16.
300. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med. 2010;363(8):711–23.
301. Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, Lebbe C, Baurain JF, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. N Engl J Med. 2011;364(26):2517–26.
302. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, Patt D, Chen TT, et al. Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. J Clin Oncol. 2015;33(17):1889–94.
303. Eggermont AM, Chiarion-Sileni V, Grob JJ, Dummer R, Wolchok JD, Schmidt H, Hamid O, Robert C, et al. Adjuvant ipilimumab versus placebo after complete resection of high-risk stage III melanoma (EORTC 18071): a randomised, double-blind, phase 3 trial. Lancet Oncol. 2015;16(5):522–30.
304. Flies DB, Sandler BJ, Sznol M, Chen L. Blockade of the B7-H1/PD-1 pathway for cancer immunotherapy. Yale J Biol Med. 2011;84(4):409–21.
305. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science. 2011;331(6024):1565–70.
306. Topalian SL, Sznol M, McDermott DF, Kluger HM, Carvajal RD, Sharfman WH, Brahmer JR, Lawrence DP, et al. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. J Clin Oncol. 2014;32(10):1020–30.
307. Weber JS, D'Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B, Hoeller C, Khushalani NI, et al. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. Lancet Oncol. 2015;16(4):375–84.
308. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, Hassel JC, Rutkowski P, et al. Nivolumab in previously untreated melanoma without BRAF mutation. N Engl J Med. 2015;372(4):320–30.
309. Postow MA, Chesney J, Pavlick AC, Robert C, Grossmann K, McDermott D, Linette GP, Meyer N, et al. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. N Engl J Med. 2015;372(21):2006–17.
310. Hodi FS, Chesney J, Pavlick AC, Robert C, Grossmann KF, McDermott DF, Linette GP, Meyer N, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. Lancet Oncol. 2016;17(11):1558–68.
311. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob JJ, Cowey CL, Lao CD, Wagstaff J, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. N Engl J Med. 2017;377(14):1345–56.
312. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell. 2016;165(1):35–44.
313. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci Transl Med. 2017;9(379) pii: eaah3560. https://doi.org/10.1126/scitranslmed.aah3560.
314. Hugo W, Shi H, Sun L, Piva M, Song C, Kong X, Moriceau G, Hong A, et al. Non-genomic and immune evolution of melanoma acquiring MAPKi resistance. Cell. 2015;162(6):1271–85.
315. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, Torrejon DY, Abril-Rodriguez G, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. N Engl J Med. 2016;375(9):819–29.
316. Shin DS, Zaretsky JM, Escuin-Ordinas H, Garcia-Diaz A, Hu-Lieskovan S, Kalbasi A, Grasso CS, Hugo W, et al. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. Cancer Discov. 2017;7(2):188–201.
317. Rosenberg SA, Packard BS, Aebersold PM, Solomon D, Topalian SL, Toy ST, Simon P, Lotze MT, et al. Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. N Engl J Med. 1988;319(25):1676–80.
318. Wu R, Forget MA, Chacon J, Bernatchez C, Haymaker C, Chen JQ, Hwu P, Radvanyi LG. Adoptive T-cell therapy using autologous tumor-infiltrating lymphocytes for metastatic melanoma: current status and future outlook. Cancer J. 2012;18(2):160–75.
319. Deniger DC, Kwong ML, Pasetto A, Dudley ME, Wunderlich JR, Langhan MM, Lee CR, Rosenberg SA. A pilot trial of the combination of vemurafenib with adoptive cell therapy in patients with metastatic melanoma. Clin Cancer Res. 2017;23(2):351–62.
320. Cooper ZA, Frederick DT, Juneja VR, Sullivan RJ, Lawrence DP, Piris A, Sharpe AH, Fisher DE, et al. BRAF inhibition is associated with increased clonality in tumor-infiltrating lymphocytes. Oncoimmunology. 2013;2(10):e26615.
321. Frederick DT, Piris A, Cogdill AP, Cooper ZA, Lezcano C, Ferrone CR, Mitra D, Boni A, et al. BRAF inhibition is associated with enhanced melanoma antigen expression and a more favorable tumor microenvironment in patients with metastatic melanoma. Clin Cancer Res. 2013;19(5):1225–31.
322. Ott PA, Fritsch EF, Wu CJ, Dranoff G. Vaccines and melanoma. Hematol Oncol Clin North Am. 2014;28(3):559–69.

323. Lennerz V, Fatho M, Gentilini C, Frye RA, Lifke A, Ferel D, Wolfel C, Huber C, et al. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. Proc Natl Acad Sci U S A. 2005;102(44):16013–8.
324. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, Prickett TD, Gartner JJ, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat Med. 2016;22(4):433–8.
325. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, Parkhurst MR, Ankri C, et al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. J Clin Invest. 2015;125(10):3981–91.
326. Prickett TD, Crystal JS, Cohen CJ, Pasetto A, Parkhurst MR, Gartner JJ, Yao X, Wang R, et al. Durable complete response from metastatic melanoma after transfer of autologous T cells recognizing 10 mutated tumor antigens. Cancer Immunol Res. 2016;4(8):669–78.
327. Pritchard AL, Burel JG, Neller MA, Hayward NK, Lopez JA, Fatho M, Lennerz V, Wolfel T, et al. Exome sequencing to predict neoantigens in melanoma. Cancer Immunol Res. 2015;3(9):992–8.
328. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature. 2017;547(7662):217–21.
329. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, Bukur V, Tadmor AD, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017;547(7662):222–6.
330. Palm NW, de Zoete MR, Flavell RA. Immune-microbiota interactions in health and disease. Clin Immunol. 2015;159(2):122–7.
331. Vetizou M, Pitt JM, Daillere R, Lepage P, Waldschmitt N, Flament C, Rusakiewicz S, Routy B, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science. 2015;350(6264):1079–84.
332. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359(6371):97–103.
333. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science. 2018;359(6371):104–8.
334. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillere R, Fluckiger A, Messaoudene M, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science. 2018;359(6371):91–7.

Part V

Genomic Applications in Inherited and Infectious Diseases, Pharmacogenomics, and the Microbiome

Genomic Applications in Inherited Genetic Disorders

35

Bryan L. Krock, Rong Mao, Tatiana Tvrdik, D. Hunter Best, and Elaine Lyon

Introduction

The enhanced capabilities of next-generation sequencing (NGS) have changed how clinical laboratory geneticists approach test design and application. For instance, some disorders display extreme locus heterogeneity and can be caused by alterations in one of many genes. Nonsyndromic retinitis pigmentosa (RP) is an example of locus heterogeneity and is caused by germline variants in as many as 86 different genes [30]. Traditionally, molecular genetic testing for this disorder would use a tiered approach where initial testing would prioritize the most frequently causative gene based on an individual's clinical picture and family history. For an individual with suggested autosomal dominant RP, testing would likely have begun with *RHO* gene sequencing, which accounts for 20–30% of autosomal dominant retinitis pigmentosa [30]. Patients without a causative variant identified in *RHO* gene would be tested for the next most common gene and so on until either a causative variant was found or the costs associated with additional testing outweighed the likelihood of success. The latter scenario is quickly realized for many patients, leaving them without a genetic diagnosis. This phenomenon has commonly been referred to as the "diagnostic odyssey," which can be frustrating for patients and families as well as complicating genetic counseling for family members. Additionally, in the pre-NGS era, genetic testing often was not available for very rare causes of a disorder, because these extremely low-volume tests are not financially viable for most clinical laboratories. NGS-based tests often represent the only diagnostic option for some patients, particularly those who are clearly affected with a genetic disorder but display an atypical presentation or have a rare, poorly-studied, genetic disorder. In these cases, the shotgun approach to genetic testing frequently proves to be an indispensable tool for clinical diagnosis. The clinical utility of NGS-based diagnostic testing has been clearly demonstrated for RP and, more broadly, inherited retinal disease [27].

Academic and commercial laboratories have rapidly adopted NGS technology for clinical diagnostics, and it has surpassed traditional methods for essentially all clinical indications where panel-based testing is appropriate. This is due, in part, to the ability of NGS to simultaneously sequence all genes associated with a disorder at prices and turnaround times that are far superior to traditional Sanger sequencing. Improvements in technology, wet-bench automation, computational tools, accessibility of the expertise and technology, and the maturation of clinical NGS in diagnostic laboratories have driven the widespread use for clinical NGS testing. The current applications of NGS genetic tests demonstrate its broad utility, as they can be appropriate to every stage of life – from preconception or fetal screening to newborn, childhood, and adult-onset disorders.

Clinical NGS-based tests for inherited diseases can focus on a subset of genes, all coding regions (exome), or even the whole genome. Each of these approaches has advantages and disadvantages that should be considered with regard to both test development and clinical application. Multigene panels generally focus on a specific disorder that displays genetic heterogeneity or a group of disorders that have overlapping genetics and/or phenotypes (see Table 35.1). NGS gene panels commonly have been developed for genetically heterogeneous disorders, as they are able to sequence hundreds of genes at very high coverage, imparting high sensitivity and specificity to the assay. This high depth of coverage, along with the digital nature of NGS, allows identification of somatic mosaicism with greater sensitivity than Sanger sequencing. This advantage has clear utility for disorders commonly associated with somatic mosaicism, such as tuberous sclerosis or Proteus syndrome [57, 113]. In addi-

B. L. Krock
The Children's Hospital of Philadelphia Division of Genomic Diagnostics, Perelman School of Medicine at the University of Pennslyvania, Department of Pathology and Laboratory Medicine, Philadelphia, PA, USA

R. Mao · T. Tvrdik · D. H. Best · E. Lyon (✉)
ARUP Laboratories, Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA
e-mail: lyone@aruplab.com

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_35

Table 35.1 Examples of clinically available NGS tests for inherited or germline disorders

NGS test	Age of application		
	Fetal	Infant/child	Adult
Amyotrophic lateral sclerosis			X
Aortopathies		X	X
Autism spectrum disorder		X	
Brain malformations/holoprosencephaly	X	X	
Cardiomyopathy		X	X
Chromosomal aneuploidies from cffDNA	X		
Primary ciliary dyskinesia	X	X	
Congenital orders of glycosylation		X	
Epilepsy		X	X
Eye disorders		X	X
Immunodeficiency		X	X
Inherited cancer			X[a]
Mitochondrial disorders		X	X
Neuromuscular		X	X
Noonan syndrome and related disorders	X	X	X
X-linked intellectual disability		X	
Skeletal dysplasia	X	X	
Clinical exome sequencing	X	X	X
Clinical genome sequencing	X	X	X

Cff cell-free fetal
[a]For adult cancers. Children may be tested for pediatric or early-onset cancers

tion, recent reports suggest somatic mosaicism may be more common than previously appreciated [7, 86]. For instance, a patient with Cowden syndrome – a disorder not commonly associated with mosaicism – had tested negative for *PTEN* variants by Sanger sequencing DNA from peripheral blood. Application of a NGS panel containing *PTEN* to the same sample revealed a low frequency frameshift *PTEN* variant that was subsequently found in the heterozygous state in skin fibroblasts, confirming somatic mosaicism [86]. Large-scale studies have revealed apparent somatic mosaicism is relatively common in individuals affected with intellectual disability, autism, or epilepsy at a frequency ranging from 3.5% to 7% [2, 35, 104]. The utility of the quantitative nature of NGS extends to family members subjected to testing as either the result of trio-based testing or targeted follow-up testing, as apparent parental mosaicism has been observed in up to 10% of parents [104]. Importantly, Sanger-based methodologies would likely have missed low-level mosaic variants in parents and probands, resulting in missed diagnoses and incorrect recurrence risk estimates for families already dealing with often devastating pediatric disease. Undoubtedly, continued application of NGS testing will likely improve our understanding of mosaicism in many inherited genetic disorders.

In addition to genetically heterogeneous disorders, diseases that display overlapping phenotypic spectra are also good candidates for NGS panels, as sequencing of multiple genes may be required to clarify a clinical diagnosis. The aortopathies are an example and will be discussed in greater detail later in this chapter. While an NGS panel still costs more than sequencing of medium and small single genes, gene panels have largely replaced historical cascading genetic testing algorithms. In some cases, NGS panels have become part of a testing algorithm and complement existing Sanger sequencing tests. For example, pathogenic variants in *GJB2* and deletions of *GJB6* account for up to 20% of non-syndromic hearing loss cases [23]. Accordingly, sequential *GJB2* sequencing and *GJB6* targeted deletion analysis prior to an NGS panel offers a cost-effective solution and avoids unnecessary testing costs for individuals with the most common genetic etiology of hearing loss. In addition to their advantages over Sanger-based methods, NGS panels also have specific benefits over clinical exome sequencing (CES) and clinical genome sequencing (CGS). First, bioinformatic processing and test interpretation for NGS panels are less challenging, as fewer genes are analyzed, and fewer variants are recovered. Additionally, NGS panels will seldom yield incidental findings, making clinical reports more straightforward. Next, CES and CGS tests have gaps in coverage, even in known disease-causing genes. These gaps in NGS panels can be filled with Sanger sequencing to achieve complete coverage of targeted genes. Also, depth of sequencing is greater for NGS panels, which are therefore more sensitive for detecting somatic mosaicism. Finally, the ability to multiplex more samples and use smaller capacity platforms makes NGS panels cheaper and faster than CES and CGS tests. Because fewer genes are analyzed and fewer variants are recovered, bioinformatics processing and test interpretation are straightforward and will seldom yield incidental findings. Sequencing depth is greater, thus achieving greater sensitivity for detecting somatic mosaicism while still maintaining the ability to multiplex more samples, thereby reducing costs and time to results.

Exome sequencing (ES) involves selective enrichment for and sequencing of the entire coding region of the genome. While it represents only approximately 1.5% of the total genome, pathogenic variants in the exome account for about 85% of all known disease-causing alterations [96]. By selectively sequencing the coding region of the genome, clinical molecular geneticists can interrogate all known genes at a reasonable cost. ES has been applied in a research setting for gene discovery with great success, often enabling the identification of a disease-causing gene by studying a single individual or family. In this way, CES can blur the lines between a clinical test and research assay; novel disease-causing genes are routinely identified in patients referred for CES. Although there are disadvantages of CES compared to NGS panels, its principal advantage is that it provides a broader view of a patient's genome sequence. Along these lines, CES does not require a priori suspicion of a specific

genetic disorder, and accordingly it represents the only molecular diagnostic option for some patients. CES may also yield unexpected genetic diagnoses that shape the clinical diagnosis of patients. For instance, CES may identify two or more disorders contributing to the overall phenotype, which has been observed at rates ranging from 3% to 5% in large-scale studies [93, 130]. In an early demonstration of the power of ES, a pair of siblings studied for the genetic cause of Miller syndrome was shown to have two recessive disorders: Miller syndrome with the concurrent identification of the responsible gene and primary ciliary dyskinesia [76].

CES has been more widely adopted for clinical use than CGS, which involves sequencing the vast majority of the annotated genome. This is mainly because genetic alterations in the coding portion of the genome (exome) are easier to interpret than intragenic and intergenic regions, particularly when they cause direct changes to the protein sequence or alter canonical splice donor/acceptor sites. Nonetheless, CGS is available at multiple commercial and academic diagnostic laboratories. In contrast to CES and NGS panels, CGS does not involve target enrichment and, accordingly, provides more even sequence coverage with fewer gaps than CES [71]. Moreover, current enrichment technologies tend to preferentially enrich for reference alleles at heterozygous sites, potentially producing false-negative variant calls [73]. These advantages of CGS make it slightly superior to exome sequencing for the identification of coding SNVs [8]. However, cost constraints have prevented the wholesale transition from CES to CGS. The technical advantages suggest CGS should exhibit improved diagnostic yield over CES, though appropriate large-scale studies have not been reported to fully quantify the difference. Recent studies of GS are consistent with its suspected advantages, as it exhibits improved diagnostic yield over capture-based methodologies [28]. While dramatic technological advances have enabled clinical geneticists to probe the entire genome, the vast majority of noncoding regions are uninterpretable, and the large amounts of extra data take up tremendous storage space and computational time. Given the difficulty in interpreting variants outside of coding and intron/exon boundary regions, CGS is typically analyzed in a manner similar to an exome. If the causative variant(s) is/are not found, exploring variants in other genomic regions could be attempted, though the interpretation of these variants will be hindered by the lack of functional evidence to support clinical interpretation. Despite these challenges, many view a transition to CGS as an inevitable next step in clinical diagnostics, though the point at which CGS costs become equivalent to CES may be the primary factor governing this transition. Due to the similar nature of CES and CGS applications, they will be discussed together for the remainder of this chapter.

Similar to other clinical genetic tests, selecting the appropriate tests and patients for NGS-based studies is critical to maximize diagnostic yield and minimize unproductive genetic testing. As NGS panels are targeted to a phenotype or group of disorders, patients with a clear clinical presentation consistent with a particular disorder are good candidates for a disease-specific NGS panel. In this way, patients for NGS panels are evaluated in a similar manner to traditional genetic tests. Some clinicians may prefer CES instead of gene panels because of the broader coverage and rationalize that CES will cover all of the genes in an NGS panel. However, most NGS panels will achieve complete coverage of the target genes (although filling in with Sanger sequencing may be needed) and should detect known variants (provided that the type of variant is detectable by NGS), giving a higher negative predictive value. In contrast, CES may have low coverage in clinically relevant genes and miss important pathogenic variants as a result. Moreover, poorly covered genes may not be clearly conveyed in a clinical report, potentially giving a false confidence in a test result. The American College of Medical Genetics and Genomics (ACMG) has outlined indications for CES and CGS diagnostic testing [1]. CES/CGS should be considered for patients in whom the phenotype or family history strongly implicates a disorder with a genetic etiology, but the phenotype does not correspond with a specific disorder for which a genetic test targeting a specific gene is available on a clinical basis. CES/CGS is also indicated for defined genetic disorders that demonstrate a high degree of genetic heterogeneity, making analysis of multiple genes simultaneously a more practical approach. Additionally, patients with an apparent genetic disorder, but who have failed to obtain a genetic diagnosis with available genetic tests, are candidates for CES/CGS. Lastly, a fetus with a likely genetic disorder that has not obtained a diagnosis with specific genetic tests available for the phenotype is also a candidate for CES/CGS. Given the costs and technical/analytical limitations of CES/CGS, patients should be tested for copy number variations by a cytogenomic microarray prior to CES, as this is typically a more cost-effective testing paradigm. As copy number calling from NGS-based tests improves, and lower sequencing costs reduce the overall prices of these tests, it may become more practical for CES/CGS to serve as a first-tier diagnostic test. These criteria will obviously select for patients with novel genetic disorders or pathogenic alterations in genes not previously associated with a disease. In spite of this challenge, the diagnostic yield for clinical exome sequencing is remarkably high, typically of 25–31% across a broad spectrum of clinical indications [53, 93, 129, 130].

Although NGS represents an extremely powerful technology for genetic diagnostics, its limitations must be considered before its clinical application. First, CES and NGS panels cannot reliably detect chromosomal translocations, large inversions, and other copy-neutral structural alterations. Numerous approaches to assess copy number variations

from targeted NGS, CES, and CGS have been developed. The reported sensitivity and specificity of these approaches have been generally variable, though careful validation has revealed these approaches can exhibit high sensitivity and are appropriate for clinical use in targeted NGS, CES, and CGS [29, 94]. Significant technical and analytical challenges arise from this practice which should be approached with high rigor and orthogonal confirmation where appropriate. NGS, similar to traditional Sanger-based methods, often cannot amplify GC-rich regions and can suffer from allele dropout. Accordingly, clinically relevant promoter or exon 1 variants can potentially be missed. Long repetitive elements, such as those observed in disorders with triplet repeat expansions, are particularly problematic for short-read NGS, both for the sequencing chemistry and bioinformatic sequencing alignment. When testing for disorders that can be caused by both repeat expansion and sequence variants, it is important to rule out the repeat expansion before an NGS test is ordered if that methodology is not used as an ancillary assay with the NGS panel. For instance, in those patients tested for X-linked intellectual disability, triplet repeat expansion in *FMR1* should be ruled out prior to NGS-based tests. NGS panels and exome sequencing also rely on the enrichment of targeted exonic regions, the process of which is not 100% efficient, leaving areas with low or no coverage. Although Sanger sequencing can supplement deficient regions in NGS panels to achieve 100% coverage, this approach is not feasible for CES. Clinical groups have sought to enhance the coverage of CES by augmenting the enrichment reagents for known clinically relevant genes. These approaches represent improvements over initial clinical exome sequencing, and similar advancements in commercially available reagents are likely to continue. Interpretation of rare and novel sequence variants is already a significant challenge for traditional Sanger sequencing-based tests. As the number of genes sequenced by NGS panels and by CES/CGS is orders of magnitude greater than targeted variant studies and Sanger sequencing, there is a proportional increase in novel variants in NGS-based assays. These variants are challenging to evaluate and take significant amounts of time for manual review; as a result, they represent a significant interpretive burden that will be encountered with nearly all CES/CGS cases. The inevitable increase in variants of uncertain significance (VUS) in clinical reports represents another difficulty, as they may be understood differently by ordering physicians. One approach to minimize reporting of VUS is to perform CES/CGS on trios to enable the bioinformatics sorting of variants based on the mode of inheritance. This approach clearly enhances the diagnostic yield of CES/CGS [93]. However, in our experience with CES, testing is often ordered only for the proband. This results in an increased number of variants that have to be manually evaluated, increasing the time required for analysis. Additionally, follow-up testing of family members to confirm de novo or biparental inheritance can be challenging for proband-only CES. While not insurmountable, these challenges increase the turnaround times and costs associated with these diagnostic tests. The medical literature is evaluating novel variants in both known and novel genes at a record pace, which will likely alter the classification of previously reported variants and reveal clinically relevant insights that were unknown at the time of the original CES/CGS analysis. Integrating these new findings into previously reported cases is and will remain challenging, as standards of best practice have not been established, and the tools to do so are not commonly available. Recent studies have shown that approximately 10% of previously negative exome cases find a molecular diagnosis upon reanalysis, demonstrating a clinical need for this service [124]. Clearly in the development and implementation of methods accurately and sustainably to implement clinical exome, reanalysis will be important for clinical laboratories and clinicians alike going forward.

Expanded Carrier Testing Applications

Carrier testing for inherited disorders has traditionally been performed for monogenetic, autosomal recessive disorders with very high allele frequencies, such as cystic fibrosis, or in populations with elevated risk for genetic disease, based on either family history or ethnicity, most notably the Ashkenazi Jewish (AJ) population [39]. Appropriate application of preconception screening can be a profoundly effective public health tool, as demonstrated by the >90% reduction in the incidence of Tay-Sachs disease in the American AJ population [45, 50, 74]. However, carrier screening outside of the few aforementioned situations has been impractical for technical and logistical reasons – until recently. NGS has enabled clinicians and laboratory professionals to consider expanded panethnic carrier screening, where prospective parents could be screened for the vast majority of known deleterious recessive variants in a time- and cost-effective manner. This new approach could reduce the overall prevalence of severe genetic disorders, which collectively account for ~20% of infant mortality [18], and facilitate the prenatal diagnosis of genetic disorders for at-risk pregnancies, leading to early intervention and improved outcomes for affected infants. Proof of concept studies were first reported in 2011, where a targeted NGS panel was designed to screen 7717 genomic regions for known disease-causing variants in 437 genes that cause 448 severe recessive childhood diseases [9]. Preconception screening was performed in 104 unrelated individuals, and the authors found the average carrier burden for severe pediatric recessive variants was 2.8 per individual, consistent with previous estimates [52]. Importantly, the authors presented cost analyses that projected an overall analytical cost of $378, equating to less than $1 per condition

tested. However, this figure does not include costs associated with test interpretation, reporting, and other ancillary costs associated with a complex genetic test. Another study included comparison of traditional targeted genotyping assays and NGS for carrier screening in 506 individuals, of whom 288 were reported as carriers of at least one condition and eight couples were carriers for the same disorder. Although no additional pathogenic variants were detected by NGS in diseases routinely screened in the Ashkenazi Jewish population, 26.5% of carrier results and two carrier couples were identified only by NGS [3]. A simulated model study of 1,000, 000 couples showed that carrier screening using NGS would reduce costs by approximately $13 million and offer greater benefit in clinical outcomes and lower total healthcare cost as compared with traditional genotyping [4]. A randomized control trial has been designed to evaluate genomic clinical sequencing for women and partners considering a pregnancy that will compare cost, utilization, and psychosocial impacts of usual care vs. genomic carrier screening [46].

In view of the large scale of preconception screening with the potential to dramatically reduce the incidence of mortality and morbidity due to genetic disease, significant hurdles to the broad application of this screening paradigm remain. Variant classification is a substantial challenge for all NGS-based tests, even when clear phenotypes and modes of inheritance are known. Interpretation of novel variants in the absence of a phenotype adds another layer of complexity to this already difficult process. One study found that 27% of pathogenic variants cited in the literature were found to be common polymorphisms or misannotated [9], a critical problem for all NGS-based testing that necessitates careful examination of all variants reported. This will add to the already significant time and cost of data analysis. When screening asymptomatic carriers, rare nonconservative and nonsynonymous variants will likely be identified in many patients and scored as VUS. The increased psychological stress for parents and perhaps increased reliance upon preimplantation genetic diagnosis due to VUS reports need to be carefully considered for expanded carrier screening. Moreover, the concept of variable penetrance needs to be clearly conveyed to patients concerning appropriate genetic disorders. One approach to minimize many of the aforementioned concerns is to restrict analysis to known pathogenic variants, using NGS as a multigene, multivariant, but targeted panel. However, rare, novel pathogenic variants may be missed with this method. If a pathogenic variant is identified in a targeted panel, full gene analysis should be considered for the reproductive partner but with the same analytic and interpretive issues previously described. Although a comprehensive single test, this technology in its current form will not be able to identify some of the most frequent disease-causing variants, of which triplet repeat expansion fragile X syndrome is an example. This limitation of NGS needs to be clearly communicated to both patients and ordering physicians. Ultimately, expanded carrier screening is an area likely to undergo dramatic changes due to NGS. In the face of drastic change, clinical guidelines need to be established to ensure appropriate application, reporting, and counseling of NGS-based expanded carrier screening.

Fetal Applications

Cell-Free Fetal DNA in Maternal Plasma

Testing of fetal genotypes has traditionally relied upon invasive sampling of fetal cellular material through amniocentesis or chorionic villus sampling, both of which carry a small but significant risk for fetal loss. Therefore, a noninvasive means of fetal DNA sampling for genetic evaluation has long been pursued. Fetal lymphocytes are present in very small numbers in maternal blood, but their extreme rarity, challenges to their purification, and concerns about persistence after birth have precluded their use in clinical testing. Cell-free fetal DNA (cffDNA) was identified in maternal plasma and serum, suggesting that this could be an easily accessible, noninvasive source of fetal DNA for genetic testing [60]. This source of cffDNA is the result of normal placental cell apoptosis, which releases highly fragmented DNA representative of the fetal genotype into the maternal circulation. Importantly, cffDNA can be detected from 4 weeks gestation until birth [44], making it amenable to genetic testing in at-risk pregnancies. The entire genome is represented in cffDNA [59], suggesting that this platform is appropriate for molecular testing for the vast majority of inherited disorders. Moreover, cffDNA is highly unstable in the maternal circulation and is cleared soon after birth, meaning that a sample will not be contaminated with fetal DNA from prior pregnancies [62]. Fetal DNA represents 5–10% of total plasma DNA, with the remainder maternal in origin [66]. While the fraction of cffDNA increases with fetal age, pure fetal DNA cannot be extracted from maternal serum, and the maternal DNA background has been the major hurdle to the use of cffDNA for molecular diagnostics. Accordingly, the first application of cffDNA testing was to detect fetal Y chromosome sequences in maternal plasma, circumventing the issue of contaminating maternal DNA [60]. Other early clinical applications of cffDNA also reflect this constraint, as they include determining fetal Rh D status in Rh D negative mothers and detecting paternally inherited autosomal dominant and recessive variants [22]. Initial efforts to broaden the applicability of cffDNA relied on allelic heterozygosity between the fetus and mother to determine fetal chromosomal dosage in testing for fetal aneuploidy, primarily trisomies 21, 18, and 13 [61]. Traditional methods, however, require

heterozygosity at many loci on each chromosome, making clinical assay design challenging. Moreover, locus-specific approaches require large amounts of DNA for multiple PCR reactions to achieve analytical precision [17], making them unrealistic for clinical application. NGS alleviates some of the constraints encountered by traditional technologies for cffDNA testing. The massively parallel nature of NGS can intrinsically sequence hundreds of thousands of sites with great depth in a quantitative manner. This enables very small changes in chromosome DNA frequency to be detected, as would be predicted for a trisomy 21 fetus contributing 5% of the total plasma DNA content. Thus, by counting NGS sequencing reads from maternal plasma mapping to chromosomes 21, 18, and 13 and assessing the over- or underrepresentation of these chromosomes in maternal plasma, two initial groups were able to accurately identify trisomy 21 fetuses and then expand to identify trisomy 13 and trisomy 18, demonstrating this approach is applicable to other common chromosomal aneuploidies [17, 31]. Recently, microdeletion syndromes have been included in testing. Options for detecting chromosome 22q11.2 deletions associated with DiGeorge syndrome, as well as others such as cri du chat (chromosome 5p), and Angelman/Prader Willi regions (chromosome 15q11-q13) are available. However, given the rarity of these conditions in the population, the positive predictive value is low, reportedly 44.2% for 22q11.2 and 31.7% combined for four other microdeletion syndromes [68]. Several genetic diagnostic companies now offer noninvasive prenatal testing of the common fetal aneuploidies by NGS analysis of cffDNA, as well as known microdeletion syndromes.

Fetal screening for severe, early-onset diseases has been described. Many of the diseases included are typically caused by dominant, de novo pathogenic variants. Alternately, paternally inherited variants are easily discriminated, allowing for an initial screen for recessive disorders. These lend themselves to cffDNA testing because they are not expected to be present in the mother. Conditions such as craniosynostosis, skeletal dysplasia, cardiac defects, and multiple congenital anomalies may be suspected by ultrasound findings, so although offered as a screening test in the absence of symptoms, this test may be used as diagnostic testing when clinical findings are present [6].

Research into methylation patterns differing between the mother and the fetus has been described to expand the number of disorders detected [117]. Such research may expand noninvasive testing to other disorders.

Remarkably, the entire genome of an 18.5-week gestation fetus has been sequenced from cffDNA isolated from maternal plasma [47]. This approach required extensive, specialized bioinformatic processing and integration of maternal and paternal genomic data and is not likely to become clinically available in the current form. However, it demonstrates that with the appropriate techniques, any region of the fetal genome can be queried by NGS. Accordingly, NGS-based testing of cffDNA is likely to expand to gene sequencing and copy number variant detection in the future.

Skeletal Dysplasia

Skeletal dysplasias are a heterogeneous group of disorders characterized by abnormal bone or cartilage growth. Over 300 types of skeletal dysplasias have been described with causal variants in over 200 genes [120]. Although rare disorders individually, they have an overall prevalence of approximately 3 per 10,000 births and 20.0 per 10,000 stillbirths [103]. They vary greatly in severity, with severe forms such as thanatophoric dysplasia being lethal in the prenatal or neonatal period and mild forms such as hypochondroplasia not detected until childhood. Routine ultrasound monitoring can identify skeletal dysplasias during fetal development; however, a specific diagnosis is challenging due to the limitations of noninvasive imaging procedures. Although some dysplasias are diagnosed by ultrasound and confirmed by molecular analysis prenatally, many may remain undiagnosed. Yet a diagnosis is important for appropriate prenatal and postnatal management, as well as determining recurrence risk. Most importantly, an accurate diagnosis can differentiate between lethal and nonlethal conditions. NGS panels for skeletal dysplasias have the potential to provide molecular diagnosis for those who would not have traditionally received a prenatal diagnosis. A gene panel designed for skeletal dysplasias detected prenatally by ultrasound includes gene families that cover the most commonly observed disorders, such as fibroblast growth factor receptor 3 (*FGFR3*), pathogenic variants in which cause thanatophoric dysplasia, achondroplasia, hypochondroplasia, and the Crouzon and Muenke craniosynostosis syndromes. *FGFR2* pathogenic variants also cause Crouzon syndrome, as well as Apert, Pfeiffer, and Jackson-Weiss syndromes, and should also be included in a skeletal dysplasia panel. The collagen gene *COL2A1* is responsible for achondrogenesis, spondyloepiphyseal dysplasia congenital, and Stickler syndrome, while *COL1A1* and *COL1A2* variants (as well as pathogenic alterations in other genes) are responsible for many types of osteogenesis imperfecta, another common skeletal dysplasia. An advantage of an NGS gene panel is that the numerous genetic causes of rare skeletal dysplasias can be simultaneously evaluated along with common causes. Given the ultrasound findings of a skeletal dysplasia, the positivity rate is higher, with over 50% detection of pathogenic or likely pathogenic variants that can explain the clinical scenario. Accordingly, genes such as *SOX9* (campomelic dysplasia), *SLC26A2* (diastrophic dysplasia, achondrogenesis), and *ALPL* (hypophosphatasia) can be tested in fetuses with abnormal skeletal findings. As with all prenatal (fetal) testing, a prenatal skeletal dysplasia panel

ideally will have a quick turnaround time and be suited for amniotic fluid or chorionic villus samples. With this consideration, careful thought should be given to the appropriate NGS platform for fetal testing, with a preference given to faster sequencing systems.

Noonan Syndrome

Noonan syndrome (NS) is an autosomal dominant condition that is caused by hyperactivation of the RAS/MAPK signaling pathway [92]. It is a relatively common disorder, with an estimated incidence as high as 1:1000 live births. NS displays both phenotypic and genetic overlaps with several related disorders that are also caused by inappropriate RAS/MAPK activity, including neurofibromatosis type 1, cardiofaciocutaneous syndrome, Costello syndrome, Noonan syndrome with multiple lentigines (NSML, formerly known as LEOPARD syndrome), and Legius syndrome. These disorders are collectively referred to as RASopathies and are characterized by short stature, congenital heart defects, facial dysmorphia, developmental delay, cryptorchidism, variable skeletal abnormalities, and variable tumor predisposition. NS displays significant variability in clinical presentation, in part due to genetic heterogeneity, as it can be caused by pathogenic variants in at least seven genes. Several additional genes associated with a Noonan-like phenotype have also been identified. This, along with the phenotypic overlap with other RASopathies, can make a definitive diagnosis challenging in some cases. Indeed, some individuals are only diagnosed after they have children who are affected more severely. In children with a RASopathy phenotype who do not display clear symptoms of one of the related syndromes, a genetic diagnosis can guide the clinical diagnosis. Prenatal studies suggest NS phenotypes can manifest in early fetal development, with features that include increased nuchal translucency, cystic hygroma, hydrops fetalis, distended jugular lymphatic sacs, and congenital heart disease [77]. These findings have led some clinicians to consider prenatal molecular testing for NS and related disorders based on ultrasound abnormalities [19, 54]. In the prenatal setting, a genetic diagnosis of a RASopathy may be the most significant finding that results in a clinical diagnosis and can guide patient management as a result. The efficacy of prenatal genetic testing for NS was examined in a study that performed prenatal testing for the most common genetic causes of NS, namely, pathogenic variants in *PTPN11*, *KRAS*, *SOS1*, and *RAF1*, in pregnancies with increased nuchal translucency and one other abnormality [19]. De novo pathogenic variants in these genes were detected in 17.3% of cases, indicating that prenatal genetic testing for NS can greatly aid in an early diagnosis of this disease. Of note, the authors only assessed the most commonly mutated genes for NS suggesting that some cases might be missed with their testing strategy. An NGS panel for NS would therefore presumably have greater diagnostic yield than sequencing of a few genes. NGS panels for NS and related disorders are currently available clinically, although there have been no published reports of their prenatal application. Some parents may elect to use a prenatal diagnosis of NS to make decisions regarding termination of pregnancy. For those who do not, a prenatal NS diagnosis has clinical value, as there is a 50–80% penetrance of congenital heart disease in NS patients [95]. These defects are often not detectable by fetal ultrasound, and a prenatal NS diagnosis would facilitate early monitoring of affected newborns. In summary, NGS panels for NS and related disorders are applicable to both fetal and postnatal diagnostics.

Newborn/Infant Applications

Hearing Loss

Hearing loss is the most common sensory impairment in humans, with an incidence of approximately 1 in 1000 newborns [69, 81]. Newborn screening programs have been very successful at identifying hearing-impaired infants, the early diagnosis of which is crucial for affected patients to receive maximal benefit from hearing aids or cochlear implants for language development. Although hearing loss can be caused by environmental factors, genetic etiology accounts for more than 60% of congenital hearing loss in developed countries [80, 96]. Around 70% of hearing loss is considered nonsyndromic hearing loss and is not accompanied by other recognizable phenotypes [109]. Conversely, 30% of genetic hearing loss is syndromic, and the diagnosis of some of these disorders would predict more severe manifestations developing with age. For instance, Usher syndrome cannot be clinically distinguished from nonsyndromic hearing loss at a young age; however, affected individuals develop progressive retinitis pigmentosa in early adolescence. Early identification of the causative variant in Usher syndrome enables life planning to cope with progressive vision loss and potentially therapeutic intervention through gene therapy, which is currently in clinical trials for Usher syndrome type 1b. Hereditary hearing loss is a genetically heterogeneous disorder. Protein expression studies found over 5000 genes, and hundreds of them uniquely represented or enriched, in mouse inner ear hair cells, suggesting potential for new deafness gene discoveries [43, 98]. Diagnostic testing for the most common genetic causes of hearing loss, pathogenic variants in the *GJB2* gene and deletions of *GJB6*, which together account for approximately 20% of nonsyndromic hearing loss, is widely available [89]. However, other genetic causes of hearing loss individually account for a small percentage of cases, and their frequency in many populations

is very variable or unknown, making sequential analysis of the remaining candidate genes impractical. Recent studies employing large panels of known causative genes reported diagnostic yield in the range of 33–39% [131, 100]. Sloan-Heggen et al. noted in their study of 1119 patients with non-syndromic hearing loss that the diagnostic yield was higher in patients with positive family history or in those with congenital or symmetric hearing loss [100]. Seco et al. used exome sequencing and copy number variant analysis of 120 hearing impairment-associated genes in Dutch patients and found causative variants in USH2A, MYO15A, and STRC as the leading causes of autosomal recessive hearing loss and causative variants in MYO6 as the leading cause of dominant hearing loss. Eight patients carried large homozygous deletions in STRC, OTOA, or USH2A genes – all confirmed by a second detection method [131]. The STRC gene is a major contributor to prelingual hearing loss, and STRC deletions have been found as one of the most frequent causes of hearing loss in several studies [83, 99, 100, 116]. However, due to over 99% sequence homology with pseudogene, other detection methods – such as long-range PCR, targeted droplet digital PCR, or array comparative genomic hybridization – are necessary to accurately detect deletions in this gene [67, 75, 116]. Therefore, it seems cost-effective to prescreen patients for the common *GJB2* and *STRC* genetic causes of hearing loss by specific detection methods prior to employing an NGS hearing loss panel while taking into account the wide ethnic variability and the pseudogenic association for some genes.

Epilepsy

Epilepsy represents a group of complex neurological disorders that are united by the presence of recurring seizures. With a prevalence of three million individuals in the USA and 3% lifetime incidence, it is among the most common neurological disorders in the developed world [41]. The age of onset is most commonly in infancy and in advanced age and can be followed by a widely variable clinical course depending on the subtype [42]. For instance, benign familial neonatal-infantile seizures have a mean age of onset of 3 months, but infants usually undergo remission by 12 months and have a very low risk of seizures later in life [115]. In contrast, Dravet syndrome patients experience seizures beginning at 6 months of age and generally do not respond to therapy [42]. Other seizure types manifest between 1 and 4 years of age, and patients suffer frequent severe seizures which slow the child's development. The causes of epilepsy are as varied as their clinical course and can be the result of a range of environmental or genetic factors. The heritable forms of epilepsy can exhibit either single gene or polygenic etiology, with pathogenic variants affecting the function of ion channels representing the best-characterized pathophysiologic mechanism [122]. In addition, numerous genetic syndromes exhibit an epileptic phenotype, including but not limited to mitochondrial disorders, neuronal migration disorders, holoprosencephaly, metabolic disorders, and storage disorders. The numerous epileptic subtypes exhibit extensive phenotypic and genetic overlap with each other and with the aforementioned syndromes, which can make a definitive diagnosis exceedingly challenging in some patients. Accordingly, a molecular diagnosis can have substantial diagnostic and prognostic value for the epilepsies, as genetics is predicted to play a role in over 50% of them [24]. Progressive myoclonic epilepsies are a prime example as a molecular finding can distinguish between several highly related disorders [97]. Finding a genetic cause of epilepsy can have a significant impact on patient care. For instance, a ketogenic diet has proven effective at reducing seizures in patients with GLUT1 deficiency syndrome [48]. For the aforementioned reasons, epilepsy represents an ideal disease candidate for the application of an NGS panel. On average about 30% of investigated patients with epilepsy received molecular diagnosis [40]. A recent study of patients with early-life epilepsies showed 29% positive rate by gene panel testing and 28% by exome sequencing [10]. The list of genes involved in epilepsy is growing fast. Wang J et al. (2017) in their review of epilepsy genes listed 84 epilepsy genes, 73 neurodevelopment and epilepsy-associated genes, and 536 epilepsy-related genes [118]. Yet, more genes still await their official association with epilepsy. As an example of new gene-phenotype association findings, our laboratory recently identified a de novo variant in a gene not currently associated with an epileptic disorder. However, a number of other potentially causative variants in the same gene are necessary to support causality [42]. Due to the phenotypic and genotypic variability of epileptic disorders, targeted exome sequencing has become a viable test choice in patients with epilepsy. In one such study of 63 trios, Wang Y et al. (2017) reported pathogenic variants in 24% of patients and also identified possibly causative variants in several new candidate epilepsy genes [119]. Another exome sequencing study on 40 patients with focal epilepsy targeted 64 epilepsy genes and identified pathogenic variants in 12.5% of them [82].

One of the challenges of exome sequencing is potential loss of information due to less-than-optimal coverage for parts of the exome. On the other hand, targeted gene panel NGS technology with high read depth has proven very instrumental, especially in detecting low-level mosaicism in a number of epilepsy-associated genes. Stosser et al. identified mosaic pathogenic variants in nine genes associated with epilepsy, accounting overall for 3.5% detection frequency in their study [104]. Thus, patients or parents of the probands with previous negative results by Sanger sequencing

in whom the genetic etiology is still suspected may benefit from targeted gene panel NGS testing due to the possibility of detecting pathogenic variants in a mosaic form, with the caveat that alternative tissue samples, other than blood, may need to be analyzed.

Mitochondrial Disorders

Mitochondrial dysfunction underlies a group of disorders that have a reported incidence of 1 in 5000 live births [26]. Mitochondria-related disorders have a broad range of clinical presentation, display locus, and allelic heterogeneity and can be autosomal recessive, dominant, sex-linked, and sporadic, making them challenging to diagnose and thereby obvious candidates for NGS panels [21]. Further complicating matters, mitochondria are uniquely composed of proteins and RNA encoded by the nuclear and mitochondrial genomes, pathogenic variants in both of which can lead to mitochondrial dysfunction and disease. The mitochondrial genome is 16.6 kb and contains 37 genes that encode mitochondrial enzymes and transfer and ribosomal RNAs. The nuclear genome contributes over 1000 genes to mitochondrial function – pathogenic variants in at least 100 of which are associated with human disease [49, 79]. Although there is clear clinical utility for an NGS mitochondrial disorder panel, sequencing these disease genes poses some technical challenges. Mitochondrial genes have numerous nuclear-encoded pseudogenes, which can be challenging to distinguish from the targeted coding gene given the short reads currently employed by NGS platforms. Additionally, a cell or population of cells can have more than one mitochondrial genome, a phenomenon known as heteroplasmy. A recent NGS-based study has shown that the incidence of heteroplasmy varies between 10% and 50%, which is significantly higher than previously appreciated [102]. Disease-causing mitochondrial variants are frequently heteroplasmic, and the proportion of the pathogenic variant allele directly impacts disease manifestation and severity. While NGS is uniquely suited to detect heteroplasmy, NGS panels need deep coverage of the mitochondrial genome to detect sensitively low-level heteroplasmic variants. Moreover, heteroplasmy can complicate data analysis, interpretation, and reporting, as there is currently no consensus as to what threshold of mutation burden is clinically significant. This is confounded by the fact that the cell population analyzed, usually peripheral blood, may not accurately represent the mitochondrial DNA (mtDNA) profile of the diseased tissue. Moreover, Sanger sequencing is incapable of confirming low-level heteroplasmic variants, and more sensitive methods such as allele-specific PCR or pyrosequencing are needed to confirm true positives. For rare pathogenic variants, the need to confirm low-level heteroplasmic variants presents a significant burden for clinical laboratories. In light of these challenges, laboratories should consider validating samples from other tissues when developing NGS-based panels for mitochondrial disorders. Additionally, developing and validating mitochondrial NGS panels on multiple NGS platforms would provide a robust method for confirmation of heteroplasmic variants. Several computational methods have been developed to detect and quantify mtDNA deletions from next-generation sequencing experiments. A recent study described a software that detects mtDNA deletions below 1% heteroplasmy levels with a low false-positive rate [12].

Multiple groups have reported the effective development and application of NGS-based mitochondrial gene panels [15, 21, 112]. They have employed either a hybrid capture or a combination of long-range PCR for selective enrichment of the mitochondrial genome and emulsion PCR for nuclear genes. The first proof-of-concept study applied an NGS panel containing the mitochondrial genome and 1000 nuclear-encoded genes to 42 unrelated infants with clinical and biochemical evidence of mitochondrial disease [15]. Ten patients were found to have clear disease-causing variants in known genes, while 13 patients had pathogenic variants in nuclear-encoded mitochondrial genes that had not been previously linked to disease. A second group reported a validation study of a clinical NGS panel, where 13 clinical samples and a group of 16 validation samples with known mitochondrial and nuclear variants were sequenced [21]. All known variants in the validation samples were observed, while five clinical samples were found to have disease-causing variants. These studies collectively demonstrate that NGS panels are effective for the molecular diagnosis of mitochondrial disorders and should prove to be a valuable tool for clinicians. At least two studies utilizing exome sequencing to evaluate both the mitochondrial and nuclear genomes suggested that mitochondrial disorders can be effectively diagnosed [25, 87]. As mitochondrial disorders have significant phenotypic overlap with numerous other disorders – including epilepsy, hearing loss, and retinitis pigmentosa – evaluation of the mitochondrial genome with CES is an important consideration.

Rapid NICU NGS Testing Panel

Genetic diseases constitute a major cause of infant mortality [70]. In neonatal intensive care of critically ill infants, the application of rapid, comprehensive next-generation sequencing to identify the molecular causes of disease has become increasingly valuable for clinical management and family planning. Conventional genetic testing approaches are time-consuming and often fail to provide diagnosis. Clinicians are left to provide only supportive therapies,

possibly missing treatments targeting the specific molecular cause, and families have to make difficult decisions with incomplete information. Many genetic disorders are phenotypically not fully expressed early in life; clinical signs are subtle, highly unspecific, or masked, especially in the setting of critical conditions. In a proof-of-concept pilot study, infants without a diagnosis presenting with respiratory and/or cardiovascular failure, encephalopathy, profound hypotonia, severe metabolic disturbance, multiple major congenital anomalies, or multisystem organ failure likely due to genetic causes were tested together with their parents by a 4500 known disease-causing gene panel [14]. Of the 12 tested infants, causative variants were identified in 8 (67%), and preliminary and final results were delivered after a mean of 9 and 16 days, respectively. In none of the eight positive cases was the diagnosis strongly suspected prior to testing with the rapid gene panel [14]. Another targeted analysis of 3426 known disease genes performed by using proband-only whole-genome sequencing data with copy number variant detection provided a genetic diagnosis for seven critically ill infants (30%) that had no clear a priori diagnosis, with a median turnaround time of 12 days. Based on the diagnosis, unsuccessful intensive care treatment was withdrawn in five of the seven children diagnosed. The results also led to more informed decisions for the parents as they learned about possibilities of prenatal testing [110]. Recently, a cohort trio exome sequencing resulted in a molecular diagnosis in 32 of 63 infants (50.8%) with a mean turnaround time of 13 days. The indications for rapid exome testing included neuromuscular diseases; syndromic congenital cardiovascular malformations; hypertrophic cardiomyopathy with an assessment for cardiac transplant, skeletal malformations, and/or dysplasia; neonatal cholestasis and liver failure; lung disease including alveolar capillary dysplasia; cystic renal disease; and metabolic disorders with persistent lactic acidosis. Clinical care was altered by the diagnosis in 23 of 32 patients (72%) [72]. Overall, the study reported molecular diagnosis in 102 infants (36.7%) by clinical exome sequencing and affected medical management for 53 infants (52.0%). About 38% of diagnosed patients had an atypical or unrecognized infantile presentation of genetic disorders [72]. The benefits of utilizing exome sequencing at an early point of the diagnostic process in children with genetically heterogeneous or overlapping conditions have been also underscored in an Australian study that assessed 61 children with a mean age of 28 months. A diagnosis was achieved in 23 (52%) singleton patients and was unexpected in 8 of those 23 (35%), and clinical management was altered in 6 of the 23 (26%). The study also provided compelling evidence for cost-effectiveness of exome sequencing at an early stage in the diagnostic trajectory [106].

The Future of Genomic Newborn Screening

With the advancement of genomic technology and its utility to effectively identify genetic causes of rare pediatric disorders, a theoretical possibility emerges to use this technology in newborn screening. However, before it can be implemented, clinical utility; cost-effectiveness; improvements in the interpretation of genomic data (including predictive, diagnostic, therapeutic, and prognostic value of genomic information); policies on disclosures of variants of uncertain significance; secondary and other incidental findings; informed consent and parental education; ownership, storage, and sharing of the data; public health policies; ethical and psychosocial parental concerns; and infrastructure must be addressed. A number of working groups have been created to research and drive the complex projects addressing key questions of the potential and challenges of genomic newborn screening [11]. The Pediatric Task Team of Global Alliance for Genomics and Health's Regulatory and Ethics Working Group has developed eight recommendations, with the overarching guiding principle that the best interest of children should be the basis for the decisions about implementing genomic newborn screening [36].

Adult or Young Adult-Onset Applications

Aortopathy

Aneurysm and dissection of the aorta is a common cause of mortality, accounting for 1–2% of deaths every year [58]. Diseases of the aorta are collectively known as the aortopathies and can involve dilation, aneurysm, or malformation of the aorta. These aortic disorders result in dissection of the aorta, most commonly the thoracic aorta, and thus these disorders have also been termed thoracic aortic aneurysms and dissections (TAA/TAAD). Abdominal aortic aneurysms (AAA) are likely caused by interactions of multiple predisposing genes and environmental risk factors [38]. There is a strong genetic contribution to thoracic aortic aneurysms, with reportedly about 20% of patients having an affected first-degree relative [114]. One of the most common inherited aortopathies is Marfan syndrome, a connective tissue disorder with a prevalence of approximately 1 in 5000 [90]. Affected individuals display a pleiotropic phenotype that primarily involves the skeletal and cardiovascular systems, with cardiovascular phenotypes being the primary cause of mortality. Pathogenic variants in the fibrillin gene (*FBN1*) cause 90–95% of Marfan cases, with over 3000 different pathogenic variants in the large, 65-exon *FBN1* gene described [128]. Out of these, large single- or multi-exonic deletions in *FBN1* are a rare cause of Marfan syndrome as only 37 have so far been described [56]. *FBN1* pathogenic variants

have also been identified in patients with isolated ectopia lentis [78], with bicuspid aortic valve [37], as well as in isolated nonsyndromic aortic aneurysm and dissections [105]. Loeys-Dietz syndrome (LDS) displays a wide spectrum of multisystem involvement [108] and patients with LDS experience progressive aortic aneurysmal disease with aneurysms that grow faster than 1.0 cm/year [38]. Mean age of death at 26 years in patients with LDS was reported in one study [64]. The typical triad of clinical symptoms in patients with LDS includes hypertelorism, cleft palate or bifid uvula, and arterial tortuosity [111]. Importantly, minimal diagnostic criteria have not been established for LDS, and a definitive diagnosis often relies mainly on molecular testing due to the significant degree of clinical overlap with other syndromes such as Beals and Marfan syndrome [63, 126]. Numerous other disorders that involve aneurysm of the aorta have been described and include congenital contractural arachnodactyly/Beals syndrome, Shprintzen-Goldberg syndrome, Ehlers-Danlos syndrome, cutis laxa, arterial tortuosity syndrome, and familial thoracic aortic aneurysms and dissections. These disorders are all caused by pathogenic variants in at least 29 genes involved in TGFβ signaling or cytoskeletal organization and, as a result, can be challenging to differentiate from one another [13]. Early clinical and molecular diagnosis of these disorders is essential, as there is a trend toward gene-tailored management strategies [13]. For instance, surgical intervention is recommended for LDS patients with pathogenic variants in *TGFBR1*, *TGFBR2*, or *SMAD3* when their ascending aorta reaches a diameter of 40–45 mm. Thoracic aortic aneurysms are thus typically clinically silent but can be fatal due to dissection or rupture, whereas recommendations for patients with LDS type 5 with *TGFB3* pathogenic variants are to undergo prophylactic surgery when their ascending aorta reaches a diameter of 50–55 mm [13]. Considering the overlapping phenotypic and genetic spectrum of the aortopathies, this group of disorders is an excellent candidate for targeted NGS, and the resultant molecular diagnoses have high clinical utility.

Several studies have reported results of NGS testing in patents with thoracic aortic disorders. Campens et al. [16] sequenced 7 genes in 264 patients and identified pathogenic variants in 13% of the cases. Poninska et al. [84] analyzed 10 genes in 51 patients and found causative variants in 18% of the cases. By testing 10 genes in 175 patients by NGS and comparative genomic hybridization, Wooderchak-Donahue et al. [125] identified pathogenic variants in 10% of the samples. Similarly, Fang et al. [32] tested 11 genes in 70 patients by NGS and target capture array and identified pathogenic variants in 19% of them. Proost et al. [88] found causal variants in 27% of the tested individuals analyzing 14 genes by NGS and MLPA (multiple ligation probe amplification) for large deletions and duplications in 55 patients. Ziganshin et al. [132] performed exome sequencing on 102 patients and analyzed targeted 21 genes and found pathogenic variants in 4% of the patients. The number of genes tested continues to increase, as additional disorders or new genetic causes of disease are included.

Cardiomyopathy

Inherited cardiac disorders are a relatively common group of diseases that collectively affect about 1 in 390 people [91]. In contrast to age-related cardiac disease, inherited cardiomyopathies occur much earlier in life, ranging from adolescence to early adulthood. Hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and arrhythmogenic right ventricular cardiomyopathy (ARVC) are the most common forms of inherited cardiomyopathy and, together, represent a major cause of heart disease in all age groups [121]. These disorders predispose to sudden cardiac death, but also may progress with age, leading to heart failure. Accordingly, individuals with inherited cardiomyopathies represent critical cases for early diagnosis so that appropriate clinical management can be employed to reduce morbidity and mortality associated with these disorders. Although cardiomyopathies are diagnosed based on clinical presentation, a genetic diagnosis of the proband greatly facilitates the genetic counseling and risk assessment for family members [123]. This is of particular importance for this group of disorders, which can go undiagnosed for years.

Inherited cardiomyopathies represent an excellent application of NGS panels, as they exhibit locus heterogeneity and both genetic and phenotypic overlap [107]. For instance, HCM is caused by pathogenic variants in approximately 8 genes, although over 20 genes have published clinical associations with HCM. Similarly, DCM is caused by pathogenic variants in approximately 10 genes, though around 40 genes have published association with DCM. Lastly, ARVC is less genetically heterogeneous with six genes clearly causing disease and an additional four disease-associated genes. In total, these disorders have been associated with over 50 genes, which certainly preclude comprehensive Sanger sequencing.

An NGS cardiomyopathy panel can also facilitate a clinical diagnosis in some cases, for instance, later-stage HCM is morphologically similar to dilated cardiomyopathy. Cardiomyopathies, moreover, are a feature of numerous other syndromic conditions, including metabolic disorders, Noonan syndrome, and assorted myopathies. In these cases, a genetic diagnosis can greatly facilitate differentiation between these disorders. In a recent study, a 41-gene NGS cardiomyopathy panel was evaluated for 223 unrelated patients presenting with HCM [65]. Published disease-causing variants were found in 33.6% of patients, while an additional 23.8% of patients had novel variants that were predicted to be pathogenic. For the cardiomyopathies,

TTN represents a challenging gene for variant analysis, as it encodes the largest known human protein with 34,350 amino acids. Protein-truncating variants located in highly conserved exons have been observed in about 1 in 5 patients with dilated cardiomyopathy but also in about 1 in 200 individuals in general population [34]. Thus, current understanding is that truncating TTN variants identified in patients with clinical diagnosis of dilated cardiomyopathy, located in the constitutively expressed exons, are likely major contributors to the pathogenesis of the disease; however, the pathogenicity of such truncating variants in asymptomatic carriers is largely unknown and will require long-term follow-up of these unaffected individuals [34].

Clinical Exome and Genome

Clinical exome and genome tests are unique among sequencing-based tests in that they can be applied to the diagnosis of essentially any genetic disorder. As with most new technologies, it was first applied on a research basis with great success in the realm of gene discovery. As the majority of patients subjected to CES/CGS have been clinically evaluated and often tested for common and/or classic genetic disorders, the diagnostic yield of CES/CGS was initially unclear. In the absence of large-scale studies, the first initial case reports focused on successful applications of CES/CGS. A notable success story of CES was the report of a young male child with intractable inflammatory bowel disease [127]. The patient suffered from severe stunting and malnutrition, bacterial sepsis, and other severe complications related to severe gastrointestinal inflammation. An ES study revealed a missense variant in the *X-Linked Inhibitor of Apoptosis* gene (*XIAP*), which had not previously been associated with bowel disease but had been implicated in the proinflammatory response and bacterial sensing through the nucleotide-binding oligomerization domain (NOD) signaling pathway. Functional studies revealed defects in the NOD signaling pathway, and subsequently an allogeneic hematopoietic progenitor cell transplant was performed on the patient based on these findings. Remarkably, the patient was reported to be disease-free 42 days posttransplant. This study represents a unique case where gene discovery and novel therapy were made possible by the application of ES. A second study examined fraternal twins who suffered from dopa (3,4-dihydroxyphenylalanine)-responsive dystonia (DRD) [5]. They had no identified deleterious variants in the two genes known to cause DRD for which clinical tests were available but were not tested for two other known causative genes for which clinical tests were unavailable. CGS identified compound heterozygous variants in the *SPR* gene, a gene previously implicated in DRD. This genetic diagnosis led to the supplementation of their current therapy with a second agent that resulted in clinical improvements for both twins. This study represents a more likely positive outcome in CES/CGS studies, where a pathogenic alteration is found in a known gene for which no clinical tests are available.

Widespread clinical adoption of CES/CGS has resulted in studies that consistently report diagnostic yields in the range of 25–31% in individuals with diverse clinical indications [53, 93, 129, 130]. Based on these large studies, CES/CGS is more effective for certain phenotypes than others. The primary clinical indications of disorders of hearing, vision, skeletal muscle, skeletal system, and multiple congenital anomalies find molecular diagnoses at an above-average rate, while nonsyndromic autism and malignancy have much lower diagnostic rates [93]. The reasons for these discrepancies are complex but likely involve a mixture of several factors that include variable understanding of the genetic etiologies of studied phenotypes, increased prevalence of environmental or other non-Mendelian causes of certain phenotypes, and unappreciated reduced penetrance and variable expressivity that may confound CES/CGS interpretation. Regardless of the underlying etiology, further study will hopefully provide detailed empirical evidence to guide test selection for those with suspected Mendelian disease. In addition to these confirmed molecular diagnoses, clinical laboratories report the so-called candidate genes at rates typically ranging from 6% to 9% but as high as 24% [33, 93]. Candidate genes are often defined as genes that are not currently known to be associated with a Mendelian phenotype, but for which there is compelling biological, signaling pathway, animal model, and genomic data that support proposing a candidate gene as a potential new genetic etiology of a Mendelian phenotype. These findings are often corroborated in peer-reviewed publications, with a clinical laboratory reporting 38–52% of candidates described in the scientific literature within a year of the original reporting [33]. Clearly, candidate genes significantly increase the diagnostic yield of clinical exome sequence, but they pose an interpretive challenge for laboratories and clinicians. Data-sharing efforts pioneered by GeneMatcher, an online tool for clinical laboratories, clinicians, and researchers to share information on rare disease genes and phenotypes, has accelerated new gene discovery and helped transform candidate genes reported on clinical tests into bona fide Mendelian disease genes [101]. Continued and enhanced collaboration between clinical laboratories, clinicians, patient families, and researchers will be essential to improve the overall yield of genomic test and help bring answers to more families in need.

Widespread clinical adoption of CES/CGS has also revealed multiple Mendelian diagnoses in 3–5% of cases with a molecular diagnosis [93, 129]. The resulting blended phenotypes of these patients explain why they had previously gone without a molecular diagnosis through traditional means. Moreover, they represent the high clinical utility of

this testing approach that could identify a molecular etiology where no other approach could. While CES/CGS has been most readily applied in pediatrics, there is clearly a utility for adult patients with a suspected genetic disorder. The largest clinical-grade cohort reported a diagnostic rate of 17.5% in adults, though most of the diagnostic results were due to pediatric onset disease [85]. As clinical adoption for CES/CGS continues to expand for adult-onset conditions, it will be interesting to observe the diagnostic value of this testing methodology in this population. In conclusion, CES/CGS has made a rapid transition from a research assay to clinical diagnostic test. The impressive diagnostic yields across broad clinical indications clearly illustrate the clinical utility of this testing approach. Advances in our understanding of the genetic etiology of Mendelian disease and normal genomic variation, fueled in part from diagnostic testing itself, will undoubtedly lead to higher rates of diagnoses and less uncertainty for patients. Challenges remain, but the future is bright for this testing modality, which will continue to experience rapid growth in the coming years.

Conclusions

The clinical application of NGS tests has already led to an improved understanding of Mendelian disease, both in discovering new disease-causing genes and for novel therapeutic strategies. Both rare and common inherited disorders have benefitted from the increased capabilities NGS has provided to clinical laboratories and the striking success stories reported in the literature have fanned the flames of interest in the medical community. NGS brings the promise of improved healthcare for people in all stages of life. From couples seeking preconception genetic screening to noninvasive prenatal diagnosis to disorders suffered by young and old, we may soon realize a time when every person is able to have a whole-genome test. It is clear that this technology will enjoy explosive growth over the coming decade.

As clinical NGS matures, technical improvements in both sequencing capabilities and bioinformatic analysis will impact the future of this testing methodology. The clinical field will likely move toward a single NGS assay that combines copy number analysis and sequence variant detection. As a CES or CGS test, this assay would integrate all the relevant genetic information in a single test. As such, it would reduce turnaround times and could be consolidated into a single report that would be easier for clinical laboratories and clinicians alike. Longer NGS read lengths and/or genome partitioning methods will lead to longer distance haplotyping, a capability that would facilitate NGS testing of cffDNA. As NGS sequencing improves in quality and declines in cost, CES/CGS may someday be performed at high depth without sequencing gaps. In this scenario, NGS panels may become obsolete, and the field may move toward CES/CGS as a singular testing methodology that is analyzed in a targeted fashion based on clinical phenotype. Indeed, some clinical laboratories already offer virtual diagnostic panels derived from an exome sequence. As the clinical adoption of CGS increases, new tools for the interpretation of the immense numbers of noncoding variants will be required. These will likely include computational tools but will also require functional evidence to facilitate prioritization of noncoding variants. This may come in the form of RNA-seq, which early studies have demonstrated can significantly enhance diagnostic yield in undiagnosed muscular disease [20]. Streamlined integration of these ever-expanding datasets into clinical analysis and interpretation will be key to harnessing the power of the next wave of genomic diagnostics.

As discussed previously, variant classification is a challenging, time-consuming endeavor. Clinical laboratories are combining efforts and sharing variant information in centralized databases such as ClinVar [51]. Collating variant interpretations has already provided significant resources for both clinical and research laboratories; however, improvements could be made to enhance the clinical utility of this resource. Variants would ideally be linked to phenotypic data so that this large data set can be queried to assess rare variants in an informed context. Such an effort, if widely adopted, will be a tremendous tool for both researchers working toward gene discovery and clinical laboratories evaluating patient NGS tests. While recent efforts to improve the depth and breadth of human genomics databases has dramatically improved the clinical laboratory's ability to interpret sequence variants, some ethnicities are still not well represented in the large-scale genome sequencing efforts such as the 1000 Genomes, Exome Aggregation Consortium, and Genome Aggregation Consortium databases [55]. As a result, CES tests of these patients yield many rare, novel variants that reflect their ancestry but are unrelated to their clinical presentation. Expansion of these large-scale sequencing projects to include these underrepresented ethnicities will greatly facilitate the evaluation of rare variants and also enhance the quality of clinical reports generated from CES/CGS testing.

References

1. ACMG Board of Directors. Points to consider in the clinical application of genomic sequencing. Genet Med. 2012;14:759–61.
2. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hoischen A, Vissers LE, Gilissen C. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. Am J Hum Genet. 2015;97(1):67–74.
3. Arjunan A, Litwack K, Collins N, Charrow J. Carrier screening in the era of expanding genetic technology. Genet Med. 2016;18(12):1214–7.
4. Azimi M, Schmaus K, Gerger V, Neitzel D, Rochelle R, Dinh T. Carrier screening by next-generation sequencing: health benefits and cost effectiveness. Mol Genet Genomic Med. 2016;4(3):292–302.

5. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, Reid JG, Fink JK, Morgan MB, Gingras MC, Muzny DM, Hoang LD, Yousaf S, Lupski JR, Gibbs RA. Whole-genome sequencing for optimized patient management. Sci Transl Med. 2011;3:87re3.
6. Baylor Genetics Laboratories. PreSeek non-invasive prenatal gene sequencing screen. https://www.bcm.edu/research/medical-genetics-labs/test_detail.cfm?testcode=21200. Accessed 25 Jan 2018.
7. Beicht S, Strobl-Wildemann G, Rath S, Wachter O, Alberer M, Kaminsky E, Weber LT, Hinrichsen T, Klein HG, Hoefele J. Next generation sequencing as a useful tool in the diagnostics of mosaicism in Alport syndrome. Gene. 2013;526:474–7.
8. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473–8.
9. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW, Kingsmore SF. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med. 2011;3:65ra4.
10. Berg AT, Coryell J, Saneto RP, Grinspan ZM, Alexander JJ, Kekis M, Sullivan JE, Wirrell EC, Shellhaas RA, Mytinger JR, Gaillard WD, Kossoff EH, Valencia I, Knupp KG, Wusthoff C, Keator C, Dobyns WB, Ryan N, Loddenkemper T, Chu CJ, Novotny EJ Jr, Koh S. Early-life epilepsies and the emerging role of genetic testing. JAMA Pediatr. 2017;171(9):863–71.
11. Berg JS, Agrawal PB, Bailey DB Jr, Beggs AH, Brenner SE, Brower AM, Cakici JA, Ceyhan-Birsoy O, Chan K, Chen F, Currier RJ, Dukhovny D, Green RC, Harris-Wai J, Holm IA, Iglesias B, Joseph G, Kingsmore SF, Koenig BA, Kwok PY, Lantos J, Leeder SJ, Lewis MA, AL MG, Milko LV, Mooney SD, Parad RB, Pereira S, Petrikin J, Powell BC, Powell CM, Puck JM, Rehm HL, Risch N, Roche M, Shieh JT, Veeraraghavan N, Watson MS, Willig L, Yu TW, Urv T, Wise AL. Newborn sequencing in genomic medicine and public health. Pediatrics. 2017;139(2):1–13.
12. Bosworth CM, Grandhi S, Gould MP, LaFramboise T. Detection and quantification of mitochondrial DNA deletions from next-generation sequence data. BMC Bioinformatics. 2017;18(Suppl 12):407.
13. Brownstein AJ, Ziganshin BA, Kuivaniemi H, Body SC, Bale AE, Elefteriades JA. Genes associated with thoracic aortic aneurysm and dissection: an update and clinical implications. Aorta (Stamford). 2017;5(1):11–20.
14. Brunelli L, Mao R, Jenkins SM, Bleyl SB, Dames SA, Miller CE, Ostrander B, Tvrdik T, Andrews S, Flores J, Patel S, Gudgeon JM, Schaefer S. A rapid gene sequencing panel strategy to facilitate precision neonatal medicine. Am J Med Genet A. 2017;173(7):1979–82.
15. Calvo SE, Compton AG, Hershman SG, Lim SC, Lieber DS, Tucker EJ, Laskowski A, Garone C, Liu S, Jaffe DB, Christodoulou J, Fletcher JM, Bruno DL, Goldblatt J, Dimauro S, Thorburn DR, Mootha VK. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Sci Transl Med. 2012;4:118ra10.
16. Campens L, Callewaert B, Muiño Mosquera L, Renard M, Symoens S, De Paepe A, Coucke P, De Backer J. Gene panel sequencing in heritable thoracic aortic disorders and related entities – results of comprehensive testing in a cohort of 264 patients. Orphanet J Rare Dis. 2015;10:9.
17. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, Foo CH, Xie B, Tsui NB, Lun FM, Zee BC, Lau TK, Cantor CR, Lo YM. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A. 2008;105:20458–63.
18. Costa T, Scriver CR, Childs B. The effect of Mendelian disease on human health: a measurement. Am J Med Genet. 1985;21:231–42.
19. Croonen EA, Nillesen WM, Stuurman KE, Oudesluijs G, Van De Laar IM, Martens L, Ockeloen C, Mathijssen IB, Schepens M, Ruiterkamp-Versteeg M, Scheffer H, Faas BH, Van Der Burgt I, Yntema HG. Prenatal diagnostic testing of the Noonan syndrome genes in fetuses with abnormal ultrasound findings. Eur J Hum Genet. 2013;21:936–42.
20. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O'Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A, Kang PB, Beggs AH, North KN, Straub V, Dowling JJ, Muntoni F, Clarke NF, Cooper ST, Bönnemann CG, DG MA. Improving genetic diagnosis in mendelian disease with transcriptome sequencing. Sci Transl Med. 2017;9(386):1–11.
21. Dames S, Chou LS, Xiao Y, Wayman T, Stocks J, Singleton M, Eilbeck K, Mao R. The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. J Mol Diagn. 2013;15:526–34.
22. Daniels G, Finning K, Martin P, Massey E. Noninvasive prenatal diagnosis of fetal blood group phenotypes: current practice and future prospects. Prenat Diagn. 2009;29:101–7.
23. Del Castillo I, Moreno-Pelayo MA, Del Castillo FJ, Brownstein Z, Marlin S, Adina Q, Cockburn DJ, Pandya A, Siemering KR, Chamberlin GP, Ballana E, Wuyts W, Maciel-Guerra AT, Alvarez A, Villamar M, Shohat M, Abeliovich D, Dahl HH, Estivill X, Gasparini P, Hutchin T, Nance WE, Sartorato EL, Smith RJ, Van Camp G, Avraham KB, Petit C, Moreno F. Prevalence and evolutionary origins of the del(GJB6-D13S1830) mutation in the DFNB1 locus in hearing-impaired subjects: a multicenter study. Am J Hum Genet. 2003;73:1452–8.
24. Dhiman V. Molecular genetics of epilepsy: a clinician's perspective. Ann Indian Acad Neurol. 2017;20(2):96–102.
25. Dinwiddie DL, Smith LD, Miller NA, Atherton AM, Farrow EG, Strenk ME, Soden SE, Saunders CJ, Kingsmore SF. Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. Genomics. 2013;102:148–56.
26. Distelmaier F, Koopman WJ, Van Den Heuvel LP, Rodenburg RJ, Mayatepek E, Willems PH, Smeitink JA. Mitochondrial complex I deficiency: from organelle dysfunction to clinical disease. Brain. 2009;132:833–42.
27. Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, Panda B, Sergouniotis PI, Gillespie RL, Daiger SP, Hall G, Gale T, Lloyd IC, Bishop PN, Ramsden SC, Black GC. Molecular findings from 537 individuals with inherited retinal disease. J Med Genet. 2016; https://doi.org/10.1136/jmedgenet-2016-103837. pii: jmedgenet-2016-103837. [Epub ahead of print].
28. Ellingford JM, Barton S, Bhaskar S, Williams SG, Sergouniotis PI, O'Sullivan J, Lamb JA, Perveen R, Hall G, Newman WG, Bishop PN, Roberts SA, Leach R, Tearle R, Bayliss S, Ramsden SC, Nemeth AH, Black GC. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. Ophthalmology. 2016;123(5):1143–50.
29. Ellingford JM, Campbell C, Barton S, Bhaskar S, Gupta S, Taylor RL, Sergouniotis PI, Horn B, Lamb AJ, Michaelides M, Webster AR, Newman WG, Panda B, Ramsden SC, Black GC. Validation of copy number variation analysis for next-generation sequencing. Eur J Hum Genet. 2017;25(6):719–24.
30. Fahim AT, Daiger SP, Weleber RG. Nonsyndromic retinitis pigmentosa overview. Gene Reviews [Internet]. 2000. [Updated 2017 Jan 19].

31. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A. 2008;105:16266–71.
32. Fang M, Yu C, Chen S, Xiong W, Li X, Zeng R, Zhuang J, Fan R. Identification of novel clinically relevant variants in 70 southern Chinese patients with thoracic aortic aneurysm and dissection by next-generation sequencing. Sci Rep. 2017;7(1):10035.
33. Farwell Hagman KD, Shinde DN, Mroske C, Smith E, Radtke K, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Alcaraz WA, Helbig KL, Sajan SA, Rossi M, Lu HM, Huether R, Li S, Wu S, Nuñes ME, Tang S. Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. Genet Med. 2017;19(2):224–35.
34. Fatkin D, Huttner IG. Titin-truncating mutations in dilated cardiomyopathy: the long and short of it. Curr Opin Cardiol. 2017; https://doi.org/10.1097/HCO.0000000000000382. [Epub ahead of print].
35. Freed D, Pevsner J. The contribution of mosaic variants to autism spectrum disorder. PLoS Genet. 2016;12(9):e1006245.
36. Friedman JM, Cornel MC, Goldenberg AJ, Lister KJ, Sénécal K, Vears DF. Global Alliance for Genomics and Health Regulatory and Ethics Working Group Paediatric Task Team. Genomic newborn screening: public health policy considerations and recommendations. BMC Med Genomics. 2017;10(1):9.
37. Giusti B, Sticchi E, De Cario R, Magi A, Nistri S, Pepe G. Genetic bases of bicuspid aortic valve: the contribution of traditional and high-throughput sequencing approaches on research and diagnosis. Front Physiol. 2017;8:612.
38. Goyal A, Keramati AR, Czarny MJ, Resar JR, Mani A. The genetics of aortopathies in clinical cardiology. Clin Med Insights Cardiol. 2017;11:1179546817709787.
39. Grody WW, Thompson BH, Gregg AR, Bean LH, Monaghan KG, Schneider A, Lebo RV. ACMG position statement on prenatal/preconception expanded carrier screening. Genet Med. 2013;15:482–3.
40. Hardies K, Weckhuysen S, De Jonghe P, Suls A. Lessons learned from gene identification studies in mendelian epilepsy disorders. Eur J Hum Genet. 2016;24(7):961–7.
41. Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935–1984. Epilepsia. 1993;34:453–68.
42. Helbig I, Scheffer IE, Mulley JC, Berkovic SF. Navigating the channels and beyond: unravelling the genetics of the epilepsies. Lancet Neurol. 2008;7:231–45.
43. Hickox AE, Wong AC, Pak K, Strojny C, Ramirez M, Yates JR 3rd, Ryan AF, Savas JN. Global analysis of protein expression of inner ear hair cells. J Neurosci. 2017;37(5):1320–39.
44. Illanes S, Denbow M, Kailasam C, Finning K, Soothill PW. Early detection of cell-free fetal DNA in maternal plasma. Early Hum Dev. 2007;83:563–6.
45. Kaback MM. Population-based genetic screening for reproductive counseling: the Tay-Sachs disease model. Eur J Pediatr. 2000;159(Suppl 3):S192–5.
46. Kauffman TL, Wilfond BS, Jarvik GP, Leo MC, Lynch FL, Reiss JA, Richards CS, McMullen C, Nickerson D, Dorschner MO, Goddard KA. Design of a randomized controlled trial for genomic carrier screening in healthy patients seeking preconception genetic testing. Contemp Clin Trials. 2017;56:100–5.
47. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, Gammill HS, Rubens CE, Santillan DA, Murray JC, Tabor HK, Bamshad MJ, Eichler EE, Shendure J. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med. 2012;4:137.
48. Klepper J, Scheffer H, Leiendecker B, Gertsen E, Binder S, Leferink M, Hertzberg C, Nake A, Voit T, Willemsen MA. Seizure control and acceptance of the ketogenic diet in GLUT1 deficiency syndrome: a 2- to 5-year follow-up of 15 children enrolled prospectively. Neuropediatrics. 2005;36:302–8.
49. Koene S, Smeitink J. Mitochondrial medicine: entering the era of treatment. J Intern Med. 2009;265:193–209.
50. Kronn D, Jansen V, Ostrer H. Carrier screening for cystic fibrosis, Gaucher disease, and Tay-Sachs disease in the ashkenazi jewish population: the first 1000 cases at New York University Medical Center, New York. NY Arch Intern Med. 1998;158:777–81.
51. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980–5.
52. Lazarin GA, Haque IS, Nazareth S, Iori K, Patterson AS, Jacobson JL, Marshall JR, Seltzer WK, Patrizio P, Evans EA, Srinivasan BS. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. Genet Med. 2012;15:178–86.
53. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical exome sequencing for genetic identification of rare mendelian disorders. JAMA. 2014;312(18):1880–7.
54. Lee KA, Williams B, Roza K, Ferguson H, David K, Eddleman K, Stone J, Edelmann L, Richard G, Gelb BD, Kornreich R. PTPN11 analysis for the prenatal diagnosis of Noonan syndrome in fetuses with abnormal ultrasound findings. Clin Genet. 2009;75:190–4.
55. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, De Pristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91.
56. Li J, Wu W, Lu C, Liu Y, Wang R, Si N, Liu F, Zhou J, Zhang S, Zhang X. Gross deletions in FBN1 results in variable phenotypes of Marfan syndrome. Clin Chim Acta. 2017;474:54–9.
57. Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, Turner J, Cannons JL, Bick D, Blakemore L, Blumhorst C, Brockmann K, Calder P, Cherman N, Deardorff MA, Everman DB, Golas G, Greenstein RM, Kato BM, Keppler-Noreuil KM, Kuznetsov SA, Miyamoto RT, Newman K, Ng D, O'brien K, Rothenberg S, Schwartzentruber DJ, Singhal V, Tirabosco R, Upton J, Wientroub S, Zackai EH, Hoag K, Whitewood-Neal T, Robey PG, Schwartzberg PL, Darling TN, Tosi LL, Mullikin JC, Biesecker LG. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med. 2011;365:611–9.
58. Lindsay ME, Dietz HC. Lessons on the pathogenesis of aneurysm from heritable conditions. Nature. 2011;473:308–16.
59. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, Zheng YW, Leung TY, Lau TK, Cantor CR, Chiu RW. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2:61.
60. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350:485–7.

61. Lo YM, Tsui NB, Chiu RW, Lau TK, Leung TN, Heung MM, Gerovassili A, Jin Y, Nicolaides KH, Cantor CR, Ding C. Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med. 2007;13:218–23.
62. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet. 1999;64:218–24.
63. Loeys BL, Dietz HC. Loeys-Dietz Syndrome. Gene Reviews [Internet]. 2008. [Updated 2013 Jul 11].
64. Loeys BL, Schwartze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Manouvrier S, Roberts AE, Faravelli F, Greco MA, Pyeritz RE, Milewicz DM, Coucke PJ, Cameron DE, Braverman AC, Byers PH, De Paepe AM, Dietz HC. Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med. 2006;355(8):788–98.
65. Lopes LR, Zekavati A, Syrris P, Hubank M, Giambartolomei C, Dalageorgou C, Jenkins S, Mckenna W, UK10K Consortium, Plagnol V, Elliott PM. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. J Med Genet. 2013;50:228–39.
66. Lun FM, Chiu RW, Allen Chan KC, Yeung Leung T, Kin Lau T, Dennis Lo YM. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem. 2008;54:1664–72.
67. Mandelker D, Amr SS, Pugh T, Gowrisankar S, Shakhbatyan R, Duffy E, Bowser M, Harrison B, Lafferty K, Mahanta L, Rehm HL, Funke BH. Comprehensive diagnostic testing for stereocilin: an approach for analyzing medically important genes with high homology. J Mol Diagn. 2014;16(6):639–47.
68. Martin K, Iyengar S, Kalyan A, Lan C, Simon AL, Stosic M, Kobara K, Ravi H, Truong T, Ryan A, Demko ZP, Benn P. Clinical experience with a single-nucleotide polymorphism-based non-invasive prenatal test for five clinically significant microdeletions. Clin Genet. 2017; https://doi.org/10.1111/cge.13098. [Epub ahead of print].
69. Mason JA, Herrmann KR. Universal infant hearing screening by automated auditory brainstem response measurement. Pediatrics. 1998;101:221–8.
70. Matthews TJ, MacDorman MF, Thoma ME. Infant mortality statistics from the 2013 period linked birth/infant death data set. Natl Vital Stat Rep. 2015;64(9):1–30.
71. Meienberg J, Bruggman R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet. 2016;135(3):359–62.
72. Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Ventrini F, Zhang J, He W, Dharmadhikari AV, Qu C, Ward P, Braxton A, Narayanan S, Ge X, Tokita MJ, Santiago-Sim T, Dai H, Chiang T, Smith H, Azamian MS, Robak L, Bostwick BL, Schaaf CP, Potocki L, Scaglia F, Bacino CA, Hanchard NA, Wangler MF, Scott D, Brown C, Hu J, Belmont JW, Burrage LC, Graham BH, Sutton VR, Craigen WJ, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Muzny DM, Miller MJ, Wang X, Leduc MS, Xiao R, Liu P, Shaw C, Walkiewicz M, Bi W, Xia F, Lee B, Eng CM, Yang Y, Lalani SR. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. JAMA Pediatr. 2017;171(12):e173438.
73. Meynert AM, Bicknell LS, Hurles ME, Jackson AP, Taylor MS. Quantifying single nucleotide variant detection sensitivity in exome sequencing. BMC Bioinformatics. 2013;14:195.
74. Mitchell JJ, Capua A, Clow C, Scriver CR. Twenty-year outcome analysis of genetic screening programs for Tay-Sachs and beta-thalassemia disease carriers in high schools. Am J Hum Genet. 1996;59:793–8.
75. Moteki H, Azaiez H, Sloan-Heggen CM, Booth K, Nishio SY, Wakui K, Yamaguchi T, Kolbe DL, Iwasa YI, Shearer AE, Fukushima Y, Smith RJ, Usami SI. Detection and confirmation of deafness-causing copy number variations in the STRC gene by massively parallel sequencing and comparative genomic hybridization. Ann Otol Rhinol Laryngol. 2016;125(11):918–23.
76. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2009;42:30–5.
77. Nisbet DL, Griffin DR, Chitty LS. Prenatal features of Noonan syndrome. Prenat Diagn. 1999;19:642–7.
78. Overwater E, Floor K, van Beek D, de Boer K, van Dijk T, Hilhorst-Hofstee Y, Hoogeboom AJM, van Kaam KJ, van de Kamp JM, Kempers M, Krapels IPC, Kroes HY, Loeys B, Salemink S, Stumpel CTRM, Verhoeven VJM, Wijnands-van den Berg E, Cobben JM, van Tintelen JP, Weiss MM, Houweling AC, Maugeri A. NGS panel analysis in 24 ectopia lentis patients: a clinically relevant test with a high diagnostic yield. Eur J Med Genet. 2017;60(9):465–73.
79. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, Hill DE, Vidal M, Evans JG, Thorburn DR, Carr SA, Mootha VK. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008;134:112–23.
80. Pandya A. Genetic hearing loss: the journey of discovery to destination – how close are we to therapy? Mol Genet Genomic Med. 2016;4(6):583–7.
81. Parving A. The need for universal neonatal hearing screening – some aspects of epidemiology and identification. Acta Paediatr Suppl. 1999;88:69–72.
82. Perucca P, Scheffer IE, Harvey AS, James PA, Lunke S, Thorne N, Gaff C, Regan BM, Damiano JA, Hildebrand MS, Berkovic SF, O'Brien TJ, Kwan P. Real-world utility of whole exome sequencing with targeted gene analysis for focal epilepsy. Epilepsy Res. 2017;131:1–8.
83. Plevova P, Paprskarova M, Tvrda P, Turska P, Slavkovsky R, Mrazkova E. STRC deletion is a frequent cause of slight to moderate congenital hearing impairment in the Czech Republic. Otol Neurotol. 2017;38(10):e393–400.
84. Poninska JK, Bilinska ZT, Franaszczyk M, Michalak E, Rydzanicz M, Szpakowski E, Pollak A, Milanowska B, Truszkowska G, Chmielewski P, Sioma A, Janaszek-Sitkowska H, Klisiewicz A, Michalowska I, Makowiecka-Ciesla M, Kolsut P, Stawinski P, Foss-Nieradko B, Szperl M, Grzybowski J, Hoffman P, Januszewicz A, Kusmierczyk M, Ploski R. Next-generation sequencing for diagnosis of thoracic aortic aneurysms and dissections: diagnostic yield, novel mutations and genotype phenotype correlations. J Transl Med. 2016;14(1):115.
85. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, Dhar S, Wiszniewski W, Akdemir ZH, Gambin T, Xia F, Person RE, Walkiewicz M, Shaw CA, Sutton VR, Beaudet AL, Muzny D, Eng CM, Yang Y, Gibbs RA, Lupski JR, Boerwinkle E, Plon SE. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016;18(7):678–85.
86. Pritchard CC, Smith C, Marushchak T, Koehler K, Holmes H, Raskind W, Walsh T, Bennett RL. A mosaic PTEN mutation causing Cowden syndrome identified by deep sequencing. Genet Med. 2013;15:1004–7.
87. Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucińska-Więckowska A, Pajdowska M, Jurkiewicz E, Halat P, Kosińska J, Pollak A, Rydzanicz M, Stawinski P, Pronicki M, Krajewska-Walasek M, Płoski R. New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric Centre. J Transl Med. 2016;14(1):174.
88. Proost D, Vandeweyer G, Meester JAN, Salemink S, Kempers M, Ingram C, Peeters N, Saenen J, Vrints C, Lacro R, Roden D, Wuyts W, Dietz H, Mortier G, Loeys B, Van Laer L. Performant mutation identification using targeted next-generation sequencing of 14 thoracic aortic aneurysm genes. Hum Mutat. 2015;36(8):808–14.

89. Putcha GV, Bejjani BA, Bleoo S, Booker JK, Carey JC, Carson N, Das S, Dempsey MA, Gastier-Foster JM, Greinwald JH Jr, Hoffmann ML, Jeng LJ, Kenna MA, Khababa I, Lilley M, Mao R, Muralidharan K, Otani IM, Rehm HL, Schaefer F, Seltzer WK, Spector EB, Springer MA, Weck KE, Wenstrup RJ, Withrow S, Wu BL, Zariwala MA, Schrijver I. A multicenter study of the frequency and distribution of GJB2 and GJB6 mutations in a large North American cohort. Genet Med. 2007;9:413–26.
90. Pyeritz RE. Marfan syndrome: current and future clinical and genetic management of cardiovascular manifestations. Semin Thorac Cardiovasc Surg. 1993;5:11–6.
91. Raju H, Alberg C, Sagoo GS, Burton H, Behr ER. Inherited cardiomyopathies. BMJ. 2011;343:d6966.
92. Rauen KA. The RASopathies. Annu Rev Genomics Hum Genet. 2013;14:355–69.
93. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bai R, Suchy S, Friedman B, Tahiliani J, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18(7):696–704.
94. Retterer K, Scuffins J, Schmidt D, Lewis R, Pineda-Alvarez D, Stafford A, Schmidt L, Warren S, Gibellini F, Kondakova A, Blair A, Bale S, Matyakhina L, Meck J, Aradhya S, Haverfield E. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. Genet Med. 2015;17(8):623–9.
95. Sarkozy A, Conti E, Seripa D, Digilio MC, Grifone N, Tandoi C, Fazio VM, Di Ciommo V, Marino B, Pizzuti A, Dallapiccola B. Correlation between PTPN11 gene mutations and congenital heart defects in Noonan and LEOPARD syndromes. J Med Genet. 2003;40:704–8.
96. Scriver C, Beaudet AL, Sly WS, Valle D, editors. The metabolic and molecular bases of inherited disease. New York: McGraw-Hill; 1995.
97. Shahwan A, Farrell M, Delanty N. Progressive myoclonic epilepsies: a review of genetic and therapeutic aspects. Lancet Neurol. 2005;4:239–48.
98. Shearer AE, Deluca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, Scheetz TE, Smith RJ. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc Natl Acad Sci U S A. 2010;107:21104–9.
99. Shearer AE, Kolbe DL, Azaiez H, Sloan CM, Frees KL, Weaver AE, Clark ET, Nishimura CJ, Black-Ziegelbein EA, Smith RJ. Copy number variants are a common cause of non-syndromic hearing loss. Genome Med. 2014;6(5):37.
100. Sloan-Heggen CM, Bierer AO, Shearer AE, Kolbe DL, Nishimura CJ, Frees KL, Ephraim SS, Shibata SB, Booth KT, Campbell CA, Ranum PT, Weaver AE, Black-Ziegelbein EA, Wang D, Azaiez H, Smith RJ. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. Hum Genet. 2016;135(4):441–50.
101. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36(10):928–30.
102. Sosa MX, Sivakumar IK, Maragh S, Veeramachaneni V, Hariharan R, Parulekar M, Fredrikson KM, Harkins TT, Lin J, Feldman AB, Tata P, Ehret GB, Chakravarti A. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. PLoS Comput Biol. 2012;8:e1002737.
103. Stevenson DA, Carey JC, Byrne JL, Srisukhumbowornchai S, Feldkamp ML. Analysis of skeletal dysplasias in the Utah population. Am J Med Genet A. 2012;158A:1046–54.
104. Stosser MB, Lindy AS, Butler E, Retterer K, Piccirillo-Stosser CM, Richard G, McKnight DA. High frequency of mosaic pathogenic variants in genes causing epilepsy-related neurodevelopmental disorders. Genet Med. 2017; https://doi.org/10.1038/gim.2017.114. [Epub ahead of print].
105. Tan L, Li Z, Zhou C, Cao Y, Zhang L, Li X, Cianflone K, Wang Y, Wang DW. FBN1 mutations largely contribute to sporadic non-syndromic aortic dissection. Hum Mol Genet. 2017;26(24):4814–22.
106. Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, Chong B, Phelan D, Brett GR, Creed E, Jarmolowicz A, Yap P, Walsh M, Downie L, Amor DJ, Savarirayan R, McGillivray G, Yeung A, Peters H, Robertson SJ, Robinson AJ, Macciocca I, Sadedin S, Bell K, Oshlack A, Georgeson P, Thorne N, Gaff C, White SM. Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. JAMA Pediatr. 2017;171(9):855–62.
107. Teekakirikul P, Kelly MA, Rehm HL, Lakdawala NK, Funke BH. Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era. J Mol Diagn. 2013;15:158–70.
108. Valenzuela I, Fernández-Alvarez P, Munell F, Sanchez-Montanez A, Giralt G, Vendrell T, Tizzano EF. Arthrogryposis as a neonatal presentation of Loeys-Dietz syndrome due to a novel TGFBR2 mutation. Eur J Med Genet. 2017;60(6):303–7.
109. Van Camp G, Willems PJ, Smith RJ. Nonsyndromic hearing impairment: unparalleled heterogeneity. Am J Hum Genet. 1997;60:758–64.
110. Van Diemen CC, Kerstjens-Frederikse WS, Bergman KA, de Koning TJ, Sikkema-Raddatz B, van der Velde JK, Abbott KM, Herkert JC, Löhner K, Rump P, Meems-Veldhuis MT, Neerincx PBT, Jongbloed JDH, van Ravenswaaij-Arts CM, Swertz MA, Sinke RJ, van Langen IM, Wijmenga C. Rapid targeted genomics in critically ill newborns. Pediatrics. 2017;140(4):1–11.
111. Laer V, Dietz H, Loeys B. Loeys-Dietz syndrome. Adv Exp Med Biol. 2014;802:95–105.
112. Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. Genome Med. 2009;1:100.
113. Verhoef S, Bakker L, Tempelaars AM, Hesseling-Janssen AL, Mazurczak T, Jozwiak S, Fois A, Bartalini G, Zonnenberg BA, Van Essen AJ, Lindhout D, Halley DJ, Van Den Ouweland AM. High rate of mosaicism in tuberous sclerosis complex. Am J Hum Genet. 1999;64:1632–7.
114. Verloes A, Sakalihasan N, Koulischer L, Limet R. Aneurysms of the abdominal aorta: familial and genetic aspects in three hundred thirteen pedigrees. J Vasc Surg. 1995;21:646–55.
115. Vigevano F. Benign familial infantile seizures. Brain and Development. 2005;27:172–7.
116. Vona B, Hofrichter MA, Neuner C, Schröder J, Gehrig A, Hennermann JB, Kraus F, Shehata-Dieler W, Klopocki E, Nanda I, Haaf T. DFNB16 is a frequent cause of congenital hearing impairment: implementation of STRC mutation analysis in routine diagnostics. Clin Genet. 2015;87(1):49–55.
117. Wang HD, Liu L, Zhao HR, Hou QF, Yan JB, Shi WL, Guo QN, Wang L, Liao SX, Zhu BF. Detection of fetal epigenetic biomarkers through genome-wide DNA methylation study for non-invasive prenatal diagnosis. Mol Med Rep. 2017;15(6):3989–98.
118. Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, He N, Liao WP. Epilepsy-associated genes. Seizure. 2017;44:11–20.
119. Wang Y, Du X, Bin R, Yu S, Xia Z, Zheng G, Zhong J, Zhang Y, Jiang YH, Wang Y. Corrigendum: genetic variants identified from epilepsy of unknown etiology in Chinese children by targeted exome sequencing. Sci Rep. 2017;7:46520.
120. Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, Lemerrer M, Mortier G, Mundlos S, Nishimura G, Rimoin DL, Robertson S, Savarirayan R, Sillence D, Spranger J, Unger S, Zabel B, Superti-Furga A. Nosology and classification of genetic skeletal disorders: 2010 revision. Am J Med Genet A. 2011;155A:943–68.

121. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. N Engl J Med. 2011;364:1643–56.
122. Wei F, Yan LM, Su T, He N, Lin ZJ, Wang J, Shi YW, Yi YH, Liao WP. Ion channel genes and epilepsy: functional alteration, pathogenic potential, and mechanism of epilepsy. Neurosci Bull. 2017;33(4):455–77.
123. Wilde AA, Behr ER. Genetic testing for inherited cardiac disease. Nat Rev Cardiol. 2013;10:571–83.
124. Williams E, Richard G, Cho M, et al. Diagnostics yield from reanalysis of whole exome sequencing data. Poster presented at: 2016 ACMG Annual Clinical Genetics Meeting; 8–12 March 2016; Tampa, FL.
125. Wooderchak-Donahue WL, O'fallon B, Furtado LV, Durtschi JD, Plant P, Ridge PG, Rope AF, Yetman AT, Bayrak-Toydemir P. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel-clinical diagnostics perspective. BMC Med Genet. 2012;5:50.
126. Woolnough R, Dhawan A, Dow K, Walia JS. Are patients with Loeys-Dietz syndrome misdiagnosed with Beals syndrome? Pediatrics. 2017;139(3):e20161281.
127. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13:255–62.
128. Xiao Y, Liu X, Guo X, Liu L, Jiang L, Wang Q, Gong B. A novel FBN1 mutation causes autosomal dominant Marfan syndrome. Mol Med Rep. 2017;16(5):7321–8.
129. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369:1502–11.
130. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WF, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312(18):1870–9.
131. Zazo Seco C, Wesdorp M, Feenstra I, Pfundt R, Hehir-Kwa JY, Lelieveld SH, Castelein S, Gilissen C, de Wijs IJ, Admiraal RJ, Pennings RJ, Kunst HP, van de Kamp JM, Tamminga S, Houweling AC, Plomp AS, Maas SM, de Koning Gans PA, Kant SG, de Geus CM, Frints SG, Vanhoutte EK, van Dooren MF, van den Boogaard MH, Scheffer H, Nelen M, Kremer H, Hoefsloot L, Schraders M, Yntema HG. The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. Eur J Hum Genet. 2017;25(3):308–14.
132. Ziganshin BA, Bailey AE, Coons C, Dykas D, Charilaou P, Tanriverdi LH, Liu L, Tranquilli M, Bale AE, Elefteriades JA. Routine genetic testing for thoracic aortic aneurysm and dissection in a clinical setting. Ann Thorac Surg. 2015;100(5):1604–11.

36 Sequencing Cell-Free DNA in the Maternal Circulation to Screen for Down Syndrome, Other Common Trisomies, and Selected Genetic Disorders

Glenn E. Palomaki and Robert G. Best

Introduction and Brief History

Screening is different from most forms of medical testing in that the individuals opting for screening are generally healthy and/or not known to be at increased risk for the disorder being tested. Those offering or recommending screening tests need to ensure that the screening test performance is well characterized and acceptable, usually by determining the detection rate or sensitivity (the proportion of affected individuals with a positive test) and the false-positive rate (proportion of unaffected individuals with a positive test, or 1-specificity) are favorable compared with other means of screening for the same disorder. A working definition of screening has been proposed by Wald [1] and will be adhered to in this chapter.

> Screening is the systematic application of a test or enquiry to identify individuals at sufficient risk of a specific disorder to warrant further investigation or direct preventive action, amongst persons who have not sought medical attention on account of symptoms of that disorder.

Although Down syndrome was first described in 1862 by Langdon Down [2], the cause of this disorder (an extra chromosome 21) was not determined until 1959 [3]. This led directly to the ability to prenatally diagnose Down syndrome via amniocentesis and karyotyping. Given the well-known association between increasing birth prevalence and advancing maternal age, the first screening test for Down syndrome was the question "How old will you be at delivery?" Until 2007 [4], this question formed the basis of most prenatal screening for Down syndrome, although laboratory tests apart from karyotyping were, by then, available.

In the early 1980s, women routinely had second trimester maternal serum alpha-fetoprotein (AFP) measured as a screening test for open neural tube defects, specifically open spina bifida. In 1984, a woman asked her physician whether her very low AFP measurement might be related to the diagnosis of trisomy 18 in her newborn. This query prompted a line of investigation that culminated in the discovery that second trimester AFP correlates not only with fetal trisomy 18 but also with Down syndrome [5]. The precise reason for these reductions in AFP levels remains unknown but may reflect differences in liver and placental metabolism between aneuploid and euploid fetuses. Discovery of these associations was quickly verified [6] and a screening algorithm devised that combined maternal age and AFP levels to provide patient-specific estimates of risk [7]. Further research identified additional early second trimester serum markers [8–10] leading to the development of the "triple test," which includes maternal age in combination with AFP, unconjugated estriol (uE3), and human chorionic gonadotropin (hCG) measurements. The current, best serum-based second trimester combination (the triple test with dimeric inhibin-A measurements) is commonly called the "quadruple" test for Down syndrome and has a detection rate of 80% at a false-positive rate of 5% [10, 11]. During the same time period, additional markers were identified that could be measured in the late first trimester. These included not only the serum markers (PAPP-A [12] and free beta hCG [13]) but also an ultrasound measurement of the translucent space between the spine and skin of the fetus (nuchal translucency or NT) [14]. Together with maternal age, these markers formed the basis of the "combined" first trimester test [15]. Detection and false-positive rates for the combined test are similar to, or slightly better than, the quadruple test.

Throughout the 1990s, first and second trimester tests were offered independently as alternative methods for aneuploidy screening. Many screening programs elected to offer testing in the first or second trimester, but not both. In 1999,

G. E. Palomaki (✉)
Division of Medical Screening and Special Testing, Department of Pathology and Laboratory Medicine, Women and Infants Hospital/Alpert Medical School of Brown University, Providence, RI, USA

Savjani Institute for Health Research, Windham, ME, USA
e-mail: gpalomaki@ipmms.org

R. G. Best
Biomedical Sciences, University of South Carolina School of Medicine/Greenville Health System, Greenville, SC, USA

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_36

the concept of the "integrated" test was introduced [16], in which the best first and second trimester markers were combined into a single second trimester risk estimate. This "integrated" test has the best performance of any widely offered prenatal screening test based on biochemical and ultrasound measurements, with a detection rate of 90% at a false positive rate of 3% [17].

As part of an external proficiency testing program administered by the College of American Pathologists (CAP) and the American College of Medical Genetics and Genomics (www.cap.org), a survey of participating prenatal screening laboratories has been conducted within the first of three distributions each year, beginning in 2011. This survey asks participants what tests they offer for prenatal screening and how many of each are performed each year. Results from the 2011 and 2012 survey were compiled [18], and the results are summarized in Table 36.1. Overall, the survey documents that 72% of the 4.24 million pregnancies each year in the USA have been screened for Down syndrome by a total of 123 laboratories. Testing most commonly occurs in the second trimester (60% of tested women), followed by various forms of the integrated test (21%) and combined testing (19%). These numbers do not include women who opt directly for invasive testing (amniocentesis or chorionic villus sampling) and diagnostic testing (karyotype or microarray). These data were updated in 2015 and showed a general increase in serum-/ultrasound-based screening to 77% of the pregnancies in the USA [19].

Serum markers are readily available throughout the USA with measurements made using FDA-cleared instruments and reagents. The results are subject to external proficiency testing that shows that they can be implemented reliably in a variety of high-complexity laboratory settings. The costs of reagents are relatively low with test charges ranging from $100 to $200 or more depending on the test. Ultrasound measurements are operator dependent, but several organizations provide training and oversight (http://www.fetalmedicineusa.com/FMF/US, https://www.ntqr.org/). The turnaround time for combined or quadruple testing is 1 or 2 days. One implementation in the UK is the One Stop Clinic for Assessment of fetal Risk, or OSCAR clinic [20], where the ultrasound measurement, biochemistry testing, reporting of results, counseling, and the offer of a diagnostic test, if indicated, can be made within an hour.

Although the detection and false-positive rates of screening tests for Down syndrome, and to some extent trisomy 18 and trisomy 13, have greatly improved, the prenatal care community has two main objectives: (1) prenatal screening and diagnosis should occur early in pregnancy, preferably being completed by the end of the first trimester, and (2) the number of invasive procedures should be kept to a minimum to reduce the possibility of procedure-related loss. Optimally, a noninvasive prenatal diagnostic (NIPD) test or set of tests might be developed that could identify not only Down syndrome and the other common trisomies but also other prenatally diagnosable conditions, many of which can now only be identified after invasive procedures such as amniocentesis or chorionic villus sampling (CVS) and using diagnostic techniques such as arrayed comparative genomic hybridization (aCGH). Testing cell-free nucleic acids in maternal circulation is a major step forward in the evolution of prenatal screening and diagnoses.

Table 36.1 Types and numbers of Down syndrome screening tests performed in US laboratories in 2012

Type of test	Laboratories	Median (*N*)	Number (%)
First trimester[a]	34	3000	565,692 (19)
Second trimester	122	2538	1,770,024 (60)
AFP only	85	720	235,492
Triple test	44	402	90,132
Quadruple test[b]	118	2400	1,443,900
Integrated	30	4176	583,416 (21)
Fully integrated	22	2136	102,972
Serum integrated	21	888	119,760
Sequential	24	2436	405,144
All	123[c]	3660	2,963,592 (100)

[a]Includes all first trimester tests, including those using serum measurements of total/intact hCG, free beta hCG and dimeric inhibin-A
[b]Includes tests with five second trimester serum markers
[c]Does not add up, as some laboratories are counted in multiple "types of tests"

Discovery, Identification, and Initial Uses of cfDNA

Identification of cfDNA from the Fetus in Maternal Plasma

In 1997, a landmark publication [21] reported the presence of fetal/placental DNA in maternal plasma and serum. This finding was based on earlier work showing circulating cell-free cancer DNA in the plasma of affected patients [22–24]. In that 1997 study, pregnant and nonpregnant women provided serum, plasma, and whole blood samples for testing. The fetal sex was determined at birth or by karyotype after amniocentesis. After extraction, DNA was tested for presence of a Y-chromosome sequence (DYS14). Of the 30 pregnancies with a male fetus, plasma testing identified 24 (80%); none of the 13 female fetuses were misclassified. Fewer males were identified using serum and whole blood, but again, none of samples from the 13 female fetuses were positive for the Y probe. The 80% detection rate for males was, in hindsight, likely due to the small plasma sample size (10 μL), as most of the missed male fetuses were sampled earlier in gestation when the circulating cfDNA is lower.

Using cfDNA in Maternal Plasma to Determine Fetal Sex

Soon after the report by Lo and colleagues [21], these findings were confirmed and expanded by other groups. Early fetal sexing can be useful to resolve ambiguous genitalia, manage X-linked conditions, and help identify some single-gene disorders (e.g., congenital adrenal hyperplasia). Noninvasive testing could thus be used in place of invasive cytogenetic testing. Because early fetal sex determination via ultrasound is not reliable for this use in the first trimester [25], cfDNA testing might be a more reliable alternative. A summary of 57 published studies on the use of cfDNA to identify fetal sex was published in 2011 [26]. Overall, 3524 male and 3017 female pregnancy samples were included. The screening performance was better after 7 weeks of gestation, when plasma rather than serum was used and when real-time quantitative polymerase chain reaction (RQ-PCR) was employed. At 7–20 weeks' gestation, the use of RQ-PCR results in an estimated detection rate for identifying a male fetus of 98%, with a corresponding false positive rate of 0.9%.

Use of cfDNA in Maternal Plasma to Determine Fetal Rh Status

Among RhD-negative women, RhD-positive fetal cells crossing the placenta can cause the mother to make anti-RhD antibodies. This Rh incompatibility can lead to fetal complications. Rh incompatibility is preventable by providing injection of Rh immunoglobulin to RhD-negative women in the second trimester. However, treatment is unnecessary if the fetus is also RhD negative. Testing cfDNA can identify these pregnancies and avoid unnecessary treatment. In a nationwide study in Denmark [27], 2312 RhD-negative women were tested both by routine genotyping and by testing of cfDNA for RhD status at 25 weeks' gestation. Overall, the cfDNA test had a 99.9% detection rate with 96.5% accuracy. A total of 862 of these women avoided unnecessary treatment; 39 women with an RhD-negative fetus still received treatment (unnecessary treatment), and two women with RhD-positive fetuses were not detected and the mothers were not treated (false negatives).

Sequencing cfDNA to Identify Down Syndrome: Preliminary Studies

Sequencing cfDNA as a prenatal screening test for Down syndrome is a disruptive innovation that exploits recent dramatic improvements in the speed and cost of DNA sequencing and analysis. Although the circulating fetal DNA has been under study since 1997, it wasn't until 2008 when reductions in testing costs and growth in bioinformatic capabilities provided the conditions necessary for proof of principle in a clinical prenatal population. Such testing was initially thought to be diagnostic, so the term "noninvasive prenatal diagnosis" (NIPD) was sometimes applied to early results. Later, after it was clear that both false-positive and false-negative results occurred, the test was identified as "noninvasive prenatal testing" (NIPT). Later the term "noninvasive prenatal screening" (NIPS) was recommended to avoid confusion related to diagnostic versus screening distinctions. Unfortunately, neither of these terms clearly distinguish between serum/ultrasound-based screening and cfDNA-based screening. To avoid confusion, this chapter will use the phrase "cfDNA screening" for Down syndrome and other genetic disorders.

Serum and ultrasound screening for common trisomies is based on phenotypic findings. For example, infants with Down syndrome are known to have thickening of the skin on the back of the neck. Sonographers tried to identify this finding during pregnancy and eventually discovered that increased nuchal translucency thickness measurements in the late first trimester were strongly associated with Down syndrome. Testing cfDNA, instead, focuses on genotypic markers that are directly measuring the number of chromosome 21's in order to screen for Down syndrome. This is a radical departure from earlier prenatal screening tests and should result in higher overall test performance.

This chapter focuses on the initial development, validation, and introduction of three distinctly different methodologies used to screen for the common autosomal aneuploidies using cfDNA obtained from maternal plasma. Specifically, these methods are based on (1) shotgun sequencing, (2) chromosome-specific sequencing, and (3) SNP-based pattern matching. This section is devoted to preliminary academic proof-of-concept and in-house validation studies performed by commercial companies. Later sections will describe the results from collaborations with academic sites that performed external validation studies and results of offering cfDNA screening as a clinical test.

Massively Parallel Shotgun Sequencing (MPSS)

In late 2008, two studies [28, 29] demonstrated the potential for shotgun sequencing of cfDNA in maternal plasma to identify common autosomal trisomies. Both groups used similar sequencing and counting methodologies. After sequencing random cfDNA fragments, the results were mapped to the human genome to identify the chromosome of origin. The number of fragments aligned to chromosome 21 was then compared to the sum of fragments aligned to all other autosomal chromosomes. A normalizing function was then applied to account for the varying number of total matched reads per sample. These normalized values were then compared with reference ranges from known euploid samples.

In the study by Fan and colleagues [29], nine samples from Down syndrome pregnancies showed "normalized sequence

tag densities" for chromosome 21 that were 4–18% higher than the corresponding densities in samples from euploid pregnancies. Two trisomy 18 samples were also detected, but one trisomy 13 sample was only slightly elevated. On average, about ten million DNA fragments were sequenced per patient, and about five million were aligned and suitable for analysis. This study was limited in that most trisomic samples were collected after amniocentesis (which could influence the results), some samples were collected in the third trimester, no trisomic samples were collected prior to 14 weeks' gestation, and the number of control samples (six) was very limited.

In the study by Chiu et al. [28], all 14 samples from Down syndrome pregnancies had chromosome 21 *z*-scores of 5 or higher, while none of the 14 samples from euploid pregnancies had values above 3. The *z*-score measures the difference between the chromosome 21% in the patient minus the average percentage in euploid pregnancies, divided by the standard deviation of chromosome 21 percentages in known euploid pregnancies. All samples were collected at 20 weeks' gestation or earlier. On average, about 11 million DNA fragments were sequenced and about 2.5 million were aligned and suitable for analysis. Limitations of this study include the collection of samples after an invasive procedure for 11 of 14 cases, and no trisomy 18 or trisomy 13 samples were tested.

In a subsequent larger scale collaborative study [30], samples from three sites were available for testing. In order to scale the test for larger throughput, this group multiplexed eight samples per flow cell lane (Illumina, San Diego, CA), allowing for up to 64 patients per run. Seven hundred and fifty-three samples were run, including 86 samples from Down syndrome pregnancies. Unfortunately, there was an average of only 300,000 aligned reads per patient, resulting in a detection rate of 79% with a false-positive rate of 1.1%. A subset of the samples was available to rerun at a 2-plex, with an average of 2.3 million aligned reads per patient. The Down syndrome detection rate improved to 100%, with a 2.1% false-positive rate. For the first time, a study reported that it was not possible to obtain results on a subset of samples (1.4%) that passed specimen quality requirements, but not sequencing quality metrics. This "failure" rate is an important consideration when examining the performance of this type of testing. The fetal fraction (measured by Y probes among male fetuses) had a median value of 15%. This was also the first report that showed the clear relationship between fetal fraction (measured by Y probes among male fetuses) and the chromosome 21 *z*-score. The median fetal fraction was 15%. Limitations of the study include the unblinded re-running of the samples at a greater depth (2-plex) and the lack of a training set of known euploid samples prior to test interpretation.

A publication [31] from a commercial company (Sequenom, Inc., San Diego, CA) provided results from three small preliminary studies as well as a larger clinical validation of a laboratory developed test that could be suitable for introduction into practice. In the larger study, a total of 40 Down syndrome cases were matched with 440 euploid control pregnancies. Thirteen samples were not suitable for testing (e.g., too little plasma, sample dropped, or broken), and another 18 failed quality control (e.g., low fetal DNA, low library concentration, too few aligned counts), including 1 case of Down syndrome. The classification system relied on a single *z*-score cutoff of 3.0. The detection rate was 100% (39/39) with no false positives (0/410). The failure rate was 4%. At least 12 million matched reads were available for each patient. The median gestational age was 15 weeks, but some late second and third trimester samples were included, along with some cases that had been collected after diagnostic testing.

Another study [32] from a privately held company (Verinata, Inc., now wholly owned by Illumina, San Diego, CA) described a laboratory developed test based on shotgun sequencing that was aimed at detecting Down syndrome, trisomy 18, and trisomy 13, as well as selected sex aneuploidies. Their method differs slightly in the result interpretation. It relies on comparing target chromosome results with (a) specific comparison chromosome(s). For example, the chromosome 21 counts were compared against counts from chromosome 9 (in the methodology described in the previous paragraphs, it would have been compared to the counts from all other autosomes). Comparator chromosome(s) were chosen based on minimizing the variances of the ratio over runs. This approach may reduce variability because the comparator chromosome is similar in GC content to the chromosome of interest. The mean and standard deviation of these ratios in known euploid pregnancies can then be used to assign a "normalized chromosome value" which can be interpreted as a *z*-score. After training, a set of 48 samples were blindly tested. On average, ten million matched reads per patient were available for analysis. For Down syndrome, the detection rate was 100% (13/13). The classification method used by this group employs two cutoffs: results below 2.5 are considered euploid while those above 4.0 are considered trisomic. Those in-between the cutoffs are classified as "no-calls." For this group, the no-calls were not test failures. Rather, testing provide a result that was not clearly positive or negative but borderline. In practice, borderline calls would be considered screen positive.

Alternatives to Shotgun Sequencing

The previous sections focused on shotgun sequencing methodologies. If the testing remains focused on chromosomes 21, 18, 13, X, and Y, then one might consider preferentially amplifying targets located on these chromosomes as a way to reduce "unnecessary" sequencing time and expense. These next two methodologies apply targeted sequencing to maternal plasma DNA sequencing for common aneuploidies.

Chromosome-Specific Sequencing (CSS)

Two preliminary papers described a targeted sequencing methodology for common aneuploidies. One [33] uses CSS coupled with a simple interpretive method, while another [34] replaces that simple method with the "FORTE" algorithm that allows for patient-specific risks. According to this second publication [34], several hundred non-polymorphic loci were identified on each chromosome of interest (chromosomes 21 and 18), and a subset of the most predictive loci was then identified using the training set. After normalizing the counts for assay and patient biases, the chromosome 21 results were compared to the chromosome 18 results (under the assumption that no fetus will be trisomic for both chromosomes). The result was expressed as a normalized *z*-score. The *z*-score was then transformed into a risk by multiplying a prior risk for Down syndrome (based on mother's age) and a likelihood ratio derived from the overlapping distributions of *z*-scores in affected and unaffected pregnancies, after accounting for fetal fraction. The test set consisted of samples from 123 euploid and 36 Down syndrome pregnancies. Some cases were collected after invasive testing, and the control pregnancies were women at low risk of aneuploidy. Some samples were collected in the third trimester. All 36 cases were identified (100% detection rate), and no false-positive results occurred (0% false positive rate). There were no failures. Similar results were reported using only the *z*-score [33].

SNP-Based Pattern Matching (SNP)

The test methodology behind this publication [35] relies on a highly multiplexed PCR reaction that amplifies thousands of single nucleotide polymorphisms (SNPs) on chromosomes 13, 18, 21, X, and Y. Testing of the plasma was supplemented by maternal genotyping of the same SNPs using the sample's buffy coat. A complex matrix of potential genotypes based on crossover frequencies, source of the extra chromosome, single nucleotide polymorphism (SNP) copies (mono, di, and tri), and fetal fraction are modeled and the paired genotypes fitted using a Bayesian maximum likelihood method. The most likely model is the sum of the likelihoods and is reported as an accuracy score (from 0% to 100%). The algorithm is capable of incorporating the paternal genotype as well, if available. A total of 161 samples were collected, most between 9 and 25 weeks of gestation. Most aneuploidies were collected after invasive testing. Some euploid samples were collected prior to confirmatory invasive testing, but many were simply assumed to be euploid. Both maternal and paternal genotypes were available for analysis. Overall, 20 samples (12%) failed quality parameters, many with low fetal fractions. Among the remaining samples, 11 Down syndrome cases were correctly identified (100% detection rate), and all of the control samples were negative (0% false-positive rate).

Summary of Proof-of-Concept and Preliminary Examination of Laboratory Developed Tests

These preliminary studies all had important weaknesses including small sample sizes, samples taken at gestational ages not relevant for routine prenatal screening, samples taken after (rather than before) an invasive procedure, a limited range of abnormalities, and the need for training sets of known euploid pregnancies. However, they all provide important information regarding required elements for such testing. Clearly, samples with low fetal fraction will be more difficult to interpret correctly, the total number of matched reads will be important, and there will be variability in how results are interpreted (two or three categories or patient-specific risk assessment).

Down Syndrome Screening: Clinical Validation Studies

This section focuses on clinical validity studies which require that a locked down assay be blindly tested. The focus is on screening for Down syndrome only. Other aneuploidies that may have been part of the reports will be discussed in a later section. Seven structured reviews and associated meta-analyses of cfDNA-based screening for Down syndrome have been published since 2014 [36–42]. Although the study inclusion criteria and timeframes for these reviews differ, they provide an excellent overview of the detection and false-positive rates for cfDNA screening for Down syndrome (Table 36.2). It is impor-

Table 36.2 Summary of seven structured reviews of Down syndrome screening among cfDNA samples with successful testing

Review	Studies	DS	Euploid	DR (%)[a]	FPR (%)[a]
Gil et al. [36]	18	809	11,646	99.9 (98.2–99.6)	0.07 (0.03–0.12)
Gil et al. [37]	24	1051	21,608	99.2 (98.5–99.6)	0.13 (0.07–0.20)
Taylor-Phillips et al. [38]	40	2245	184,099	99.3 (98.9–99.6)	0.10 (0.00–0.10)
Mackie et al. [39]	31	721	111,544	99.4 (98.3–99.8)	0.10 (0.00–0.10)
Iwarsson et al. [40]	26[b]	1847	105,627	99.8 (98.0–99.9)	0.10 (0.10–1.00)
	6[b]	157	62,144	99.3 (95.5–98.7)	0.10 (0.10–0.20)
Gil et al. [41]	30	1963	225,032	99.7 (99.1–99.9)	0.04 (0.02–0.08)
Jin et al. [42]	44	2097	166,080	99.7 (98.3–99.9)	0.00 (0.00–0.10)

DS Down syndrome, *DR* detection rate (sensitivity), *FPR* false-positive rate (1-specificity)
[a]Reported as point estimate and 95% confidence interval
[b]Results from the 26 studies of high-risk pregnancies are separated from the results in 6 low-risk studies

tant to understand that these results apply only when the cfDNA testing is successful (e.g., test failure, inadequate samples were excluded). Some Down syndrome pregnancies will occur among these failed tests, and the impact of this on the overall screening performance will be discussed later. In general, the reviews published prior to 2016 include studies of high-risk pregnancies that would undergo diagnostic testing. The later reviews would often include or restrict inclusion to cohort studies where only a subset of patients had diagnostic testing and follow-up of pregnancy outcome may have been incomplete. In general, these reviews did not find differences in overall test performance when results were stratified by test methodology (shotgun, CSS, SNP), by gestational age, by high-risk/low-risk classification, or by other potential covariates. Overall, the median pooled detection rate for the seven structured reviews is 99.5% with an associated median false positive rate of 0.1%.

Other Common Autosomal Aneuploidies Currently Detected by Maternal Plasma DNA Testing

Trisomy 18

Trisomy 18 appears to be more difficult to detect than Down syndrome. The GC content of chromosome 18 makes the counting techniques more challenging [28–32]. This problem can be addressed by improvements in the sequencing chemistry [28] and through bioinformatics [43]. In addition, it is now apparent that the fetal fractions may also be lower in trisomy 18 pregnancies due to a smaller placenta, resulting in more test failures and reduced average separation between euploid and trisomy 18 scores. Six of the seven structured reviews also examined the literature on the screening performance of cfDNA for the detection of trisomy 18 in singleton pregnancies (see the previous section on Down syndrome for information about how these review might differ and the interpretation of their results). Table 36.3 shows the resulting summary pooled detection and false-positive rates for the six relevant structured reviews. Overall, a reasonable overall detection rate is 97.5% with an associated false-positive rate of 0.1%.

Trisomy 13 (Patau Syndrome)

Trisomy 13 also appears to be more difficult to detect than Down syndrome. The GC content may be part of the reason and would be addressed in the same way as when testing for trisomy 18 (or Down syndrome). All but one [42] of the seven structured reviews also examined the literature on the screening performance of cfDNA for the detection of trisomy 13 in singleton pregnancies (see the previous section on Down syndrome for information about how these reviews might differ and the interpretation of their results). Table 36.4 shows the resulting summary pooled detection and false-positive rates. Given the much smaller number of trisomy 13 cases observed, the 95% confidence intervals are wider, and the detection rate estimates vary considerably from 90.6% to 99.0%. A reasonable estimate of the detection rate for the six structured reviews is 95% with an associated false-positive rate of about 0.1%.

Combined cfDNA Screening for Down Syndrome, Trisomy 18, and Trisomy 13

The three previous section examined the performance of screening for each of the three common autosomal trisomies in isolation. When combined, the false-positive rates will be additive. Figure 36.1 shows the potential impact of cfDNA screening on a general population of 100,000 pregnancies, when restricted to these three disorders. The maternal age distribution is from the 2015 US birth records (https://wonder.cdc.gov/controller/datarequest/D66), and the prevalence and fetal loss rates from the first trimester to term are from recent large studies [44]. For Down syndrome, trisomy 18, and trisomy 13, the detection rates among singleton pregnancies with successful testing are 99.5%, 97.5%, and 95.0%, respectively. The corresponding false-positive rates are each 0.1%, for a total of 0.3%. First trimester prevalence of the three disorders are 1:340, 1:1130, and 1:3510, respectively.

Table 36.3 Summary of six structured reviews of trisomy 18 screening among cfDNA samples with successful testing

Review	Studies	T18	Euploid	DR (%)[a]	FPR (%)[a]
Gil et al. [36]	15	301	11,646	96.8 (94.5–98.4)	0.15 (0.08–0.25)
Gil et al. [37]	21	389	21,306	96.3 (94.3–97.9)	0.13 (0.07–0.20)
Taylor-Phillips et al. [38]	33	683	184,419	97.4 (95.8–98.4)	0.10 (0.00–0.10)
Mackie et al. [39]	24	444	146,496	97.7 (95.2–98.9)	0.01 (0.00–0.02)
Iwarsson et al. [40]	26	157	62,144	97.7 (95.8–98.7)	0.01 (0.00–0.02)
Gil et al. [41]	25	563	222,013	97.9 (94.9–99.1)	0.04 (0.03–0.07)

DS Down syndrome, *DR* detection rate (sensitivity), *FPR* false-positive rate (1-specificity)
[a]Reported as point estimate and 95% confidence interval

Table 36.4 Summary of six structured reviews of trisomy 13 screening among cfDNA samples from singleton pregnancies with successful testing

Review	Studies	T13	Euploid	DR (%)[a]	FPR (%)[a]
Gil et al. [36]	11	85	8339	92.1 (85.9–96.7)	0.20 (0.04–0.46)
Gil et al. [37]	18	138	18,059	91.0 (84.0–95.6)	0.13 (0.05–0.26)
Taylor-Phillips et al. [38]	24	167	160,325	97.4 (86.1–99.6)	0.10 (0.00–0.10)
Mackie et al. [39]	16	85	134,606	90.6 (82.3–95.8)	0.01 (0.00–0.01)
Iwarsson et al. [40]	18	144	137,555	97.5 (81.9–99.7)	0.10 (0.10–0.10)
Gil et al. [41]	23	119	212,883	99.0 (65.8–100)	0.04 (0.02–0.07)

T13 trisomy 13 (Patio syndrome), *DR* detection rate (sensitivity), *FPR* false-positive rate (1-specificity)
[a]Reported as point estimate and 95% confidence interval

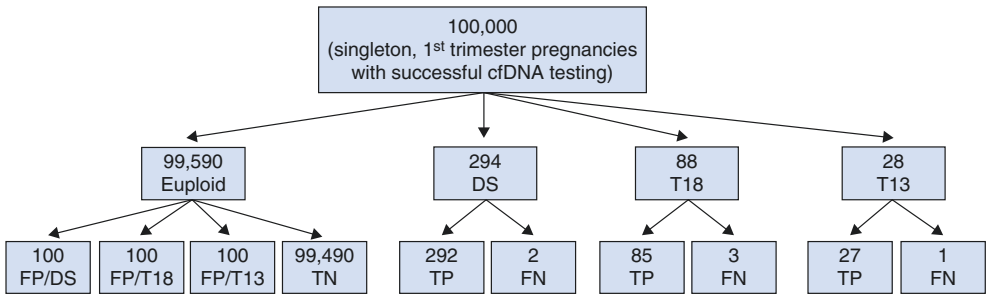


Fig. 36.1 Flow diagram showing the impact of screening a hypothetical cohort of 100,000 women from the general pregnancy population for Down syndrome, trisomy 18, and trisomy 13 using cfDNA from the maternal circulation

The positive predictive values (PPV) for the three disorders in a general pregnancy population are estimated to be 74% (odds 2.9:1), 46% (1:1.2), and 13% (1:3.3), respectively. The weighted overall detection rate for the three trisomies is 98.5% at a 0.3% false-positive rate, leading to an overall PPV of 57% (1.3:1). These are considerably lower PPVs than have been reported in the literature where PPVs of 90% or higher have been reported. Given the well-known impact of prevalence on PPV, this is an expected finding, as most such testing has occurred in high-risk populations where the prevalence of these trisomies is much higher than in a general pregnancy population. Modeling results in high-risk populations would be dependent on how the "high risk" was defined and would be expected to vary from setting to setting. Among the 6 false-negative cases occurring in the 100,000 screened pregnancies, 2 were Down syndrome, 3 were trisomy 18, and 1 was trisomy 13.

Sex Aneuploidies and Fetal Sex

Prenatal screening tends to be focused on Down syndrome. Trisomy 18 and trisomy 13 would not, by themselves, be sufficiently common or have the long-term survival to warrant organizing a stand-alone screening program just for these disorders. Common sex aneuploidies would most likely also not warrant a screening program of their own. These disorders are less serious, and there have been arguments that they should not be part of routine prenatal screening. Currently, however, most sex aneuploidies are identified around puberty or via presentation with reproductive issues. Identifying them at that time in life is too late for them to receive hormone treatments that might improve their quality of life. Given that there is no newborn screening for sex aneuploidies, prenatal screening may be one way to identify these individuals. Early experience indicates that 60–90% of women opt for sex aneuploidy testing when given the choice. A high proportion of women with screen-positive results for sex aneuploidies do not undergo subsequent diagnostic testing. Rather, they choose to delay testing until after the child is born. Some professional organizations recommend that if sex aneuploidy testing is offered, it should be made optional. Among those receiving a prenatal diagnosis of a sex aneuploidy, recent information indicated that 15% or fewer will terminate the pregnancy [45]. Additional patient information would be needed for those choosing to be screened for sex chromosome aneuploidies [46].

Screening for Monosomy X (Turner Syndrome)

Serum/ultrasound screening has found that increased first trimester nuchal translucency and/or generalized hydrops are strong markers for monosomy X (Turner syndrome). It is also possible to identify monosomy X via cfDNA testing. The first step in identifying the common sex aneuploidies is to first establish the fetal sex. One of the seven structured reviews [39] summarized the literature on cfDNA and the identification of fetal sex. Among 60 studies (11,179 tests), the detection rate was 98.9% (98.0–99.4) with a false-positive rate of 0.04% (0.02–0.11). Not all of the included studies used the methods described in this chapter, but the

results are consistent with the reports from cfDNA screening for sex chromosomes.

Table 36.5 provides data from four of the structured reviews [36, 37, 39, 41] regarding screening for monosomy X which is relatively common in the first trimester and the most severe sex aneuploidy being identified in the first trimester. A high proportion of these pregnancies will be spontaneously lost, and those identified via increased nuchal translucency/hydrops are at an even high risk of loss. There appears to be a trend toward higher performance as later studies are included, but the last review tended to focus on general population studies with incomplete follow-up which may explain the higher detection rates. Monosomy X is overrepresented in many studies because the focus on the first trimester. More importantly, many enrolled women were identified as "high risk" because of increased nuchal translucency.

False-positive rates for monosomy X can occasionally be caused by the maternal genotype. For example, as women age, an increasing proportion has a small percentage of cells that have lost an X chromosome, and these can lead to a false-positive cfDNA result [47]. In such phenotypically normal women, the loss of a portion of the X-chromosome signal on cfDNA testing would be attributed to the fetus and reported as fetal Turner syndrome. Follow-up fetal diagnostic testing would identify a euploid fetus, and eventual testing of the mother would show a low percentage mosaicism. For this reason, it is important that patient educational materials included information that warns the woman that testing may result in information about her genetic makeup.

Screening for Sex Chromosome Trisomies

The most common sex chromosome trisomies are 47,XXY (Klinefelter syndrome), 47,XYY (Jacob syndrome), and 47,XXX. Based on karyotypes of 34,000 newborns, the birth prevalences of these disorders are 1:576 and 1:851 males and 1:947 females [48]. There are few fetal losses of these trisomies during pregnancy [49], so the birth prevalence is a reasonable estimate of the first trimester prevalence. Thus, in a general pregnancy population of 100,000 pregnancies (assuming 51% male and 49% female fetuses), the expectations are for 89, 60, and 52 pregnancies with sex chromosome trisomies, respectively. Table 36.6 shows the results from four of the structured reviews with a grouped analysis of all three sex chromosome trisomies. Due to the number of included studies and the exclusion criteria, the first two reviews may provide the most reliable performance estimates of a 93% detection rate at a 0.14% false-positive rate.

Reasons for False-Negative cfDNA Screening Results

There are multiple reasons for false-negative cfDNA screening results. The following paragraphs provide an overview of the most common ones, roughly in order of their probability from high to low.

Relatively low fetal fraction A relatively low, but acceptable, fetal fraction (between 3% and 7%) results in a very small difference in the expected versus observed percentage of chromosome fragments from a given chromosome. If a sufficient number of fragments are not sequenced, this difference may not be identified, and the results will be incorrectly reported as a false negative. There is a well-known relationship between the fetal fraction and the final result (e.g., *z*-score) as shown in Fig. 36.2 [50]. The fetal fraction is on the *x*-axis and the *z*-score (number of standard deviations above the average in euploids on the *y*-axis). The small filled

Table 36.5 Summary of four structured reviews of monosomy X screening among cfDNA samples from singleton pregnancies with successful testing

Review	Studies	45,X	Euploid	DR (%)[a]	FPR (%)[a]
Gil 2014 [36]	12	139	4855	88.6 (83.0–93.1)	0.12 (0.05–0.24)
Gil 2015 [37]	16	177	9079	90.3 (95.7–94.2)	0.23 (0.14–0.34)
Mackie 2017 [39]	8	71	6641	92.9 (74.1–98.4)	0.10 (0.00–0.50)
Gil 2017 [41]	11	36	7676	95.8 (70.3–99.5)	0.14 (0.05–0.38)

45X monosomy X (Turner syndrome), *DR* detection rate (sensitivity), *FPR* false-positive rate
[a]Reported as point estimate and 95% confidence interval

Table 36.6 Summary of four structured reviews of sex chromosome trisomy screening among cfDNA samples from singleton pregnancies with successful testing

Review	Studies	SCT	Euploid	DR (%)[a]	FPR (%)[a]
Gil 2014 [36]	9	44	2802	93.8 (85.9–98.7)	0.12 (0.02–0.28)
Gil 2015 [37]	12	56	6699	93.0 (85.7–97.8)	0.14 (0.06–0.24)
Mackie 2017 [39]	5	11	5465	100 (71.5–100)	0.11 (0.01–0.24)
Gil 2017 [41]	8	17	5043	100 (83.6–100)	0.14 (0.05–0.38)

SCT sex chromosome trisomies (47,XXY, 47,XYY, and 47,XXX), *DR* detection rate (sensitivity), *FPR* false-positive rate (1-specificity)
[a]Reported as point estimate and 95% confidence interval

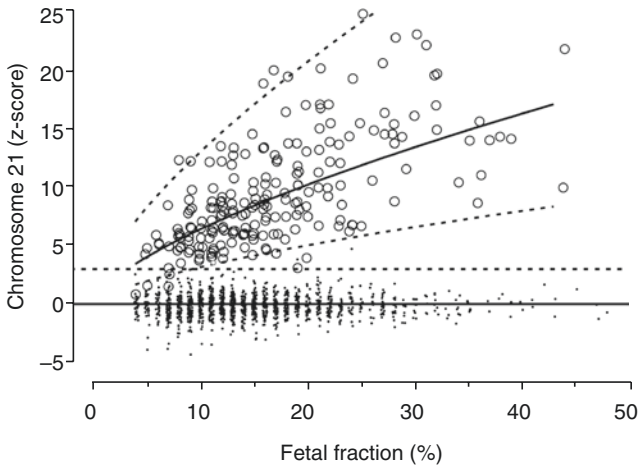


Fig. 36.2 Scatterplot of the fetal fraction versus chromosome 21 *z*-score in euploid (small closed circles) and Down syndrome (open circles) pregnancies. The horizontal dashed line is drawn at a z-score of +3; the screening cut-off

circles are the results in 1470 euploid pregnancies, and they tend to vary between −3 *z* and +3 *z*, centered at 0 *z*. The larger open circles represent the results in 212 Down syndrome pregnancies, most of which are above +3 *z* and considered screen positive. Note that as the fetal fraction decreases, the *z*-score in cases also decreases, indicating that false negatives are more likely with lower fetal fractions. Deeper sequencing is one way that it might be possible to reduce this reason for false negative results.

Confined placental mosaicism The primary source of "fetal" cfDNA is placental in origin, and this may be discordant with the fetal genome. A fetus could be aneuploid even though the karyotype of the placenta does not reflect that finding. In these cases, the cfDNA test is analytically correct (i.e., detecting those placental cells of the mosaicism which are euploid) but clinically incorrect (i.e., the fetus is aneuploid). This is recognized to occur for trisomy 13 and 18 but not trisomy 21 [51].

Maternal Copy Number Variants It is possible for a maternal deletion to cause a false-negative result [52]. However, this would be a much rarer event, as the fetus must be aneuploid and the maternal deletion would need to be relatively large and on the same chromosome.

Technical Issues Technical assay issues can make the identification of some aneuploidies more difficult. For example, the low guanine-cytosine (GC) content of chromosomes 18 and 13 renders the polymerase chain reaction (PCR) steps and subsequent sequencing counts less reliable. This results in lower detection rates than for other aneuploidies. Laboratories attempt to correct for this in their bioinformatics pipeline but are not always successful. There are also occasional sample mix-ups or other laboratory-related issues that could cause a false-negative test result.

Reasons for False-Positive cfDNA Screening Results

False-positive results are not concentrated within the group of samples, such as those with lower fetal fractions. Rather, they occur at any fetal fraction and are due to other factors such as those listed below.

Confined placental mosaicism The source of "fetal" cfDNA in the maternal circulation is placental cells. Thus, cfDNA testing provides test results that are directed to the placenta, which may be discordant with fetus. In these cases, the cfDNA test is analytically correct but clinically incorrect. Experience gained from chorionic villus sampling indicates that this may occur in up to 1–2% of pregnancies [51, 53–55].

Maternal mosaicism The cfDNA testing methods always assume the mother has a normal karyotype. This is not always the case, as shown earlier when screening for monosomy X [56]. A normal fetus/placenta might have a positive cfDNA screening test due to low level maternal mosaicism. Although less common, some women may have a full sex chromosome trisomy (e.g., 47,XXX) and appear normal [57]. Maternal mosaicism can be diagnosed by karyotyping peripheral blood lymphocytes. Educational materials and counseling about cfDNA testing for sex aneuploidies should include the possibility that a chromosome abnormality may be identified in the mother.

Deceased fetus A deceased fetus can cause a false-positive result if, for example, the deceased fetus was aneuploidy [58]. The placenta from the deceased fetus (which is also likely to be aneuploid) is intact and continues to shed DNA weeks after the fetal demise. The occurrence of an original twin gestation rather than a singleton gestation may not have been clinically apparent, if an early ultrasound was not performed, hence the term "vanishing twin." The SNP methodology can identify a vanished twin pregnancy through the remaining placental tissue. However, the interpretation of these patterns is complex, and laboratories using this method may report the results as a vanished twin, but not interpreted for aneuploidy.

Maternal cancer Rarely, cfDNA screening may be screen positive for more than one chromosome (e.g., screen positive for both Down syndrome and trisomy 18). This type of finding has been associated with maternal malignancy in which tumor DNA is shed into the maternal circulation [59–62]. In a series of over 125,000 cfDNA screening tests, more than

one common trisomy was reported in 39 (0.03% of tests) [59]. Discordancy between cfDNA test results and follow-up invasive fetal diagnostic testing was documented in 16 pregnancies (41%). Maternal cancer was subsequently diagnosed during or after pregnancy in 7 of the 39 women (18%) for a minimum rate of 1 in 18,000. However the rate could be higher as maternal follow-up was not available for all 39 women. Cancer types diagnosed included neuroendocrine, leukemia, lymphoma, colorectal, and anal.

Testing of cfDNA is not a viable screening test for maternal malignancy, given the paucity of data on this association, the potential for false-positive results, and the emotional and medical impact of such results on the patient's well-being. The appropriate clinical evaluation of such patients is unclear at this time. The most common malignancies in women of reproductive ages are cancer of the breast, cervix, ovary, colorectal cancer, leukemia, Hodgkin and non-Hodgkin lymphoma, and melanoma. Educational materials and counseling should include the possibility of a test result associated with a suspicion of maternal cancer.

Maternal copy number variants The methodology for cfDNA analysis assumes that every woman carries the same proportion of genetic material on each chromosome, but chromosomes vary slightly among individuals due to inherited or de novo copy number variants (i.e., deletion or duplication of a genomic region(s)). In these individuals, cfDNA sequencing might yield a positive result when the size of the maternal duplication was relatively large and it occurred on a chromosome of interest (e.g., maternal chromosome 21). In two studies, maternal duplications on chromosome 18 were the likely cause of trisomy 18 false-positive results in six of seven cases examined [63, 64]. Deeper sequencing (e.g., higher number of fragments sequenced) makes this form of false-positive results more likely.

Chance False-positive results can occur as a result of simple statistical chance. A common cut-off level for test results such as a z-score is +3. Therefore, 1 or 2 per 1000 euploid fetuses would be expected to have a false-positive result by chance alone. Where 100,000 tests were performed, an estimated 100 false-positive results would be expected.

Technical issues Rare sample mix-ups or other technical errors could lead to false-positive (or false-negative) test results. This is well-documented in many types of chemistry and molecular tests.

Transplant recipient If transplanted tissue (bone marrow or organ) was obtained from a male donor, cfDNA testing may incorrectly identify a female fetus as being male, due to the release of male cfDNA from the donor organ into the maternal circulation [65].

Recent blood transfusion Maternal blood transfusion from a male donor performed within the last 4 weeks prior to the blood draw for cfDNA may incorrectly identify a female fetus as being male [66].

cfDNA Test Failures

In the previous sections, the detection and false-positive rates were summarized for various disorders with the caveat that the cfDNA testing was successful. Those tests that were not successfully reported have been labeled as "indeterminate," "not reportable," "uninterpretable," or a "no call." Those requested tests that had an adequate sample and for which the testing process was started but there was no screen-positive/screen-negative result provided to the patient/provider will be considered a "test failure." The test failure rate is the third important characteristic of cfDNA screening tests, in addition to the detection and false-positive rates.

There are many reasons for test failures, but about half are due to the fetal fraction (or the quantity of fetal cfDNA being too low for reliable testing). In Fig. 36.2, for example, there are no results shown below a fetal fraction of 4% (x-axis). Results below 4% were considered test failures due to low fetal fraction. Not all laboratories measure and/or report fetal fraction. The philosophy of some laboratories is that cfDNA is a screening test with both false-negative and false-positive results expected. Thus, even if the fetal fraction were to be low, there is still a reasonable chance of detecting the few affected pregnancies in those samples, with a relatively low false-positive rate. This approach results in an important reduction in the cfDNA test failure rate at the cost of only a slight reduction in the test's detection rate and slight increase in false-positive rate. This tactic makes most sense when screening in a low-risk population. Alternatively, some laboratories adjust their quality parameters such that they minimize their false-negative rate at the expense of having more test failures. This is may be important when the cfDNA test is used as a secondary screening test (e.g., after a serum screen is positive) where a false negative is especially worrisome. One way to both reduce failures due to low fetal fraction while maintaining a high detection rates is to perform deeper sequencing. Unfortunately, this increases the costs of an already relatively expensive screening test, but as the cost of testing falls over time, this may become a more affordable option. Other reasons for test failures include insufficient number of aligned reads/SNP calls, laboratory issues (poor PCR, library quality), and special circumstances (dizygotic twins, donor egg, large regions of homozygosity, inappropriate fragment length).

Several of the structured reviews referred to earlier had summarized the test failure rates, but generally concluded that it was difficult to compare rates due to study design issues and differences in the populations and disorders included. Failure rates are strongly dependent on fetal fraction, and fetal fraction is dependent on several important variables that may also vary by study. These include maternal weight (obese women have up to a tenfold increase in test failures due to lower fetal fraction), early gestational age (samples collected at 8–9 weeks have a three- to fourfold increase in test failures), late gestational age (samples collected after 20 weeks have higher fetal fraction by more than 1% per week), twin pregnancies (fetus-specific fetal fractions are about half the total) and chromosome abnormalities (trisomy 18 pregnancies will have about half the normal fetal fraction; triploidy will have even lower fetal fractions), and what is being screened for (Down syndrome only, or all aneuploidies include those of the sex chromosomes). Published studies differ by all of these factors, including whether or not they even quantify fetal fraction for use in interpreting results. Given this, it is not surprising that the range of failure rates in these reviews is wide. In one review [37], rates of test failures were provided by 31 studies and ranged from 0% to 12.2% with 11 of those studies reporting the test failures that were solely due to low fetal fraction (range 0.1–6.1%).

Another important factor that is generally overlooked in assessing cfDNA test failure rates is that some published studies included repeat testing. Repeat testing occurs when testing of the initial sample fails, but a new blood sample is then tested. Testing a new blood sample (either a duplicate samples collected at the same time as the initial sample, or a subsequent sample collected after the initial test has been reports) results in an important reduction in test failure rates. This reduction ranges from about 50% [67–69] to 80% or more [70–72], among those having a repeat sample tested.

In the USA, it is reasonable to expect that the test failure rate on the initial sample is perhaps 1% but may be as high as 5%, depending on the test method, how many disorders are included, and the maternal weight and gestational age of the population being tested. Among the pregnancies that had a failed initial test, there appears to be no increase in the risk for Down syndrome, but there may be a doubling in risk for trisomy 18 and trisomy13. The American College of Obstetricians and Gynecologists [73] suggests that women with an initial test failure due to low fetal fraction be offered a repeat test (when allowed). Those with a test failure should be counseled and offered ultrasound and diagnostic testing. In a high-risk setting, the offer of diagnostic testing is already warranted, and repeat testing will delay diagnosis and possibly reduce reproductive options. In a general pregnancy population where the risks are much lower, even among test failures, repeat testing and high-resolution ultrasound may resolve many of the issues. However, the gestational age and related reproduction options must still be considered.

Mosaicism

Identifying fetuses mosaic for the common aneuploidies will be more difficult than for a fetus with a full aneuploidy. An additional complication is that the cfDNA methods rely on fetal DNA derived from the placenta, and, therefore, confined placental mosaicism could be identified, even if the fetal chromosomes are normal. This is a long-standing issue that also complicates the interpretation of CVS test results [74]. To help understand screening for mosaicism, consider the effective fetal fraction when a mosaicism for chromosome 21 is 50% and the overall fetal fraction is 10%. The mosaicism would indicate that about half of the fetal DNA would be derived from cells with three chromosomes 21 and half from euploid cells. Thus, the effective fetal fraction in such a pregnancy is only 5%, not 10%. This computation is complicated by the potential heterogeneity of the placenta, which would be difficult to quantify. In many instances, the lower the rate of mosaicism, the more likely the effective fetal fraction will fall below the level at which cfDNA testing is effective. This would likely result in a screen-negative cfDNA test result.

One study [75] reported on three mosaic karyotypes involving the common trisomies. Two were mosaic for chromosome 21 (29% and 44% mosaicism) and one for trisomy 18 (89% mosaicism). No test results or fetal fractions were reported, but all three were reported as being detected. A second study [76] identified five relevant mosaic karyotypes involving chromosomes 21, 18, and 13. The median fetal fraction was 13% (range 6–23%), and mosaicism ranged from 10% to 50%. The effective fetal fractions ranged from 0.6% to 10%. As expected, the two mosaics with very low effective fetal fractions were identified as euploid, while one of the two with effective fetal fractions around 5% was detected. The lone sample with a high effective fetal fraction of 10% (45% mosaicism × 23% fetal fraction) was a trisomy 18 mosaic superimposed on a full trisomy 21 karyotype. The result was positive for Down syndrome, but the *z*-score for chromosome 18 was unremarkable. Mosaicism is even more of an issue with screening for sex chromosome aneuploidies. The placenta is more likely to be mosaic and may not reflect the fetal sex chromosome makeup [74].

In summary, maternal plasma cfDNA testing in its current implementation will likely identify most high-level mosaicism when the fetal fraction is average or above average. However, it will likely not identify most of the lower level mosaicism, especially when the fetal fraction is below average. In the future, deeper sequencing would likely allow for a higher proportion of mosaic pregnancies to be identified. As

with all positive cfDNA testing for aneuploidy, any woman with a positive cfDNA screening result should be offered an invasive procedure to allow for a definitive diagnosis.

Testing in Specific Subgroups

IVF Pregnancies

Some studies explicitly excluded pregnancies achieved by in vitro fertilization (IVF), whereas others did not document IVF status during enrollment. One study [75] reported that 17 of 38 IVF pregnancies undergoing cfDNA screening had chromosomal abnormalities (all correctly identified) and that 21 were euploid (all correctly identified). No mention was made of the fetal fraction in this population or whether the normalized chromosome value (similar to a *z*-score) differed from non-IVF pregnancies.

A more comprehensive study [11] identified a subset of 632 tested women from seven enrollment sites that had information available about the use of assisted reproductive technologies (ART). Among euploid pregnancies, the 33 ART pregnancies did not differ from the 599 naturally conceived pregnancies in gestational age, maternal weight, maternal age, total cfDNA, fetal cfDNA, or fetal fraction. As expected, the mean chromosome 21 *z*-score was close to 0 (−0.13) in the naturally conceived pregnancies but was significantly (p = 0.048) higher among the ART pregnancies. Both the chromosome 18 and chromosome 13 average *z*-scores were also elevated, but only the chromosome 18 *z*-score was statistically significant (p = 0.0032). No differences in demographic or fetal fraction measurements occurred among the 10 ART and 63 naturally conceived Down syndrome pregnancies. However, the average chromosome 21 *z*-score for the ART cases was lower (8.7 versus 11.4, p = 0.14). This is an intriguing finding, given that these women seek alternatives to invasive testing that maternal plasma DNA testing provides, but these findings need to be confirmed. A more recent study [77] focusing on fetal fraction studied 10,698 singleton and 438 twin pregnancies with IVF rates of 9.5% and 56.2%, respectively. They found that IVF was associated with a highly significant reduction in the fetal fraction and in the failure rate (multivariate odds ratio 6.5, $p < 0.001$). Modeling showed the IVF effect to be strong and independent of the multiple gestation effect in reduced fetal fractions.

Twin Pregnancies

Multiple gestations are now about three times more common due to assisted reproductive technologies. The current twin pregnancy rate in the USA is 1:30 (http://www.cdc.gov/nchs/fastats/multiple.htm). Twin pregnancies discordant for a trisomy would likely be more difficult to classify correctly as, on average, only half of the fetal fragments would be derived from the placenta associated with the affected fetus. Thus, at a fetal fraction of 10%, a discordant twin pregnancy may have an effective fetal fraction of only 5% (perhaps as low as 2% or as high as 8%). There is some evidence that the total fetal fraction is higher in twin pregnancies than in singleton pregnancies, but any difference that exists is likely to be relatively small. Among 17 euploid twin pregnancies, 5 twin pregnancies discordant and 2 concordant for Down syndrome, and 1 pregnancy discordant for trisomy 13, the fetal fraction was 18.1% in twins and 13.4% in over 1500 euploid singleton pregnancies [78]. When twins are concordant for a trisomy, it is likely that testing will have similar performance to that in singleton pregnancies. None of the cfDNA testing methodologies included in this chapter can distinguish a monozygotic twin pregnancy from a singleton euploid pregnancy.

Among the seven structured reviews summarized earlier in this chapter, three provided results from cfDNA screening in twin pregnancies. These are summarized in Table 36.7. Although there are far fewer data for twin compared to singleton pregnancies, the existing data do show relatively high detection rates among twin pregnancies. As stated earlier, however, there is likely to be a higher test failure rate among these twin pregnancies that effectively reduced the actual detection rate achievable. There is little or no data on grand multiple gestations and cfDNA screening.

Table 36.7 Summary of four structured reviews of Down syndrome detection among cfDNA samples from twin pregnancies with successful testing

		Twin pregnancies			
Review	Studies	DS	Euploid	DR (%)[a]	FPR (%)[a]
Gil 2014 [36]	4	18[b]	209	94.4 (74.2–99.0)	0.00
Gil 2015 [37]	5	31[c]	399	93.7 (83.6–99.2)	0.23 (0.00–0.92)
Gil 2017 [41]	5	24[d]	1111	100 (95.2–100)	0.00

DS Down syndrome, *DR* detection rate (sensitivity), *FPR* false-positive rate (1-specificity)
[a]Reported as point estimate and 95% confidence interval
[b]Two trisomy 18 pregnancies were also correctly classified as screen positive
[c]Nine trisomy 18 and two trisomy 13 pregnancies were also correctly classified as screen positive
[d]Thirteen of 14 trisomy 18 and none of 1 trisomy 13 were also correctly classified as screen positive

Will cfDNA Testing Ever Be Diagnostic for Common Aneuploidies?

While cfDNA screening for aneuploidies is maturing, even more will be learned over the next few years as more samples are tested and the number with carefully documented outcomes increases. Detection and false-positive rates may improve for trisomy 18 and trisomy 13, as well as for the sex chromosome aneuploidies. In addition, the range of fetal disorders that could be identified will likely continue to grow. However, all of these cfDNA tests are using DNA of placental origin. Thus, even with improved cfDNA testing, the result would be a reflection of the genome of the placenta, not the fetus. To whatever extent the embryonic and extraembryonic tissues differ cytogenetically, there will remain an issue with false-positive cfDNA test results Thus, cfDNA testing will never be as accurate as a second trimester amniocentesis and diagnostic testing via karyotype or CGH array. In addition, there will likely always be infrequent causes of false-positive test results such as maternal cancer or a mother who is a low-level mosaic monosomy X [62]. While these findings represent a true pathological genomic state, the purpose of the testing was to identify fetal chromosome abnormalities, and in that context, the finding is a false positive. Another example of a false positive shows the problems with interpreting other autosomal chromosomes. In one study [79], a trisomy 22 was identified in a second trimester cfDNA test. Since fetal trisomy 22 is uncommon this late in pregnancy, the potential for a confined placenta mosaicism must be considered. The woman did not want an invasive procedure and delivered a healthy baby. Three placental samples were karyotyped, and each confirmed the trisomy 22 finding. Other rare reasons for false-positive fetal results have also been reported [80]. Thus, for the foreseeable future, amniocentesis and diagnostic testing should be offered to all women with a positive test result. This may not always remain the case for other application of cfDNA in prenatal care. Significant differences between embryonic and extraembryonic tissues to date have been limited to chromosome copy number, and there is no reason to expect dominant Mendelian gene mutations to exhibit issues with false-positive test results.

Detection of Deletions/Duplications

Genomic copy number variation (CNV) at the subchromosomal level in the form of microdeletions and microduplications is also of clinical concern, and several groups have already shown the ability to identify selected deletions and duplications using cfDNA [81–83]. All of these required deeper sequencing and/or adding targeted areas for sequencing. Copy number variation is extremely common in healthy populations, and ongoing research [84] is helping to define which deletions/duplications are pathogenic and which are benign. Specific microdeletions are associated with a number of well-defined clinical genomic syndromes, such as DiGeorge syndrome (22q11.2), Cri-du-chat syndrome (5p), 1p36 deletion syndrome, Prader-Willi/Angelman syndromes (15q), Jacobsen syndrome (11q), and Langer-Giedion syndrome (8q). Individually, these disorders are generally rare and often are associated with incomplete penetrance. The range of phenotypes can vary from severe to very mild or benign. In addition, the size of the CNV associated with each of these syndromes is quite broad, and the vast majority are small enough to have escaped detection by conventional karyotyping. Given the high level of genetic, phenotypic, and epidemiologic variability associated with smaller CNVs, substantially more work is required to validate cfDNA testing for these microdeletions, and each of these syndromes needs to be individually evaluated. More care is required in the design of effective assays to identify these more subtle genomic lesions, and thus there may be higher costs associated with their inclusion, relative to the larger and simpler whole chromosome aneuploidies. In order to reduce the false-positive rate and improve the PPV, initial test offerings have included subjecting initial screen-positive samples to a second sequencing run at increased depth. This has been shown to greatly reduce false-positive results and improve the PPV [85]. Because of these complexities, the return on investment to develop comprehensive CNV tests can be expected to be low compared with screening for common aneuploidies, and the rationale for their inclusion in screening programs is currently questionable.

If testing is not targeted to specific known duplications/deletions, the issue of variants of unknown clinical significance will need to be addressed, as it already is with aCGH interpretation [84]. In contrast with the more common aneuploidies, these determinations require the professional judgment of a trained clinical cytogeneticist and will be harder to justify within the context of prenatal population screening. Routine prenatal screening should not include strategies for identifying variants of unknown significance. Validation of laboratory-developed tests for these disorders will be difficult, given the relatively rare nature of each.

Low-Level Whole-Genome Sequencing of the Fetus

Laboratories using the shotgun methodology to identify common trisomies have sequencing information available for all of the autosomes, albeit at a relatively low depth of sequencing. Some laboratories are offering a separate test that includes a whole-genome interpretation that can identify less common autosomal trisomies (e.g., trisomy 16, trisomy

22) as well as large deletions and duplications across the genome (www.sequenom.com/tests/reproductive-health/maternit-genome; www.bgi.com/global/wp-content/uploads/sites/3/2017/04/NIFTY-brochure-0515.pdf). Since the interpretation is dependent on the depth of sequencing, such testing may include far more aligned reads that would be needed for the interpretation of common trisomies, increasing test costs. Whether the resulting information is of sufficient quality and utility remains an open question.

High-Level Exome/Genome Sequencing of the Fetus

Research groups have applied multiple methods allowing for whole-exome/whole-genome sequencing of the fetus [86–88] that is of similar sophistication as that offered to individuals. At this time, however, the resources needed to undertake such an effort appear to far outweigh any direct benefit of this technology. One study [88] utilized 900,000 SNPs over the genome and accumulated 3.9 billion reads, equivalent to an average 65-fold coverage of the genome. This was done for both the mother and biological father, in order to infer the genotype of the fetus. Another study sequenced to a depth of 270-fold and identified 85% of fetal de novo mutations with a PPV of 75% [89]. When whole-genome/whole-exome sequencing becomes more commonplace among adults, and the technology becomes less expensive, there may be pressure and interest to perform such testing in the prenatal setting. Studies are already underway to determine whether such testing is warranted as part of newborn screening [90].

Social and Ethical Issues

One issue of ethical and social concern that represents a significant shift in the development and delivery of clinical genetic testing is the externalization of the academic genetic center's voice from the creation and validation of patient and provider educational materials, service delivery, and pricing of laboratory services. While neither the commercial nor the academic sector is completely free from financial conflicts, there is an implicit fiduciary responsibility that falls to the academic practitioner and to physicians and clinicians more generally, to place the interests of the patient above selfish interests. This cannot be said to be true for commercial testing where the market serves as the prevailing force that guides decision-making. The fiduciary responsibility of the commercial laboratory lies with the shareholders and investors. What constrains pricing for the commercial entity is simply market competition. In contrast, while market forces also influence pricing within the professional community, protecting the interests of patients, both individually and in the collective, is a balancing force that has been important in the development and delivery of genetic services. What is seen more fully in cfDNA prenatal screening than in previous testing is the disappearance of the academic laboratory perspective. This is not to speak badly of either commercial or academic enterprises as both ultimately are engaged for the good of patient care, albeit in different ways and with different underlying motivations. However, the shift is likely to become evident in patient and provider education, test pricing, and delivery of services unless these elements are addressed jointly with the academic genetics community.

Cell-free DNA represents a major change in prenatal screening practice in three important ways. First, all prior tests (maternal age, serum markers, ultrasound markers) have been based on phenotypic correlations. In contrast, cfDNA is based on genotypic measures derived directly from the chromosome abnormalities that cause Down syndrome and other chromosomal syndromes. This accounts at least, in part, for the original framing of cfDNA more as a diagnostic test than a screening test. This difference between cfDNA and previous tests is not merely a quantitative matter of superior statistical performance, it is a fundamental difference in its derivation from the causal root of the disorders, except that the source of cfDNA is extraembryonic (i.e., derived from the fertilized egg but in a placental compartment separate from the actual embryo), testing by this methodology might otherwise perform at a diagnostic level. However, as this source of tissue is naturally tolerant of chromosomal variation not found in the embryo, it likely would not always be reliable as a diagnostic test. This same phenomenon of false-positive findings due to extraembryonic chromosomal variation is observed in chorionic villus sampling, for exactly the same reason, and yet we have considered CVS to be a diagnostic test and not a screening test.

Second, cfDNA represents a disruptive innovation in the classic sense. It is a break from the iterative evolution of biomarkers correlated with specific birth defects using standard laboratory methodology and calculated risks and the beginning of high–throughput DNA analysis with sophisticated bioinformatics. Christensen defines disruptive innovations as products that enter the market initially as crude alternatives to some prevailing technology that only meets the least demanding market needs, and that over time, refinement of the newer technology overtakes the prevailing technologies and displaces them altogether [91, 92]. The cfDNA technology required refinements in both DNA analytic methodologies and technologies and in the bioinformatics algorithms used in interpretation to become successful.

Third, cfDNA also represents a disruption in the clinical workforce needed to reach diagnostic closure. There exists a highly integrated team of physicians, laboratorians, and genetic counselors surrounding multiple marker screening that is tuned to prevailing practices of educating patients, collecting and shipping specimens, generating reports, communicating results with referring physicians, counseling screen-positive patients on risks and diagnostic options, con-

firming pregnancy dating for screen-positive cases, imaging for ultrasound clues, providing invasive diagnostic procedures, and analyzing karyotypes using cytogenetic procedures. Transition to cfDNA stresses the existing system as it greatly reduces the need for many of these services and drastically alters others. This produces a dramatic reduction in clinical reimbursements and a simultaneous need for reorganization of the communications and educational services surrounding prenatal testing. The shift away from the academic health center or prenatal services unit toward a commercial testing laboratory potentially separates the clinical expertise needed to redesign the patient care environment from the resources required to sustain it.

Testing in the General Population

The initial clinical validity studies provided complete ascertainment of fetal status and generally occurred in "high-risk" populations. This was because it was necessary to compare the true genotype of the fetus via genetic information only available after an invasive procedure such as an amniocentesis or CVS with the results obtained from cfDNA screening. Such procedures could not be justified in a general population for reasons of both iatrogenic risk and cost. Consequently, low-risk populations were not amenable to studying clinical validity. Because of this, several professional organizations [93–97] initially concluded that cfDNA testing should not be applied to a general pregnancy population due to lack of sufficient information.

Several groups have addressed the issue of whether there are any known factors that might indicate that the detection/false-positive rates of cfDNA testing of maternal plasma might differ in a general population versus a high-risk population. Table 36.8 lists potential differences in these two populations, along with studies that provided relevant information.

There are, however, other important differences between offering the test in a high-risk setting and a general population setting. Women classified as being at high risk are generally aware of their risk status and are likely to have already been referred for genetic counseling. In this setting, they can be provided additional information and have their questions answered. Also, most women screen positive by serum/ultrasound testing have Down syndrome risks in the range of 1 in 250 to 1 in 10. Only a very small proportion might have risks exceeding 1 in 2. This allows for more objective decision-making by focusing on the fact that a 1 in 100 risk means that 99 of 100 outcomes are not Down syndrome. Thus, the high-risk group has more access to information at a "teachable moment" without extreme risks that may complicate decision-making. Contrast this with a woman in the general population being offered cfDNA testing for whom there is little reason to acquire specific knowledge related to these chromosomal disorders. This is one of many options she is offered, and she is unlikely to be focused on the potential impact of the testing results. If the woman were to be screen positive on this test, her risks will be very high compared even to current group of "high-risk" women. This will almost certainly create tremendous anxiety that must be dealt with by the care providers quickly. It is also likely to put the couple in a difficult position to make decisions due to the high assigned risk. It might also be confusing as to whether this is a diagnostic or screening test. Despite the clarity of this argument, the near diagnostic performance of cfDNA for unanticipated and unfamiliar conditions is no different than that provided by most newborn screening tests and many other genetic testing for inherited diseases. The keys to appropriate patient care are a well-informed cadre of healthcare providers, educational materials for patients, and a program in place to provide comprehensive care to the women with screen-positive results.

Table 36.8 Potential differences between a "high-risk" population and a general population that might impact cfDNA test performance

Factor	Potential impact	Findings
Prevalence	A high-risk population will have a higher prevalence of the disorder	Although true, the higher prevalence, by itself, will not impact the detection or false-positive rates of the test. However, the predictive values (both positive and negative) will change depending on prevalence. Fortunately, this change is well described, and screening programs deal with varying prevalence as part of current prenatal screening
Fetal fraction	A high-risk population may have a higher fetal fraction leading to higher detection rates	After review of many factors (e.g., maternal age, abnormal ultrasound findings, serum marker levels), only maternal weight and some serum markers may differ. The impact is likely to be very small and is sometimes in the wrong direction (low PAPP-A associated with low fetal fraction would reduce performance in high risk, but not low-risk settings) [50, 98–101]
Maternal age	A high-risk population is older and this could impact test performance	The only known predictor of improved test performance is increasing the fetal fraction. Maternal age alone does not directly impact the test's final measure (e.g., *z*-score) [50, 98, 99]
Other test indications	Those with a positive family history or abnormal ultrasound might be easier to detect	All studies stratifying results by indication find no differences in detection or false-positive rate or, when examined, the test statistics such as *z*-score [50, 98, 99]

One recent study was undertaken to determine whether primary obstetrical care providers could transmit the needed patient information and counseling prior to cfDNA being used as the "first-line" screening test for Down syndrome [102]. This process-oriented study included no charge for testing (to simulate a setting where cfDNA testing is routinely covered by insurance), provider education, patient education materials, and a local and knowledgeable screening program that oversaw the entire screening program. Uptake of cfDNA screening during the study was higher than traditional serum/ultrasound screening in the previous 6 months. Screening performance met expectations with 0.6% screen positive with a 75% positive predictive value. Patient surveys showed good, but not perfect, knowledge related to screening, while provider surveys showed a prepared and knowledgeable staff that felt such screening could easily fit into their routine.

Educational Materials

As part of the introduction of any new prenatal screening test, education of the providers and patients is of paramount importance. There is increasing recognition of the potential for confusion and even harm to the patient that may arise from uncertainty in genetic testing. The primary difference between patient education and marketing lies with the intent and motivations of the writer. Patient and provider educational materials are intended to provide the patient and her provider with objective information that empowers decision-making. In contrast, marketing materials are intended to persuade the client to choose testing being offered by the company. As genetic medicine has become big business, the potential for these lines to become blurred has grown. Similarly, as the need to generate revenue has increased at academic centers, the potential for crossing the lines that separate marketing and education has also intensified. Consequently, careful attention must be paid not only to the accuracy of the information that is presented but also to the underlying motivation of the writer.

Most, if not all, laboratories that currently offer testing have both provider and patient information on their websites. Professional societies have recommended specific content that should be covered in patient materials [93–97]. This content is summarized in Table 36.9. We are unaware of any patient materials currently available from commercial companies offering cfDNA for common trisomies that include all of the content recommended by these societies. Companies should make clear to their clients and sales force whether materials that are provided are intended for marketing or for education. Since one is intended to empower, and the other is intended to persuade, it is not sufficient to say that because the information is accurate, it can be used for either purpose. We would understand the motivations of the professional societies to be, at least in part, a reaction to the blurring of marketing and education ideals in the way that informational materials have been developed and distributed.

Patient educational materials are recommended to be written at about reading grade level 8 in order to be understood by the broadest population of patients. The majority of referring physicians, although well-qualified in their disciplines, will be unfamiliar with the complex principles of genetics and genomics. Initially, both patients and providers are likely to carry incorrect assumptions related to previous screening modalities (e.g., neural tube defect screening can be accomplished with the same specimen being used for aneuploidy screening), and writers of educational materials for cfDNA

Table 36.9 Recommended content for patient educational material as suggested by professional organizations

ACOG	ISPD	SOGC	NSGC	ACMG	Recommended content
				X	An introductory statement about the purpose of testing
			X		Accurate and up-to-date information about the possible test results
X			X		Accurate and up-to-date information for available follow-up testing
X	X	X	X	X	The implications of a positive DNA test result
X	X	X	X	X	Explains that false-positive results can occur and that there is a need for confirmatory testing
X	X				States that high Down syndrome risks will occur with positive DNA test results
	X			X	Explains the potential stress associated with the extended wait for test results
	X	X		X	Contains information that the test results may not be informative for some patients
X	X	X	X	X	Provides information that amniocentesis/chorionic villus sampling would still be indicated in order to diagnose other disorders that the DNA test is not designed to detect

ACOG American Congress of Obstetricians and Gynecologists, *ISPD* International Society of Prenatal Diagnosis, *SOGC* Society of Obstetricians and Gynecologists of Canada, *NSGC* National Society of Genetic Counselors, *ACMG* American College of Medical Genetics and Genomics

should attempt to anticipate and address such assumptions in the context of recent history and conventions.

Cost and Value of cfDNA Screening for Down Syndrome

Tracking the actual costs of cfDNA for patients and their insurers is challenging. The list price for cfDNA screening tests for aneuploidy and the reimbursement received for it are often very different. List pricing may be in excess of $2000 per test, and there is great variability and uncertainty surrounding reimbursement, patient out-of-pocket expenses, and co-payments. In addition, there are variations in the complexity of testing and reporting in cfDNA ranging from simple and relatively inexpensive basic testing for common autosomal trisomies to complex and expensive options that provide information about a broader range of chromosomal conditions including sex aneuploidies, select microdeletions, and whole-genome reporting.

One question that must be addressed when considering cost and value is the primary purpose of screening. Down syndrome has a high prevalence, a moderate to severe phenotype and lack of effective therapy. These factors are combined with the relatively long life expectancy for Down syndrome and a relatively late maternal age profile. In many families, the child would outlive the parents. These factors resulted in Down syndrome being singled out for routine prenatal screening. Other autosomal trisomies are more rare and with a substantially shorter life expectancy. Sex chromosome aneuploidies, while common, are comparatively mild disorders. Other conditions, such as triploidy, are lethal and/or very rare, and they have not merited the development of stand-alone screening tests. However, to the extent that these other conditions may be identified in the course of screening for Down syndrome or neural tube defects, there is some clinical utility to their identification, and, consequently, they have been included in screening test offerings for Down syndrome. If Down syndrome remains the primary concern in pregnancy, as can be argued, then decisions regarding the development and implementation of cfDNA as a screening test should focus first on performance around that disorder. The extent to which other testable conditions should be included should be based on the clinical utility and impact to the overall screening performance.

Using detection and false-positive rate estimates for selected serum/ultrasound and cfDNA testing provided earlier in this chapter, the distribution of autosomal trisomies for the general population described in Fig. 36.1, cost for amniocentesis/CVS of $2000, estimated cost of quadruple and integrated screening of $200 and $500, the overall costs of multiple serum/ultrasound screening can estimated. This can then be compared with cfDNA screening using costs of $300 and $1000 per test. In the US health system, there is wide variability between the costs for different patients depending upon whether or not they are insured, and by whom. We also assume that all positive tests result in diagnostic testing with either CVS or amniocentesis and that costs for ultrasound and counseling are included in the procedural charge for invasive testing.

- For the second trimester quadruple marker test, the total cost of screening a population of 100,000 pregnancies for only Down syndrome would be $30.5 M, with 5247 amniocenteses, 66 missed cases of Down syndrome, and an average cost per case detected of $116 K. Including trisomy 18 in screening, the cost increases to $32.6 M with a decreased cost per case detected of $110 K.
- Using the integrated test, the total cost of screening of a population of 100,000 pregnancies for only Down syndrome would be $56.6 M, with 3286 amniocenteses, 33 missed cases of Down syndrome, and an average cost per case detected of $191 K. Including trisomy 18 and 13 in screening, the cost increases to $63.6 M with a decreased cost per case detected of $172 K.
- For cfDNA screening in the first trimester for only Down syndrome, (assuming a cost of $300 per test), the total cost of screening and diagnostic testing would be $30.9 M with 427 CVS procedures, 2 missed cases of Down syndrome, and an average cost per case detected of $94 K. Including trisomy 18 and 13 in screening, the cost increases to $31.4 M with 710 CVS procedures and 6 missed cases (two of each syndrome), at a decreased cost per case detected of $76 K.
- For cfDNA screening in the first trimester for only Down syndrome (assuming a cost of $1000 per test), the total cost of screening and diagnostic testing would be $100.9 M (with no change in invasive procedures or missed cases) at an average cost per case detected of $308 K. Including all three syndromes in screening, the cost increases to $101.4 M at a cost per case detected of $256 K.

In comparing different modalities for prenatal screening, there are two economic variables to use as a reference: the total cost of screening and the cost per case detected. The dominant factor in the cost of all three modalities (quadruple, integrated, and cfDNA) is the cost per screening test as this cost is incurred for all women opting for screening. The cost of diagnostic testing accounts for 33% of the cost of quad screening, 10% of the cost of the integrated test, and less than 5% of the total cost of cfDNA (depending on the price per test). The cost of screening for multiple conditions adds little to the overall cost in each case and is offset in the cost per case detected.

The differences in both overall cost and cost per case detected between quadruple marker testing and integrated testing are striking, with substantially higher costs for both using the integrated test. In order for cfDNA to match quad testing on cost per case detected screening for all three autosomal aneuploidies together, we calculate that the cost per test would need to be set at $442. If the integrated test were used as the reference on a cost per case detected basis, the price of cfDNA would need to be set at $700 per test to match on cost per case detected.

Alternatively, if matching were to be done on total cost of screening, rather than cost per case detected, cfDNA would need to be set at $311 per test to match the total cost of quad screening or $622 per test to match the total cost of the integrated test. Some value in-between that reflects the proportions of patients that choose between these two alternatives would seem reasonable. Either way, the collateral noneconomic benefits of fewer invasive tests and procedure-related loss of unaffected pregnancies, reduced patient anxiety due to avoidance of false positives, and the reduced number of missed cases of trisomies are considerable. Our calculations on costs, although based on a simpler set of assumptions, are in general agreement with the literature [103].

Finally, for the purpose of establishing priorities for prenatal screening, and given that the historical focus of prenatal screening has been directed toward patient concerns regarding Down syndrome, it seems a sensible approach to think of cfDNA first from the perspective of Down syndrome. However, some modestly increased costs are justifiable to add other chromosomal conditions that are rarer, less severe, or for which life expectancy is much shorter. A simple and inexpensive cfDNA configuration that is optimized around Down syndrome detection may actually reduce the overall cost of screening in a general pregnancy population. If the cost of DNA analysis continues to fall as expected, more complex configurations that permit simultaneous detection of microdeletions, triploidy, and lethal trisomies could be introduced, provided that the detection and false-positive rates and positive predictive values of testing are comparable so as to restrain false positives and negatives.

Given what is now known about the performance of cfDNA screening in the general population, the long-standing standard of offering Down syndrome screening to women of all ages appears reasonable. This is in light of the recommendations of ACMGG [66] and ACOG [73] acceptance of cfDNA screening for the general population. The primary remaining obstacle to implementing population-wide cfDNA screening for aneuploidies would appear to be cost. With sequencing costs expected to fall, and with the economies of scale associated with general population screening, affordable cfDNA-based screening should be achievable. Such testing can outperform other available modes of Down syndrome screening during pregnancy not only in terms of cost and cost per case identified but also in the avoidance of unnecessary invasive tests and missed cases.

External Proficiency Testing

Validating a laboratory developed next-generation sequencing test for cfDNA to identify common aneuploidies is a difficult undertaking requiring significant expertise and infrastructure. The field is also complicated by intellectual property issues [104]. External proficiency testing and oversight of these tests are needed to help ensure reliability and quality and to help harmonize the practice surrounding the clinical implementation of cfDNA screening. Currently, this testing is offered by approximately 10 laboratories in the USA, with up to 100 or more worldwide. All of the tests in the USA are laboratory-developed tests (LDTs), meaning that no two are exactly the same and variations in DNA extraction, test methodology, quality parameters, and other factors make each test unique. In 2017, the UK National External Quality Assessment Service reported their results of a pilot proficiency testing program distributed to 46 European laboratories [105]. They excluded laboratories using the SNP pattern-matching methodology as the artificial samples distributed were known not to be suitable. Two laboratories did not respond, six were not able to obtain results on any of the three samples, and two obtained results on only two samples. Among those reporting results, there were three errors (both false negative and false positive). The conclusion was that method-independent materials and/or patient material would need to be developed for universal external proficiency testing.

In 2013, the College of American Pathologists (CAP) added specific CAP checklist questions that address cfDNA screening for common trisomies. This and other checklist questions have been used routinely for CAP inspections of these laboratories for over 4 years. In 2018, CAP will be offering the "NIPT" survey that will include three challenges consisting of real patient plasma samples twice a year to enrolled laboratories for use in external proficiency assessment. This, coupled with on-site laboratory inspections by experts, is the best option for ensuring consistent and high-quality cfDNA screening for common trisomies.

Population Implications

Given the extraordinarily high detection rates, low false-positive rates, and strong rationale for application within the general population, it should not be surprising that concerns have been raised, especially within the Down syndrome population, regarding the potential elimination of Down syndrome and other genomic disorders within society [106]. As these concerns are vetted, it will be important to take stock of

the many different perspectives already present within the society regarding disabilities and how they are understood and to maintain a supportive environment in which patients and families are free to choose how best to apply this emerging technology within the context of their personal values. In the past, screening programs reported that one-quarter or so of pregnant women do not choose to be screened. Among those choosing screening who receive a positive result, another 25% or more choose to not have diagnostic testing. Among those with a definitive diagnosis of Down syndrome, 20% to 50% choose to continue their pregnancy. This "opting-out" provides some reassurance that the education/counseling is nondirective and women do make informed choices.

Global Perspectives on Ethical and Social Issues

Recognizing that ethical and social perspectives are closely tied to cultural norms, an international, interdisciplinary conference was organized to look beyond "Western" moral assumptions toward a more global perspective on cfDNA [107]. The authors of the resulting report identified eight topics for further consideration and discussion related to the implementation of cfDNA around the world. These included variations around the limits of autonomy and patient decision-making in the context of gender roles, the implications for sex selection and "family balancing," socioeconomic and cultural influences on thresholds for disability, differences in disease prevalence with respect to priorities for screening, sensitivity to differences in economic systems and capabilities, the influence of commercialization on the ethical delivery of healthcare, and the need for understandable informational resources.

Conclusion

Sequencing of cfDNA derived from maternal circulation has recently emerged as a disruptive innovation in prenatal screening that promises to replace phenotypic-based methods of screening for autosomal trisomies. It is dramatically more efficient and effective as nucleic acid-based screening derives directly from causal genomic imbalances. Both the detection and false-positive rates for cfDNA screening for common trisomies approach, but cannot equal, diagnostic testing performance. Cell-free DNA enters maternal circulation through the placenta and directly reflects the genomic constitution of extraembryonic tissues. False-positive results most often arise from differences between the embryo and extraembryonic cells. Advances in DNA sequencing and bioinformatic algorithms have resulted in several different laboratory approaches to reliably detect select genomic imbalance in the fetus. The primary focus for cfDNA, as with the biochemical tests that preceded it, derives from patient concerns surrounding Down syndrome. Accordingly, optimizing screening will require that the many other disorders that are also detectable be kept in perspective with the historical norms of medical screening. Clinical studies affirm the high reliability of cfDNA and support the rationale for extending screening to low-risk populations, provided that costs can be contained. As with essentially all powerful emerging technologies, there are a number of social and ethical questions that arise from its implementation, particularly those related to potential disruptions within a complex system of care, tension between patient/provider education and marketing, contrasting values and perspectives between academic medicine and commercial enterprise, and a set of issues involving differences in personal values held by the patients for whom screening is being offered. These issues may play out differently against differing cultural and social norms around the world. Future clinical applications can be expected to arise from research already underway surrounding the detection of more subtle and rare genomic disorders, simpler applications designed to reduce cost, extension into single-gene disorder screening or diagnosis, whole-genome/whole-exome sequencing, cancer screening, and applications beyond genetic and genomic disorders.

References

1. Wald NJ. The definition of screening. J Med Screen. 2001;8(1):1.
2. Down JLH. Observations on an ethnic classification of idiots. London Hosp Rep. 1866;3:259–62.
3. Lejeune J, Gautier M, Turpin R. Study of somatic chromosomes from 9 mongoloid children. C R Hebd Seances Acad Sci. 1959;248(11):1721–2.
4. ACOG Committee on Practice Bulletins. ACOG Practice Bulletin No. 77: screening for fetal chromosomal abnormalities. Obstet Gynecol. 2007;109(1):217–27.
5. Merkatz IR, Nitowsky HM, Macri JN, Johnson WE. An association between low maternal serum alpha-fetoprotein and fetal chromosomal abnormalities. Am J Obstet Gynecol. 1984;148(7):886–94.
6. Cuckle HS, Wald NJ, Lindenbaum RH. Maternal serum alpha-fetoprotein measurement: a screening test for Down syndrome. Lancet. 1984;1(8383):926–9.
7. Palomaki GE, Haddow JE. Maternal serum alpha-fetoprotein, age, and Down syndrome risk. Am J Obstet Gynecol. 1987;156(2):460–3.
8. Bogart MH, Pandian MR, Jones OW. Abnormal maternal serum chorionic gonadotropin levels in pregnancies with fetal chromosome abnormalities. Prenat Diagn. 1987;7(9):623–30.
9. Canick JA, Knight GJ, Palomaki GE, Haddow JE, Cuckle HS, Wald NJ. Low second trimester maternal serum unconjugated oestriol in pregnancies with Down's syndrome. Br J Obstet Gynaecol. 1988;95(4):330–3.
10. Haddow JE, Palomaki GE, Knight GJ, Foster DL, Neveux LM. Second trimester screening for Down's syndrome using maternal serum dimeric inhibin A. J Med Screen. 1998;5:115–9.
11. Lambert-Messerlian G, Kloza EM, Williams J 3rd, Loucky J, O'Brien B, Wilkins-Haug L, et al. Maternal plasma DNA testing

for aneuploidy in pregnancies achieved by assisted reproductive technologies. Genet Med. 2014;16(5):419–22.
12. Wald N, Stone R, Cuckle HS, Grudzinskas JG, Barkai G, Brambati B, et al. First trimester concentrations of pregnancy associated plasma protein A and placental protein 14 in Down's syndrome. BMJ. 1992;305(6844):28.
13. Spencer K, Macri JN, Aitken DA, Connor JM. Free beta-hCG as first-trimester marker for fetal trisomy. Lancet. 1992;339(8807):1480.
14. Szabo J, Gellen J. Nuchal fluid accumulation in trisomy-21 detected by vaginosonography in first trimester. Lancet. 1990;336(8723):1133.
15. Zimmermann R, Hucha A, Savoldelli G, Binkert F, Achermann J, Grudzinskas JG. Serum parameters and nuchal translucency in first trimester screening for fetal chromosomal abnormalities. Br J Obstet Gynaecol. 1996;103(10):1009–14.
16. Wald NJ, Watt HC, Hackshaw AK. Integrated screening for Down's syndrome on the basis of tests performed during the first and second trimesters. N Engl J Med. 1999;341(7):461–7.
17. Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM. First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS). J Med Screen. 2003;10(2):56–104.
18. Palomaki GE, Knight GJ, Ashwood ER, Best RG, Haddow JE. Screening for down syndrome in the United States: results of surveys in 2011 and 2012. Arch Pathol Lab Med. 2013;137(7):921–6.
19. Palomaki GE, Ashwood ER, Best RG, Lambert-Messerlian G, Knight GJ. Is maternal plasma DNA testing impacting serum-based screening for aneuploidy in the United States? Genet Med. 2015;17(11):897–900.
20. Bindra R, Heath V, Liao A, Spencer K, Nicolaides KH. One-stop clinic for assessment of risk for trisomy 21 at 11-14 weeks: a prospective study of 15 030 pregnancies. Ultrasound Obstet Gynecol. 2002;20(3):219–25.
21. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350(9076):485–7.
22. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology. 1989;46(5):318–22.
23. Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, Lyautey J, et al. Microsatellite alterations in plasma DNA of small cell lung cancer patients. Nat Med. 1996;2(9):1033–5.
24. Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. Nat Med. 1996;2(9):1035–7.
25. Odeh M, Granin V, Kais M, Ophir E, Bornstein J. Sonographic fetal sex determination. Obstet Gynecol Surv. 2009;64(1):50–7.
26. Devaney SA, Palomaki GE, Scott JA, Bianchi DW. Noninvasive fetal sex determination using cell-free fetal DNA: a systematic review and meta-analysis. JAMA. 2011;306(6):627–36.
27. Clausen FB, Christiansen M, Steffensen R, Jorgensen S, Nielsen C, Jakobsen MA, et al. Report of the first nationally implemented clinical routine screening for fetal RHD in D- pregnant women to ascertain the requirement for antenatal RhD prophylaxis. Transfusion. 2012;52(4):752–8.
28. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A. 2008;105(51):20458–63.
29. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A. 2008;105(42):16266–71.
30. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ. 2011;342:c7401.
31. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol. 2011;204(3):205.e1–11.
32. Sehnert AJ, Rhees B, Comstock D, de Feo E, Heilek G, Burke J, et al. Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. Clin Chem. 2011;57(7):1042–9.
33. Sparks AB, Wang ET, Struble CA, Barrett W, Stokowski R, McBride C, et al. Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. Prenat Diagn. 2012;32(1):3–9.
34. Sparks AB, Struble CA, Wang ET, Song K, Oliphant A. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. Am J Obstet Gynecol. 2012;206(4):319.e1–9.
35. Zimmermann B, Hill M, Gemelos G, Demko Z, Banjevic M, Baner J, et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. Prenat Diagn. 2012;32(13):1233–41.
36. Gil MM, Akolekar R, Quezada MS, Bregant B, Nicolaides KH. Analysis of cell-Free DNA in maternal blood in screening for aneuploidies: meta-analysis. Fetal Diagn Ther. 2014;35(3):156–73.
37. Gil MM, Quezada MS, Revello R, Akolekar R, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound Obstet Gynecol. 2015;45(3):249–66.
38. Taylor-Phillips S, Freeman K, Geppert J, Agbebiyi A, Uthman OA, Madan J, et al. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ Open. 2016;6(1):e010002.
39. Mackie FL, Hemming K, Allen S, Morris RK, Kilby MD. The accuracy of cell-free fetal DNA-based non-invasive prenatal testing in singleton pregnancies: a systematic review and bivariate meta-analysis. BJOG. 2017;124(1):32–46.
40. Iwarsson E, Jacobsson B, Dagerhamn J, Davidson T, Bernabe E, Heibert Arnlind M. Analysis of cell-free fetal DNA in maternal blood for detection of trisomy 21, 18 and 13 in a general pregnant population and in a high risk population – a systematic review and meta-analysis. Acta Obstet Gynecol Scand. 2017;96(1):7–18.
41. Gil MM, Accurti V, Santacruz B, Plana MN, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. Ultrasound Obstet Gynecol. 2017;50(3):302–14.
42. Jin J, Yang J, Chen Y, Huang J. Systematic review and meta-analysis of non-invasive prenatal DNA testing for trisomy 21: implications for implementation in China. Prenat Diagn. 2017;37(9):864–73.
43. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. PLoS One. 2011;6(7):e21791.
44. Savva GM, Walker K, Morris JK. The maternal age-specific live birth prevalence of trisomies 13 and 18 compared to trisomy 21 (Down syndrome). Prenat Diagn. 2010;30(1):57–64.
45. Gruchy N, Blondeel E, Le Meur N, Joly-Helas G, Chambon P, Till M, et al. Pregnancy outcomes in prenatally diagnosed 47, XXX and 47, XYY syndromes: a 30-year French, retrospective, multicentre study. Prenat Diagn. 2016;36(6):523–9.
46. Mennuti MT, Chandrasekaran S, Khalek N, Dugoff L. Cell-free DNA screening and sex chromosome aneuploidies. Prenat Diagn. 2015;35(10):980–5.

47. Wang Y, Chen Y, Tian F, Zhang J, Song Z, Wu Y, et al. Maternal mosaicism is a significant contributor to discordant sex chromosomal aneuploidies associated with noninvasive prenatal testing. Clin Chem. 2014;60(1):251–9.
48. Nielsen J, Wohlert M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. Hum Genet. 1991;87(1):81–3.
49. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. J Clin Endocrinol Metab. 2003;88(2):622–6.
50. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med. 2011;13(11):913–20.
51. Kalousek DK, Howard-Peebles PN, Olson SB, Barrett IJ, Dorfmann A, Black SH, et al. Confirmation of CVS mosaicism in term placentae and high frequency of intrauterine growth retardation association with confined placental mosaicism. Prenat Diagn. 1991;11(10):743–50.
52. Hartwig TS, Ambye L, Sorensen S, Jorgensen FS. Discordant non-invasive prenatal testing (NIPT) – a systematic review. Prenat Diagn. 2017;37(6):527–39.
53. Malvestiti F, Agrati C, Grimi B, Pompilii E, Izzi C, Martinoni L, et al. Interpreting mosaicism in chorionic villi: results of a monocentric series of 1001 mosaics in chorionic villi with follow-up amniocentesis. Prenat Diagn. 2015;35(11):1117–27.
54. Kalousek DK, Vekemans M. Confined placental mosaicism. J Med Genet. 1996;33(7):529–33.
55. Schreck RR, Falik-Borenstein Z, Hirata G. Chromosomal mosaicism in chorionic villus sampling. Clin Perinatol. 1990;17(4):867–88.
56. Stone JF, Sandberg AA. Sex chromosome aneuploidy and aging. Mutat Res. 1995;338(1–6):107–13.
57. Yao H, Zhang L, Zhang H, Jiang F, Hu H, Chen F, et al. Noninvasive prenatal genetic testing for fetal aneuploidy detects maternal trisomy X. Prenat Diagn. 2012;32(11):1114–6.
58. Curnow KJ, Wilkins-Haug L, Ryan A, Kirkizlar E, Stosic M, Hall MP, et al. Detection of triploid, molar, and vanishing twin pregnancies by a single-nucleotide polymorphism-based noninvasive prenatal test. Am J Obstet Gynecol. 2015;212(1):79.e1–9.
59. Bianchi DW, Chudova D, Sehnert AJ, Bhatt S, Murray K, Prosen TL, et al. Noninvasive prenatal testing and incidental detection of occult maternal malignancies. JAMA. 2015;314(2):162–9.
60. Amant F, Verheecke M, Wlodarska I, Dehaspe L, Brady P, Brison N, et al. Presymptomatic identification of cancers in pregnant women during noninvasive prenatal testing. JAMA Oncol. 2015;1(6):814–9.
61. Snyder HL, Curnow KJ, Bhatt S, Bianchi DW. Follow-up of multiple aneuploidies and single monosomies detected by noninvasive prenatal testing: implications for management and counseling. Prenat Diagn. 2016;36(3):203–9.
62. Osborne CM, Hardisty E, Devers P, Kaiser-Rogers K, Hayden MA, Goodnight W, et al. Discordant noninvasive prenatal testing results in a patient subsequently diagnosed with metastatic disease. Prenat Diagn. 2013;33(6):609–11.
63. Snyder MW, Gammill HS, Shendure J. Copy-number variation and false positive results of prenatal screening. N Engl J Med. 2015;373(26):2585.
64. Zhou X, Sui L, Xu Y, Song Y, Qi Q, Zhang J, et al. Contribution of maternal copy number variations to false-positive fetal trisomies detected by noninvasive prenatal testing. Prenat Diagn. 2017;37(4):318–22.
65. Bianchi DW, Parsa S, Bhatt S, Halks-Miller M, Kurtzman K, Sehnert AJ, et al. Fetal sex chromosome testing by maternal plasma DNA sequencing: clinical laboratory experience and biology. Obstet Gynecol. 2015;125(2):375–82.
66. Gregg AR, Skotko BG, Benkendorf JL, Monaghan KG, Bajaj K, Best RG, et al. Noninvasive prenatal screening for fetal aneuploidy, 2016 update: a position statement of the American College of Medical Genetics and Genomics. Genet Med. 2016;18(10):1056–65.
67. Eiben B, Krapp M, Borth H, Kutur N, Kreiselmaier P, Glaubitz R, et al. Single nucleotide polymorphism-based analysis of cell-free fetal DNA in 3000 cases from Germany and Austria. Ultrasound Int Open. 2015;1(1):E8–E11.
68. McLennan A, Palma-Dias R, da Silva Costa F, Meagher S, Nisbet DL, Scott F. Noninvasive prenatal testing in routine clinical practice--an audit of NIPT and combined first-trimester screening in an unselected Australian population. Aust N Z J Obstet Gynaecol. 2016;56(1):22–8.
69. Quezada MS, Gil MM, Francisco C, Orosz G, Nicolaides KH. Screening for trisomies 21, 18 and 13 by cell-free DNA analysis of maternal blood at 10-11 weeks' gestation and the combined test at 11-13 weeks. Ultrasound Obstet Gynecol. 2015;45(1):36–41.
70. Zhang H, Gao Y, Jiang F, Fu M, Yuan Y, Guo Y, et al. Non-invasive prenatal testing for trisomies 21, 18 and 13: clinical experience from 146,958 pregnancies. Ultrasound Obstet Gynecol. 2015;45(5):530–8.
71. Lau TK, Cheung SW, Lo PS, Pursley AN, Chan MK, Jiang F, et al. Non-invasive prenatal testing for fetal chromosomal abnormalities by low-coverage whole-genome sequencing of maternal plasma DNA: review of 1982 consecutive cases in a single center. Ultrasound Obstet Gynecol. 2014;43(3):254–64.
72. Zhou Q, Pan L, Chen S, Chen F, Hwang R, Yang X, et al. Clinical application of noninvasive prenatal testing for the detection of trisomies 21, 18, and 13: a hospital experience. Prenat Diagn. 2014;34(11):1061–5.
73. Committee on Practice Bulletins—Obstetrics, Committee on Genetics, and the Society for Maternal-Fetal Medicine. Practice Bulletin No. 163: screening for fetal aneuploidy. Obstet Gynecol. 2016;127(5):e123–37.
74. Grati FR, Bajaj K, Zanatta V, Malvestiti F, Malvestiti B, Marcato L, et al. Implications of fetoplacental mosaicism on cell-free DNA testing for sex chromosome aneuploidies. Prenat Diagn. 2017;37:1017.
75. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP, et al. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet Gynecol. 2012;119(5):890–901.
76. Canick JA, Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. Prenat Diagn. 2013;33(7):667–74.
77. Sarno L, Revello R, Hanson E, Akolekar R, Nicolaides KH. Prospective first-trimester screening for trisomies by cell-free DNA testing of maternal blood in twin pregnancy. Ultrasound Obstet Gynecol. 2016;47(6):705–11.
78. Canick JA, Kloza EM, Lambert-Messerlian GM, Haddow JE, Ehrich M, van den Boom D, et al. DNA sequencing of maternal plasma to identify Down syndrome and other trisomies in multiple gestations. Prenat Diagn. 2012;32(8):730–4.
79. Choi H, Lau TK, Jiang FM, Chan MK, Zhang HY, Lo PS, et al. Fetal aneuploidy screening by maternal plasma DNA sequencing: 'false positive' due to confined placental mosaicism. Prenat Diagn. 2013;33(2):198–200.
80. Pan M, Li FT, Li Y, Jiang FM, Li DZ, Lau TK, et al. Discordant results between fetal karyotyping and non-invasive prenatal testing by maternal plasma sequencing in a case of uniparental disomy 21 due to trisomic rescue. Prenat Diagn. 2013;33(6):598–601.
81. Jensen TJ, Dzakula Z, Deciu C, van den Boom D, Ehrich M. Detection of microdeletion 22q11.2 in a fetus by next-generation sequencing of maternal plasma. Clin Chem. 2012;58(7):1148–51.

82. Peters D, Chu T, Yatsenko SA, Hendrix N, Hogge WA, Surti U, et al. Noninvasive prenatal diagnosis of a fetal microdeletion syndrome. N Engl J Med. 2011;365(19):1847–8.
83. Srinivasan A, Bianchi D, Liao W, Sehnert A, Rava R. 52: Maternal plasma DNA sequencing: effects of multiple gestation on aneuploidy detection and the relative cell-free fetal DNA (cffDNA) per fetus. Am J Obstet Gynecol. 2013;208(1, Suppl):S31.
84. Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM, et al. Chromosomal microarray versus karyotyping for prenatal diagnosis. N Engl J Med. 2012;367(23):2175–84.
85. Martin K, Iyengar S, Kalyan A, Lan C, Simon AL, Stosic M, et al. Clinical experience with a single-nucleotide polymorphism-based noninvasive prenatal test for five clinically significant microdeletions. Clin Genet. 2018;93:293. https://doi.org/10.1111/cge.13098.
86. Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. Non-invasive prenatal measurement of the fetal genome. Nature. 2012;487(7407):320–4.
87. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med. 2012;4(137):137ra76.
88. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2(61):61ra91.
89. Chan KC, Jiang P, Sun K, Cheng YK, Tong YK, Cheng SH, et al. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. Proc Natl Acad Sci U S A. 2016;113(50):E8159–E68.
90. Bodian DL, Klein E, Iyer RK, Wong WS, Kothiyal P, Stauffer D, et al. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. Genet Med. 2016;18(3):221–30.
91. Christensen CM. The innovator's dilemma: when new technologies cause great firms to fail. Boston: Harvard Business Review Press; 1997.
92. Hwang J, Christensen CM. Disruptive innovation in health care delivery: a framework for business-model innovation. Health Aff. 2008;27(5):1329–35. 10.1377/hlthaff.27.5.1329
93. ACOG. Committee Opinion No. 545: noninvasive prenatal testing for fetal aneuploidy. Obstet Gynecol. 2012;120(6):1532–4.
94. Benn P, Borrell A, Crossley J, Cuckle H, Dugoff L, Gross S, et al. Aneuploidy screening: a position statement from a committee on behalf of the Board of the International Society for Prenatal Diagnosis, January 2011. Prenat Diagn. 2011;31(6):519–22.
95. Devers PL, Cronister A, Ormond KE, Facio F, Brasington CK, Flodman P. Noninvasive prenatal testing/noninvasive prenatal diagnosis: the position of the national society of genetic counselors. J Genet Couns. 2013;22(3):291–5.
96. Gregg AR, Gross SJ, Best RG, Monaghan KG, Bajaj K, Skotko BG, et al. ACMG statement on noninvasive prenatal screening for fetal aneuploidy. Genet Med. 2013;15(5):395–8.
97. Langlois S, Brock JA, Genetics C, Wilson RD, Audibert F, Brock JA, et al. Current status in non-invasive prenatal detection of down syndrome, trisomy 18, and trisomy 13 using cell-free DNA in maternal plasma. J Obstet Gynaecol Can. 2013;35(2):177–81.
98. Ashoor G, Syngelaki A, Wang E, Struble C, Oliphant A, Song K, et al. Trisomy 13 detection in the first trimester of pregnancy using a chromosome-selective cell-free DNA analysis method. Ultrasound Obstet Gynecol. 2013;41(1):21–5.
99. Brar H, Wang E, Struble C, Musci TJ, Norton ME. The fetal fraction of cell-free DNA in maternal plasma is not affected by a priori risk of fetal trisomy. J Matern Fetal Neonatal Med. 2013;26(2):143–5.
100. Nicolaides KH, Syngelaki A, Ashoor G, Birdir C, Touzet G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. Am J Obstet Gynecol. 2012;207(5):374.e1–6.
101. Nicolaides KH, Syngelaki A, Gil M, Atanasova V, Markova D. Validation of targeted sequencing of single-nucleotide polymorphisms for non-invasive prenatal detection of aneuploidy of chromosomes 13, 18, 21, X, and Y. Prenat Diagn. 2013;33(6):575–9.
102. Palomaki GE, Kloza EM, O'Brien BM, Eklund EE, Lambert-Messerlian GM. The clinical utility of DNA-based screening for fetal aneuploidy by primary obstetrical care providers in the general pregnancy population. Genet Med. 2017;19(7):778–86.
103. Benn P, Curnow KJ, Chapman S, Michalopoulos SN, Hornberger J, Rabinowitz M. An economic analysis of cell-free DNA noninvasive prenatal testing in the US general pregnancy population. PLoS One. 2015;10(7):e0132313.
104. Agarwal A, Sayres LC, Cho MK, Cook-Deegan R, Chandrasekharan S. Commercial landscape of noninvasive prenatal testing in the United States. Prenat Diagn. 2013;33(6):521–31.
105. Deans Z, Khawaja F, Hastings R, Rack K, Patton S, Gutowska-Ding W, Allen S, Jenkins L, Chitty L, Sistermans E. Measuring the quality of NIPT for aneuploidies – results from the first pilot EQA. Prenat Diagn. 2017;37(Suppl 1):6–7. (Abstract)
106. Skotko BG. With new prenatal testing, will babies with Down syndrome slowly disappear? Arch Dis Child. 2009;94(11):823–6.
107. Mozersky J, Ravitsky V, Rapp R, Michie M, Chandrasekharan S, Allyse M. Toward an ethically sensitive implementation of noninvasive prenatal screening in the global context. Hast Cent Rep. 2017;47(2):41–9.

Genomic Applications in the Clinical Management of Infectious Diseases

37

Martina I. Lefterova, Carlos J. Suarez, Niaz Banaei, and Benjamin A. Pinsky

Introduction

Next-generation sequencing (NGS) methods, also referred to as deep, ultra-deep, high-throughput, or massively parallel sequencing, comprise a number of sequencing technologies that have succeeded the traditional dideoxynucleoside chain-termination (i.e., Sanger) method. Various platforms, which differ in their sequencing chemistries, read lengths, and throughput capabilities, are available (reviewed in [1]) (Table 37.1). As these platforms have become more accessible, they have become particularly attractive to clinical microbiology laboratories that already rely on molecular methods for pathogen identification and characterization.

NGS studies of microorganisms typically follow one of two general strategies: targeted sequencing or nontargeted sequencing (Fig. 37.1) [2, 3]. The first approach typically uses target-specific primers for PCR-mediated amplification, so that the genomic regions of interest are enriched and selectively sequenced. This approach is often performed to interrogate well-characterized genomic regions (e.g., identify known drug-resistant mutants). Sequencing for de novo assembly of whole genomes, on the other hand, frequently relies on nontargeted library preparation. Whole-genome sequencing or WGS is often performed on cultured isolates, when microorganisms are unknown or the goal is to define the genomic content and functional potential of the organism under investigation. Nontargeted sequencing may also be applied to primary specimens for culture-independent pathogen identification or characterization of the microbial population. These nontargeted sequencing applications using primary specimens are termed metagenomic sequencing.

Examples of these approaches in infectious disease testing will be discussed with particular attention paid to the technical and bioinformatics challenges that arise with specific scenarios in virology and bacteriology. The use of NGS in clinical microbiology laboratories remains relatively limited, though its role in the diagnosis and management of infectious diseases continues to grow as standardized operational protocols, automation, and data analysis pipelines emerge.

Specific Applications in Diagnostic Virology

Viral Drug Resistance Mutation Testing

The emergence of drug resistance is an important factor in the management of several clinically significant viral infections. Genotypic drug resistance testing was originally performed using "population" or "bulk" sequencing, which involves amplification of specific viral genes followed by Sanger sequencing. However, Sanger methodology has limited sensitivity for minor variants when present at less than 15–20% of the viral population, while NGS methods can detect drug-resistant mutations (DRMs) present at ~1% [4, 5]. The prototypical virus for NGS-based genotypic resistance testing is HIV-1, and similar to Sanger-based methods, emerging NGS assays have used targeted sequencing of viral genomic regions known to develop resistance mutations [4]. Because it has been studied most extensively, HIV-1 will be used as a paradigm for a detailed discussion below of concepts related to NGS-based testing for viral drug resistance. Cytomegalovirus (CMV) will also be discussed. However, genotypic drug resistance testing is also utilized for the

M. I. Lefterova
Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Guardant Health, Redwood City, CA, USA

C. J. Suarez
Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

N. Banaei · B. A. Pinsky (✉)
Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, CA, USA
e-mail: bpinsky@stanford.edu

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_37

Table 37.1 Characteristics of current NGS platforms. The specifications of the instruments were obtained from the manufacturers' websites and/or company's representatives and include up-to-date information as of April 2018

Platform	Manufacturer	Sequencing chemistry	Read length (bp)	Throughput	Run time
MiniSeq	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 150	1.65–7.5 Gb	7–24 h
MiSeq/MiSeqDx	Illumina	Sequencing by synthesis –reversible terminator	Up to 2 × 300	540 Mb–15 Gb	4–56 h
NextSeq 550/ NextSeq 550Dx	Illumina	Sequencing by synthesis –reversible terminator	Up to 2 × 150	16.25–120 Gb	11–29 h
HiSeq 2500	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 250	9 Gb–1 Tb	7 h–11 days
HiSeq 3000/4000	Illumina	Sequencing by synthesis – reversible terminator	2 × 150	105–1500 Gb	<1–3.5 days
HiSeq X	Illumina	Sequencing by synthesis – reversible terminator	2 × 150	1.6–1.8 Tb	< 3 days
NovaSeq 6000	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 150	167 Gb–6 Tb	13–45 h
Ion PGM	Life Technologies[a]	Sequencing by synthesis –hydrogen ion detection	200–400	30 Mb–2 Gb	2–7 h
Ion Proton	Life Technologies[a]	Sequencing by synthesis –hydrogen ion detection	200	10 Gb	2–4 h
Ion S5	Life Technologies[a]	Sequencing by synthesis –hydrogen ion detection	200–600	15–50 Gb	6.5–19 h
PacBio RS II	Pacific Biosciences	Single-molecule real-time sequencing	Mean 14 kb	500 Mb–1 Gb	30 min–6 h
PacBio Sequel	Pacific Biosciences	Single-molecule real-time sequencing	Mean 10 kb	5–10 Gb	30 min–20 h

[a]Thermo Fisher Scientific

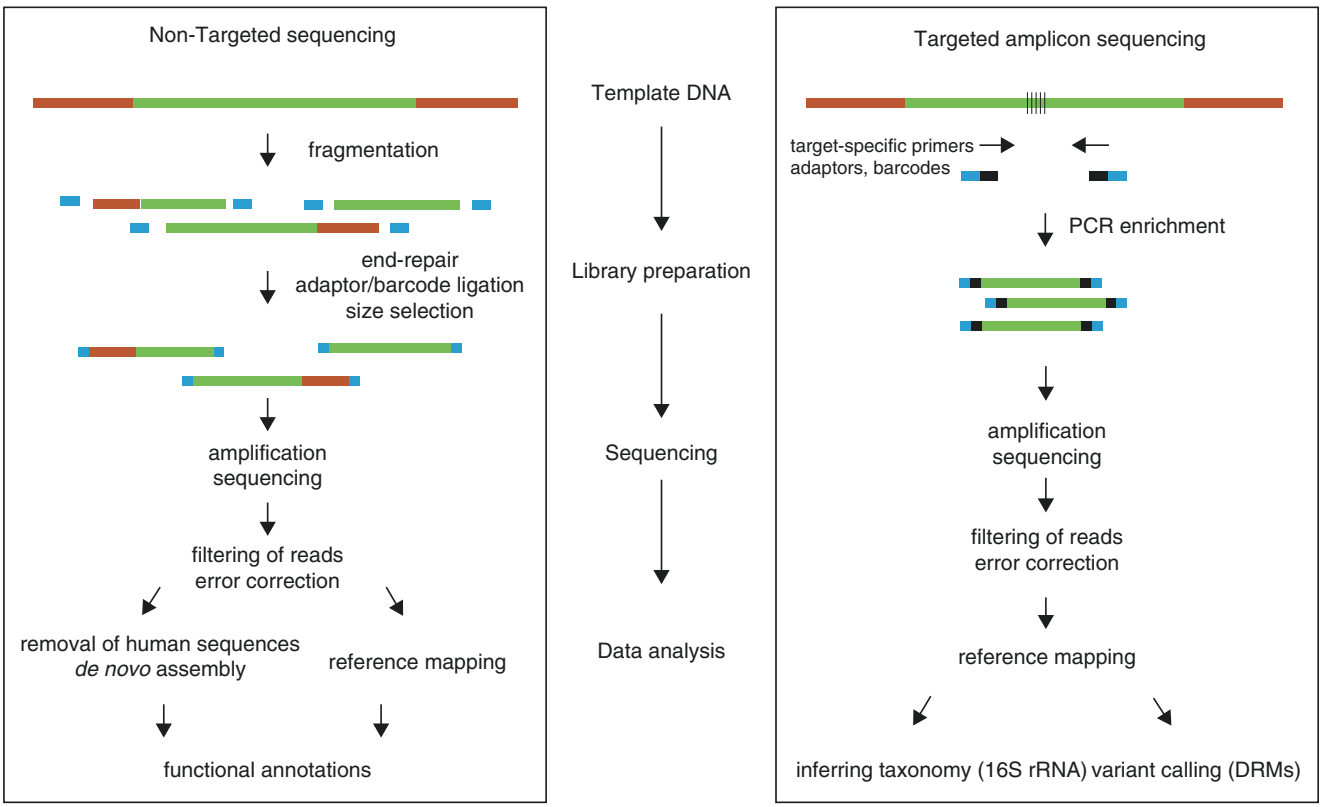


Fig. 37.1 Illustration of sequencing approaches for diagnosis and monitoring of infectious diseases. Targeted amplicon sequencing (left panel) utilizes target-specific primers for template enrichment, followed by primers that are partially complementary to the target-specific primers (black bars), and contains sequencing adaptors and bar codes (blue bars). Nontargeted or metagenonic sequencing (right panel) utilizes enzymatic or mechanical fragmentation, followed by end repair to allow ligation of primers that contain sequencing adaptors and bar codes (blue bars). Size selection allows only fragments of a predefined length to be used for sequencing. Bioinformatics removal of human sequences is required since the nucleic acids of the organism of interest frequently constitute less than 1% of the nucleic acid pool. Note that fragmentation libraries may also be made from PCR-enriched amplicons

clinical management of other viral infections, including hepatitis B, hepatitis C, and influenza, and NGS methods are applicable for these viruses as well.

Human Immunodeficiency Virus Type 1 (HIV-1)

Epidemiologic studies in HIV-1-positive patients have shown that the presence of mutations conferring resistance to highly active antiretroviral therapy (HAART) can predict treatment outcomes [6]. Therefore, genotypic testing for DRMs is currently recommended for therapy-naïve patients when they enter into clinical care and for therapy-experienced patients when they show evidence of virologic failure [7]. A number of studies have compared NGS and Sanger sequencing methods for capturing minority resistant variants, demonstrating that at least half of the DRMs identified by NGS are missed by Sanger sequencing [8, 9]. The presence of such variants has been shown to predict an increased risk for therapy failure [10].

A major consideration when assessing minor variants is distinguishing true mutations from artifacts generated during PCR amplification, library preparation, or sequencing. These include mismatches, insertions/deletions, and PCR-mediated recombination products, known as chimeric sequences [11, 12]. This is particularly problematic for clinical specimens with low virus loads because the numbers of viral copies that are used for library preparation are small and a mixed viral population may not be accurately represented, even with the use of high-fidelity polymerases. Differential amplification of some variants can skew the final PCR product mixture because of stochastic events in early PCR cycles or differences in the efficiency of primer annealing [4, 13]. One possible solution is to estimate empirical error rates for a given NGS assay and for different viral concentrations and then to set thresholds for minor variant detection safely above the empirical error rates. For instance, a plasmid of known genotype can be subjected to NGS and Sanger sequencing, with the assumption that all NGS calls not validated by the Sanger "truth" are due to library preparation and/or NGS errors [14]. Alternatively, the library preparation step can employ primers tagged with a random sequence, such that each template receives a unique identifier. This allows a consensus sequence to be generated for each original template molecule, thus correcting for random errors during library preparation and sequencing [15, 16]. Another approach for addressing PCR bias has been to perform multiple independent amplifications from the same clinical specimen and pool the products to serve as a template for library preparation [17, 18]. Novel bioinformatics tools have also been used to process NGS data in ways that reduce error rates and call authentic low-abundance viral variants [4].

A large number of HIV-1 research studies demonstrating the superior performance of NGS methods compared to Sanger sequencing [4] have resulted in the introduction of several clinical NGS HIV-1 drug resistance assays. The most comprehensive is the DEEPGEN™HIV (developed by University Hospitals Case Medical Center, Cleveland, OH), which assesses for resistance mutations in the protease, reverse transcriptase, and integrase genes, in addition to predicting HIV-1 co-receptor tropism, with mean error rates of 0.37–0.39%, sensitivity for minor variants of 5%, and capacity to multiplex up to 96 samples in a single run [5]. Though not yet available in the USA, Vela Diagnostics have obtained CE marking for their ion PGM-based Sentosa SQ HIV-1 Genotyping Assay for the automated detection of drug resistance mutations in the protease, reverse transcriptase, and integrase genes at a level of 5% [19].

Two other assays that use deep sequencing of the HIV-1 *env* V3 loop for HIV co-receptor tropism are available clinically: the HIV-1 CCR5 tropism test (V3) offered by the British Columbia Centre for Excellence in HIV/AIDS (Vancouver, Canada) [20] and the HIV-1 co-receptor tropism with reflex to ultra-deep sequencing offered by Quest Diagnostics [21]. These assays and DEEPGEN™HIV have been shown to predict non-CCR5 tropism as accurately as the phenotypic gold standard (Trofile, Monogram Biosciences) and to exhibit a higher sensitivity than Sanger sequencing for detecting minor CXCR4-tropic variants.

Importantly, the clinical significance of low-abundance HIV-1 drug resistance variants detected by NGS remains to be fully characterized. Several studies have retrospectively evaluated the impact of low-abundance resistance variants detected by NGS in treatment-naïve patients [8, 9], as well as in treatment-experienced patients with virologic failure [22, 23]. Although patients with low-abundance DRMs detected by NGS alone appear to have a modestly increased risk of failing therapy, in general, the risk of failure is substantially higher with high-abundance mutants that can be demonstrated both by NGS and Sanger sequencing [8].

Cytomegalovirus (CMV)

CMV is another virus for which genotypic drug resistance testing is clinically useful, particularly in transplant recipients [24, 25]. Rates of CMV drug resistance vary based on patient populations: 5–12.5% in solid organ transplant (SOT) recipients and 2–5% in hematopoietic stem cell transplant (HSCT) recipients [26]. Timely detection of CMV drug resistance is critical because DRMs can accumulate with continued exposure to a drug [26, 27], potentially leading to shortened graft survival and increased morbidity [28, 29]. Furthermore, rational change of therapy following

identification of drug resistance has been shown to lead to more rapid clearance of virus [30]. Mutations conferring resistance to the CMV therapeutics, ganciclovir, foscarnet, and cidofovir have been characterized in two CMV genes, the DNA polymerase *UL54* and the phosphotransferase *UL97*, together representing <6 kb of coding sequence, which makes CMV well suited for an amplicon sequencing NGS-based approach analogous to assays targeting HIV protease and reverse transcriptase. In fact, an NGS assay for CMV *UL54* and *UL97* is demonstrating low overall empirical error rate (0.189%) and reliable detection of CMV DRMs in clinical plasma specimens with a wide range of viral loads [18]. Mutations conferring resistance to the terminase inhibitor, Letermovir, FDA-cleared in 2017, have been identified in several genes encoding members of the terminase complex, primarily *UL56* and less commonly *UL89* and *UL51* [31–33]. Subsequent assays for CMV genotypic resistance will likely include *UL56* and may include other genes important for the development of Letermovir resistance.

The impact of minor-population resistant variants on clinical outcomes in CMV-positive patients has not yet been assessed in large clinical trials. However, there is emerging evidence that NGS can facilitate the detection of impending drug resistance and assist in therapy optimization [27]. NGS studies of viral drug resistance are also expected to identify novel putative DRMs, which, after appropriate phenotypic validation [27], can be incorporated into CMV DRM databases and genotypic interpretation systems, similar to those that exist for HIV-1 [34]. Such automated tools have been shown to improve sequence analysis in addition to expediting and standardizing workflow when compared to manual sequence curation [35].

Virus Identification in Clinical Specimens

Proof-of-concept studies have demonstrated the ability of nontargeted, metagenomic sequencing to identify common, clinically relevant viruses from a variety of specimen types previously shown to be positive by routine molecular testing [36]. Another area of diagnostic virology where NGS is being successfully applied is for the identification of viral pathogens in clinical scenarios where a viral agent is suspected but not detected by conventional diagnostic methods [37]. Many viruses cannot be cultured or identified by traditional molecular techniques, while other methods such as cloning and Sanger sequencing are laborious, time-consuming, and mainly applicable to sterile samples like cerebrospinal fluid [2]. Microarrays targeting highly conserved regions within viral families are capable of detecting known viruses, but they cannot identify novel pathogens without sequence similarity to oligonucleotides on the array [37]. In contrast, NGS offers an efficient, highly sensitive, and unbiased alternative for the detection of viruses in clinical specimens [2, 37]. The general approach in such studies is fundamentally different from that used in targeted sequencing. First, the virus of interest is usually not known and therefore cannot be selectively amplified with target-specific primers. Thus, specialized laboratory and bioinformatics strategies are needed to enrich viral RNA or DNA from the predominantly human nucleic acids. Second, a reference sequence may not be available for mapping of sequencing reads if the virus is novel or largely divergent from known related viruses. This necessitates de novo assembly of the viral genome.

As the nucleic acids in clinical specimens are predominantly of host origin, the enrichment of viral and/or depletion of host sequences is an important step for sensitive NGS discovery of viruses in clinical specimens. Laboratory methods for viral particle purification and enrichment include viral culture, ultracentrifugation, density gradient centrifugation, and pretreatment of the sample with nucleases in order to remove host nucleic acids, while preserving capsid-protected viral particles [2, 38]. Nucleic acid amplification methods for enrichment of viral genomes include rolling circle amplification for viruses with a circular genome [39] and use of restriction enzyme sites that are more frequently encountered in viral nucleic acids than human, followed by ligation of adaptors and PCR amplification [40, 41]. Other methods have incorporated hybridization approaches to capture viral nucleic acids with antisense oligonucleotides as baits, although bait design requires at least some prior knowledge of the pathogen [42, 43]. For example, both ViroCap [44] and VirCapSeq-VERT [45] contain probes for capture of all viruses known to infect vertebrates. Similarly, hybridization methods have been designed to deplete human nucleic acids, including methods utilizing CRISPR-based depletion [46]. Furthermore, computational tools have been developed for "subtracting" host sequences from the initial read pool containing mixed human and microbial sequences [47–49]. This filtering step is crucial because viral sequences may comprise <1% of the initial aligned reads [37, 49] (Table 37.2).

Additionally, it is frequently unknown whether a putative viral pathogen contains a DNA or RNA genome, which necessitates processing for total nucleic acid extraction. Amplification with random primers may also be necessary to generate sufficient template for library preparation. An interesting approach to this problem for RNA viruses involves reverse transcription with random primers and cDNA amplification using Phi29 bacteriophage polymerase-based multiple displacement amplification [50]. The choice of sequencing platform (Table 37.1) also requires consideration, as read length and sequence depth may impact virus detection and genome assembly [51].

Table 37.2 Select studies describing culture-independent NGS pathogen identification from primary human clinical specimens

Pathogen name	Clinical information	Specimen(s)	Method	Pathogen aligned reads	Citation(s)
Viruses					
Merkel cell polyomavirus	Rare but aggressive neuroectodermal tumor of the immune compromised	Total RNA from Merkel cell carcinoma tissues	454	0.000003% (1/382,747)	[96]
LCMV-related arenavirus	Fatal febrile illness with sepsis and encephalopathy in three solid organ transplant recipients with the same donor	Total RNA from pooled tissue	454	0.0001% (14/103,632)	[97]
Lujo virus	Hemorrhagic fever outbreak in Southern Africa	Total RNA from liver tissue and serum	454	Not Provided[a]	[98]
Enterovirus 109	Influenza-like illness	Total RNA from nasopharyngeal swab	Illumina	0.00001% (119/10,400,000)	[99]
Influenza A	Upper respiratory illness	Total RNA from nasopharyngeal swabs	Illumina	0.001–0.0001%[b]	[100]
Yellow fever virus	Fatal hemorrhagic fever	Total RNA from serum	454	0.5% (3229/599,158)	[101, 102]
Rubella virus	Anterior and intermediate uveitis	Total RNA from intraocular fluid	Illumina	0.41% (585/1,684,220)	[93]
West Nile virus	Acute meningoencephalitis in a renal transplant recipient	Total RNA from CSF	Illumina	0.001% (101/7,777,470)	[103]
Cache Valley virus	Chronic meningoencephalitis in an agammaglobulinemic patient	Total RNA from CSF	Illumina	0.00002% (5/25,069,677)	[104]
Bacteria					
Francisella tularensis	Abscess	Total DNA from abscess drainage	Illumina	0.002% (833/38,285,502)	[105]
Leptospira santarosai	Recurrent meningoencephalitis in a pediatric patient with SCID	Total DNA from CSF	Illumina	0.016% (475/3,063,784)	[76]
Brucella melitensis	Headache, nausea, and myoclonus in a pediatric patient	Total DNA from CSF	Illumina	0.0012% (277/23,638,587)	[106]
Parasites					
Balamuthia mandrillaris	Endophthalmitis and meningoencephalitis	Total RNA from cerebrospinal fluid	Illumina	0.00002% (5/19,642,962)	[107]

LCMV lymphocytic choriomeningitis virus, *HSCT* hematopoietic stem cell transplant, *FFPE* formalin-fixed, paraffin embedded, *SCID* severe combined immune deficiency

[a]Obtained coverage of 5.6 kb of a 10.4 kb genome

[b]The number of influenza A-specific and total reads varied between samples

Perhaps the most critical aspect of successful viral discovery is the choice of bioinformatics tools. When the reference genome is known, as in amplicon sequencing experiments, read alignment software typically applies stringent mismatch rules in order to minimize errors. In contrast, with unknown pathogens it may be impossible to map reads to publicly available viral databases if the target virus is highly divergent. Instead bioinformatics tools must assemble reads into contiguous sequences (contigs) by identifying overlapping sequences between reads, followed by contig assembly into genomes [52]. Sequencing methods that produce long reads and therefore linkage information facilitate contig assembly (Table 37.1). Sequences assembled this way can be compared to public databases by using algorithms with relaxed stringency in order to identify related viruses. Repetitive sequences pose a significant challenge in de novo assembly, because they can interfere with PCR amplification as well as accurate genomic mapping. Computational and experimental strategies are being developed to address such issues [2, 52]. Table 37.2 summarizes several representative studies in which the NGS approaches described above have been used to identify viral pathogens in patients with infectious syndromes of unclear etiology.

An important caveat to viral discovery is that demonstrating the presence of a virus in a patient with disease does not automatically imply pathogenicity. Traditionally, proving that a microorganism is the causative agent of disease has depended on fulfilling Koch's postulates: a putative etiologic agent is found in affected hosts but not healthy controls; it is propagated in culture and can reproduce the disease when a healthy host is inoculated. However, it is increasingly evident that many viruses cannot be cultured, which has prompted the revision of traditional approaches to prove causality for a microorganism in a disease [53]. Such guidelines eliminate the requirement for microorganism isolation but expand on the rigor with which the association between microorganism and disease is established. For example, it may be necessary to demonstrate the presence of virus in affected tissues using immunostaining or molecular methods, to establish a correlation between viral copy number and

disease severity or to show seroconversion from acute to convalescent plasma specimens.

Specific Applications in Clinical Bacteriology, Mycobacteriology, and Mycology

Identification by Targeted or Nontargeted Sequencing

Genomic approaches are also likely to assist in the diagnosis and management of bacterial, mycobacterial, and fungal infections, including pathogen identification, as well as characterization of virulence factors, strain typing, and antibiotic resistance markers. Clinical microbiology has traditionally relied on isolation of pathogens by culture followed by biochemical tests and more recently matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify the genus or species of the infecting organism [54]. These assays are inexpensive, at least on a per test basis, have a rapid turnaround time, and are therefore appropriate as first-line diagnostics. However, proof-of-concept studies have shown that NGS holds significant potential for microbial identification from primary human specimens, both by targeted amplicon sequencing of ribosomal RNA (rRNA) genes and nontargeted, metagenomics [3].

Ribosomal RNA Sequencing

rRNA gene sequencing by Sanger is routinely used for bacterial and fungal identification in clinical microbiology laboratories [55, 56]. Furthermore, rRNA sequencing by NGS is the basis for studies of the microbiome. For bacteria, these approaches employ primers targeting conserved 16S rRNA sequences, with variable intervening regions that provide sufficient sequence diversity for taxonomic assignment, often to the species level. The choice of primers is important because certain areas of 16S rRNA genes may allow amplification of a broader spectrum of bacteria than others [57]. Additionally, in some cases, classification may only be possible to the family or genus level because the amount of sequence variation may be insufficient for a species-level identification [58].

Informatics is critical for the interpretation of 16S rRNA sequencing data that is obtained by NGS. The length, number, and quality of sequencing reads, as well as possible bacterial contamination of reagents, are all factors that can impact pathogen identification and introduce bias in microbiome diversity assessments [59]. For example, commonly used DNA extraction kits are frequently contaminated with environmental bacteria, leading to overestimation of bacterial diversity in specimens with low starting bacterial loads such as cerebrospinal fluid, blood, or tissue biopsies [60]. Data analysis tools include error correction methods [58, 61] and removal of amplification-derived chimeric sequences [62, 63]. Processed data or raw sequences can be analyzed in dedicated pipelines such as QIIME [64], the Ribosomal Database Project [65] and mothur [62], which cluster similar sequences into operational taxonomic units (OTUs) based on at least 97% sequence identity [3], followed by phylogenetic analyses. In addition, efforts are under way to standardize the use of these pipelines by establishing quality-filtering parameters based on the sequencing platform and the quality of sequencing data [66]. Importantly, the accuracy of bacterial identification largely depends on the scope and completeness of reference databases used for analysis. A number of extensive databases have been created for 16S rRNA sequences: for example, SILVA (www.arb-silva.de) containing >three million small subunit and >250,000 large subunit bacterial rRNA gene sequences [67] or Greengenes (http://www.greengenes.secondgenome.com/downloads) which can calculate taxonomic relationships based on >400,000 16S rRNA sequences [68].

Targeted 16S rRNA sequencing by NGS may have immediate clinical application for the characterization of mixed infections, particularly those containing uncultivatable or nonviable organisms. This approach has been successfully applied directly to brain abscess material, lymph node biopsy tissue, cystic fibrosis (CF) sputa, and mastoid abscess material [69, 70].

Sequencing of 16S rRNA targets by NGS has also been used to study the genomic diversity of bacterial communities, or the microbiome, in health and various disease states. For example, sequencing of the 16S rRNA hypervariable region was used to study bacterial vaginosis, revealing increased bacterial heterogeneity compared to the healthy state [71]. On the other hand, studies of microbiome in the lower airway of cystic fibrosis patients [72] and stool of patients with *Clostridium difficile* infections (CDI) [73] or inflammatory bowel disease (IBD) [74] have shown that disease progression is marked by decreasing bacterial diversity, which may be related to escalating antibiotic exposures. Such results indicate that certain disease states may be driven by disturbances in the normal structure and diversity of a microbial community rather than the action of individual pathogens. Although genomic approaches may elucidate the mechanisms by which changes in the microbiome contribute to disease, the diagnostic utility of characterizing the microbiome in patient management remains to be established.

Metagenomic Sequencing

In contrast to rRNA-based NGS approaches, nontargeted approaches allow more detailed functional and taxonomic analyses, either when a cultured isolate is tested or an entire microbial community is being characterized, a field termed

metagenomics [54, 58]. WGS methods can also be helpful for bacterial pathogen discovery in patients with suspected infections where culture and other standard diagnostic methods have failed. In such scenarios, a direct patient specimen can be sequenced in a relatively unbiased way, similar to what is described above for viral discovery. The potential diagnostic utility of this approach was demonstrated in a pediatric case of severe combined immunodeficiency and recurrent meningoencephalitis, in which WGS coupled with a rapid, dedicated bioinformatics pipeline [75] detected *Leptospira santarosai* sequences in cerebrospinal fluid (CSF) within 48 h of specimen receipt [76]. Table 37.2 shows representative studies in which WGS was used for the culture-independent identification of bacterial pathogens in patients with infectious syndromes of unclear etiology.

Analysis of metagenomic data poses even more challenges than those discussed for 16S rRNA sequencing [58]. When sequencing direct clinical specimens, data need to be filtered for human sequences and sequencing errors. In addition, the putative bacterial reads have to be aligned to reference genomes or subjected to de novo assembly of contigs in order for gene predictions to be made and biological functions to be assigned [58, 75]. Examples of pipelines that have been used for clinical pathogen identification include SURPI (sequence-based ultrarapid pathogen identification) [75, 76] and Taxonomer [77], among many others. However, both taxonomic and functional annotations may be limited by the availability of reference genomes. In that respect, large endeavors exploring bacterial metagenomics in the human host, such as the Human Microbiome Project [78] and Metagenomics of the Human Intestinal Tract (MetaHit) project [79], are actively expanding bacterial genomic databases.

Several groups have assessed the feasibility of nontargeted sequencing for bacterial identification and characterization in a clinical microbiology laboratory. One study tested the feasibility of this approach for routine use by sequencing 130 cultured isolates, including aerobic and anaerobic bacteria, mycobacteria, and fungi [80]. The steps from colony harvest to acquisition of analyzable data took ~ 55 h, with most of the time attributable to the sequencing run (39 h). Comparison of these sequencing results to identification by MALDI-TOF-MS, in addition to conventional culture and biochemical methods, demonstrated good correlation: 115/130 samples (88.5%) showed concordant results, while 15/130 could not be identified due to insufficient coverage or absence of applicable reference genomes in publicly available databases (mainly for sterile molds). Thus nontargeted sequencing was able to identify the majority of organisms identified by conventional methods; however, the turnaround time was substantially slower, and a cost analysis was not performed.

Genotypic Pathogen Characterization

In addition to organism identification, NGS methods can be used to identify genotypic markers of drug resistance and virulence, as well as strain typing [54, 81]. Although phenotypic antimicrobial resistance testing is relatively well standardized, it is available for a limited number of organisms and can take up to several weeks for slow-growing organisms like *Mycobacterium tuberculosis* [82]. Molecular assays with improved sensitivity and turnaround times already exist for some resistance markers; however, resistance to an antimicrobial class can be mediated by several molecular mechanisms, necessitating multiple individual tests or panel testing [82]. Whole-genome-based genotyping, therefore, could simplify workflow and eliminate the need for individual PCR-based assays by simultaneously interrogating all possible genotypic resistance mechanisms, especially if sequencing is being performed for other purposes, such as identification, strain typing, or to detect toxin genes [83]. Use of this approach is likely to expand as new drug resistance mechanisms are characterized and catalogued in publicly available databases such as ResFinder [84] and ARG-ANNOT [85], which use BLAST to query a user-supplied sequence against a curated list of bacterial antimicrobial resistance genes. A number of proof-of-concept studies have assessed the ability of NGS to predict bacterial drug resistance patterns and have been reviewed by Koser et al. [83]. Though genotypic resistance prediction appears feasible, susceptibility determination is nuanced and challenging. Notably, a drawback of genotypic assays is that they do not provide a quantitative measure of antimicrobial susceptibility. In particular, if the presence of a resistance gene or mutation confers variable or inducible resistance, phenotypic assays will still be required [82]. Similarly, when used alone, sequencing may fail to predict a resistance pattern if it has not been characterized genetically or is absent from a given database. For these reasons, whole-genome sequencing is unlikely to replace existing cost-effective resistance assays (phenotypic or molecular) for fast-growing organisms even as cost and turnaround time for sequencing assays continue to decrease. The greatest utility of whole-genome-based drug resistance testing may be for slow-growing organisms such as *M. tuberculosis*, where multidrug regimens are used, phenotypic testing is complex and available for a limited number of drugs, and the number of genes and intergenic regions that need to be targeted for a comprehensive molecular assay is prohibitively large for a targeted approach [86, 87].

Another area where bacterial whole-genome sequencing is being implemented for clinical purposes is for strain typing in hospital outbreak investigations. Traditionally strain typing has been performed either by fragment analysis methods, e.g., pulsed field gel electrophoresis, or by sequence-

based techniques, such as multilocus sequence typing [88]. However, typing schemes exist only for a limited number of organisms, and currently typing is performed primarily in reference and public health laboratories, which means that results are frequently not available within a clinically actionable time frame. In contrast, analysis of single nucleotide polymorphisms (SNPs) based on bacterial whole-genome data can be performed during ongoing outbreaks, does not depend on the availability of established typing schemes, and has higher resolution than most existing sequence-based typing Schemes [54]. In this approach, whole-genome sequencing data are aligned to a reference genome, SNPs are identified and filtered based on preestablished quality metrics, and then phylogenetic analysis is performed to assess the relatedness of bacterial isolates. The feasibility of this approach for reconstructing transmission pathways in hospital outbreak investigations has been demonstrated in a number of studies [54]. However, it remains to be shown whether the use of whole-genome sequencing in the setting of hospital outbreaks will be cost-effective and will be associated with prevention of transmission events.

Validation, Quality Control, and Maintenance of Proficiency

The use of any diagnostic test in the clinical laboratory requires analytical and clinical validation, as well as the careful monitoring and documentation of quality control and proficiency testing (Table 37.3). In that regard, NGS performed in the clinical laboratory for patient care differs from NGS performed in the research setting, even though the sequencing methods may be the same. As such, the American College of Medical Genetics and Genomics (ACMG) has published detailed clinical laboratory standards for NGS [89]. Furthermore, the College of American Pathologists (CAP) has developed an NGS checklist for accreditation of molecular pathology laboratories performing clinical NGS testing [90]. The molecular pathology NGS checklist details requirements for documentation, validation, quality control, and quality monitoring for both the wet bench work and bioinformatics and includes guidelines for data storage, as well as the assessment and implementation of new technology and software releases. Though this checklist has been updated to include examples relevant to NGS for infectious diseases, it is anticipated that in the future, the microbiology checklist will contain a separate section for NGS tailored specifically for microbiology. To further assist in the validation of NGS-based assays for infectious diseases, the American Society for Microbiology and CAP published a manuscript describing the challenges and potential solutions for validating metagenomic pathogen detection tests in clinical laboratories [91].

Table 37.3 Assessment of the performance characteristics of NGS-based tests for clinical microbiology

Performance characteristic	Approach to evaluation
Accuracy	Use of specimens with known findings and confirmation of additional findings detected during validation by an orthogonal method
Precision	Reproducibility (between-run precision): sequencing of the same samples on different runs Repeatability (within-run precision): Sequencing of the same samples in replicates within a run
	Between library precision: sequencing different library preparations of the same samples on the same sequencing run [108]
Analytical sensitivity	Microbial variant detection: mixes of known variant strains and wild-type strains at different percentages and at low, medium, and high levels (e.g., viral loads) Microbial identification: serial dilutions of samples in an appropriate matrix containing a known pathogen(s) coupled with an estimation of the minimum coverage needed to detect the pathogen
Analytical specificity	Microbial variant detection and microbial identification: estimation of the false-positive rate at various read depths

However, the application of NGS in clinical infectious disease testing poses unique challenges that are distinct from the diagnostic settings of human inherited diseases or cancer. For example, as NGS is increasingly adopted for clinical microbiology, well-characterized and extensively sequenced reference microbial organisms will be required for use as controls and proficiency material. In order to supplement reference strains, mock sequence data may also be necessary to ensure adequate bioinformatics pipelines. These in silico controls and proficiency challenges will be particularly important for the clinical characterization of the microbial metagenome, low-level DRM detection, and the identification of organisms that are unculturable or difficult to culture.

NGS technologies that are being used for clinical infectious disease testing are currently being performed as laboratory-developed tests, as no clinical microbiology NGS tests have yet been approved by the United States Food and Drug Administration (FDA). Nevertheless, the FDA is keenly interested in the regulatory oversight of NGS in clinical microbiology, particularly for microbial identification and the detection of antimicrobial resistance markers. As such, the FDA has published a discussion paper detailing clinical applications and validation approaches for the regulatory approval/clearance of NGS diagnostic devices for clinical microbiology [92]. Of note, this document reports that the FDA is engaged in the development of a database (FDA MicroDB) comprised of >550 high-quality, "regulatory-grade" sequences from clinically relevant bacterial microorganisms to be used in the pathway for regulatory approval. The availability of FDA-approved infectious dis-

ease NGS in vitro diagnostics will likely aid in the standardization of specimen handling, library preparation and sequencing, as well as data interpretation, in order to ensure the accuracy and reproducibility of NGS-derived genotypic results. This standardization and quality assurance may be particularly important given that contaminating microbial DNA is ubiquitously found in commonly used extraction kits and reagents used for NGS, as well as "sterile" specimen transport containers [60].

Conclusions

In this chapter we have reviewed areas of clinical microbiology in which next-generation sequencing approaches have been used to identify and characterize medically important pathogens. While many of these studies have been conducted as proof-of-concept experiments or research investigations, NGS-based testing has already been adopted in select diagnostic microbiology laboratories, including academic clinical laboratories, large commercial reference laboratories, and startup companies. Routine applications are likely to increase as cost, turnaround time, and complexity decrease sufficiently to make NGS complementary to existing affordable, standardized, and considerably simpler methods. As technologies like this one are developed and evaluated, the use of NGS for infectious diseases testing may become more widespread.

Targeted NGS assays relying on amplicon sequencing, such as HIV drug resistance testing, were the first to be introduced clinically given the sensitivity advantages over Sanger sequencing and the accumulating data supporting the clinical relevance of low-abundance resistance mutations. NGS-based amplicon sequencing of ribosomal RNA genes may also become more commonly used for identification of pathogenic bacteria and fungi when there is high suspicion for infection and culture is negative or not available or when mixed infections are suspected. Metagenomic strategies may also be useful for pathogen identification in sterile specimens if testing can be optimized to provide clinically actionable data faster than culture or currently available molecular methods. Importantly, the ability of NGS methods and bioinformatics pipelines to accurately identify and characterize pathogens will need to be rigorously validated and compared with traditional diagnostic techniques [93, 94].

The greatest attraction of genomic approaches is that metagenomics sequencing could provide all relevant information about a pathogen in a single assay, including species identification, strain typing, virulence determination, and antimicrobial resistance. In practice, widespread implementation of NGS in clinical microbiology laboratories will require acquisition of costly new equipment and, in particular, the training of personnel in methods that are reliant on bioinformatics. Bioinformatics pipelines will need to provide user-friendly interfaces that allow the user to input data directly from the sequencing instrument and receive best-hit matches to comprehensive and well-curated reference genome databases [54].

Thus, at this point in time, NGS methods are expected to supplement, rather than replace, conventional diagnostic testing. An important hurdle, even in the most sophisticated of clinical laboratories, is that genotype-phenotype correlations for many clinically relevant microorganisms are unknown, although large-scale metagenomic efforts like the Human Microbiome Project will undoubtedly define numerous new associations between sequence and function. Ultimately, the tremendous promise of NGS methods for diagnostic infectious disease testing will require the successful development of clinical microbiologists capable of interpreting and evaluating NGS data and placing these data in the appropriate clinical context.

Acknowledgments Material and references were used from Lefterova MI, Banaei N, and Pinsky BA. Genomic Applications in the Clinical Management of Infectious Diseases. In: Netto GJ, Schrijver I, eds. *Genomic Applications in Pathology.* 1st Ed. Springer. 2015. The authors thank the AMP NGS in ID Work Group, a joint project of the AMP Infectious Diseases Subdivision Leadership and Clinical Practice Committee for their contributions to the previous version of this manuscript [95].

References

1. Loman NJ, Constantinidou C, Chan JZ, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol. 2012;10(9):599–606.
2. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. J Gen Virol. 2012;93(Pt 9):1853–68.
3. Weinstock GM. Genomic approaches to studying the human microbiota. Nature. 2012;489(7415):250–6.
4. Vrancken B, Lequime S, Theys K, Lemey P. Covering all bases in HIV research: unveiling a hidden world of viral evolution. AIDS Rev. 2010;12(2):89–102.
5. Gibson RM, Schmotzer CL, Quinones-Mateu ME. Next-generation sequencing to help monitor patients infected with HIV: ready for clinical use? Curr Infect Dis Rep. 2014;16(4):401.
6. Dunn DT, Coughlin K, Cane PA. Genotypic resistance testing in routine clinical care. Curr Opin HIV AIDS. 2011;6(4):251–7.
7. Cortez KJ, Maldarelli F. Clinical management of HIV drug resistance. Viruses. 2011;3(4):347–78.
8. Simen BB, Simons JF, Hullsiek KH, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis. 2009;199(5):693–701.
9. Lataillade M, Chiarella J, Yang R, et al. Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. PLoS One. 2010;5(6):e10952.
10. Li JZ, Paredes R, Ribaudo HJ, et al. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. JAMA. 2011;305(13):1327–35.

11. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. J Virol Methods. 2010;169(1):248–52.
12. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 2009;10(3):R32.
13. Mild M, Hedskog C, Jernberg J, Albert J. Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. PLoS One. 2011;6(7):e22741.
14. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17(8):1195–201.
15. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A. 2011;108(50):20166–71.
16. Clutter DS, Zhou S, Varghese V, et al. Prevalence of drug-resistant minority variants in untreated HIV-1-infected individuals with and those without transmitted drug resistance detected by Sanger sequencing. J Infect Dis. 2017;216(3):387–91.
17. Vandenbroucke I, Van Marck H, Mostmans W, et al. HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. AIDS Res Ther. 2010;7:4.
18. Sahoo MK, Lefterova MI, Yamamoto F, et al. Detection of cytomegalovirus drug resistance mutations by next-generation sequencing. J Clin Microbiol. 2013;51(11):3700–10.
19. Tzou PL, Ariyaratne P, Varghese V, et al. Comparison of an in vitro diagnostic next-generation sequencing assay with Sanger sequencing for HIV-1 genotypic resistance testing. J Clin Microbiol. 2018;56(6):pii: e00105-18.
20. Swenson LC, Mo T, Dong WW, et al. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. J Infect Dis. 2011;203(2):237–45.
21. Kagan RM, Johnson EP, Siaw M, et al. A genotypic test for HIV-1 tropism combining Sanger sequencing with ultradeep sequencing predicts virologic response in treatment-experienced patients. PLoS One. 2012;7(9):e46334.
22. Le T, Chiarella J, Simen BB, et al. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. PLoS One. 2009;4(6):e6079.
23. Codoner FM, Pou C, Thielen A, et al. Added value of deep sequencing relative to population sequencing in heavily pretreated HIV-1-infected subjects. PLoS One. 2011;6(5):e19461.
24. Kotton CN, Kumar D, Caliendo AM, et al. International consensus guidelines on the management of cytomegalovirus in solid organ transplantation. Transplantation. 2010;89(7):779–95.
25. Ljungman P, Reusser P, de la Camara R, et al. Management of CMV infections: recommendations from the infectious diseases working party of the EBMT. Bone Marrow Transplant. 2004;33(11):1075–81.
26. Lurain NS, Chou S. Antiviral drug resistance of human cytomegalovirus. Clin Microbiol Rev. 2010;23(4):689–712.
27. Chou S, Ercolani RJ, Sahoo MK, Lefterova MI, Strasfeld LM, Pinsky BA. Improved detection of emerging drug-resistant mutant cytomegalovirus subpopulations by deep sequencing. Antimicrob Agents Chemother. 2014;58(8):4697–702.
28. Li F, Kenyon KW, Kirby KA, Fishbein DP, Boeckh M, Limaye AP. Incidence and clinical features of ganciclovir-resistant cytomegalovirus disease in heart transplant recipients. Clin Infect Dis. 2007;45(4):439–47.
29. Limaye AP, Corey L, Koelle DM, Davis CL, Boeckh M. Emergence of ganciclovir-resistant cytomegalovirus disease among recipients of solid-organ transplants. Lancet. 2000;356(9230):645–9.
30. Myhre HA, Haug Dorenberg D, Kristiansen KI, et al. Incidence and outcomes of ganciclovir-resistant cytomegalovirus infections in 1244 kidney transplant recipients. Transplantation. 2011;92(2):217–23.
31. Chou S. Comparison of cytomegalovirus terminase gene mutations selected after exposure to three distinct inhibitor compounds. Antimicrob Agents Chemother. 2017;61(11)
32. Chou S. Rapid in vitro evolution of human cytomegalovirus UL56 mutations that confer letermovir resistance. Antimicrob Agents Chemother. 2015;59(10):6588–93.
33. Chou S. A third component of the human cytomegalovirus terminase complex is involved in letermovir resistance. Antivir Res. 2017;148:1–4.
34. Tang MW, Liu TF, Shafer RW. The HIVdb system for HIV-1 genotypic resistance interpretation. Intervirology. 2012;55(2):98–101.
35. Woods CK, Brumme CJ, Liu TF, et al. Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol. 2012;50(6):1936–42.
36. Petty TJ, Cordey S, Padioleau I, et al. Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. J Clin Microbiol. 2014;52(9):3351–61.
37. Chiu CY. Viral pathogen discovery. Curr Opin Microbiol. 2013;16(4):468–78.
38. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G. Applications of next-generation sequencing technologies to diagnostic virology. Int J Mol Sci. 2011;12(11):7861–84.
39. Rector A, Tachezy R, Van Ranst M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. J Virol. 2004;78(10):4993–8.
40. de Vries M, Deijs M, Canuti M, et al. A sensitive assay for virus discovery in respiratory clinical samples. PLoS One. 2011;6(1):e16118.
41. Pyrc K, Jebbink MF, Berkhout B, van der Hoek L. Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. Methods Mol Biol. 2008;454:73–89.
42. Depledge DP, Palser AL, Watson SJ, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One. 2011;6(11):e27805.
43. Duncavage EJ, Magrini V, Becker N, et al. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. J Mol Diagn. 2011;13(3):325–33.
44. Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. Genome Res. 2015;25(12):1910–20.
45. Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio. 2015;6(5):e01491-15.
46. Gu W, Crawford ED, O'Donovan BD, et al. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol. 2016;17:41.
47. Bhaduri A, Qu K, Lee CS, Ungewickell A, Khavari PA. Rapid identification of non-human sequences in high-throughput sequencing datasets. Bioinformatics. 2012;28(8):1174–5.
48. Kostic AD, Ojesina AI, Pedamallu CS, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat Biotechnol. 2011;29(5):393–6.
49. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl Trop Dis. 2012;6(2):e1485.

50. Berthet N, Reinhardt AK, Leclercq I, et al. Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. BMC Mol Biol. 2008;9:77.
51. Cheval J, Sauvage V, Frangeul L, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J Clin Microbiol. 2011;49(9):3268–75.
52. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. Genomics. 2010;95(6):315–27.
53. Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity – a personal recollection 15 years later. Nat Rev Microbiol. 2004;2(1):67–72.
54. Koser CU, Ellington MJ, Cartwright EJ, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 2012;8(8):e1002824.
55. Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17(4):840–62. table of contents
56. Balajee SA, Sigler L, Brandt ME. DNA and the classical way: identification of medically important molds in the 21st century. Med Mycol. 2007;45(6):475–90.
57. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41(1):e1.
58. Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet. 2012;13(1):47–58.
59. Cox MJ, Cookson WO, Moffatt MF. Sequencing the human microbiome in health and disease. Hum Mol Genet. 2013;22(R1):R88–94.
60. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
61. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One. 2011;6(12):e27310.
62. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.
63. Haas BJ, Gevers D, Earl AM, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 2011;21(3):494–504.
64. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–6.
65. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37(Database issue):D141–5.
66. Bokulich NA, Subramanian S, Faith JJ, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10(1):57–9.
67. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590–6.
68. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012;6(3):610–8.
69. Salipante SJ, Sengupta DJ, Rosenthal C, et al. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. PLoS One. 2013;8(5):e65226.
70. Salipante SJ, Hoogestraat DR, Abbott AN, et al. Coinfection of Fusobacterium nucleatum and Actinomyces israelii in mastoiditis diagnosed by next-generation DNA sequencing. J Clin Microbiol. 2014;52(5):1789–92.
71. Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. PLoS One. 2012;7(6):e37818.
72. Zhao J, Schloss PD, Kalikin LM, et al. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A. 2012;109(15):5809–14.
73. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. 2008;6(11):e280.
74. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.
75. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res. 2014;24(7):1180–92.
76. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med. 2014;370(25):2408–17.
77. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. Genome Biol. 2016;17(1):111.
78. Group NHW, Peterson J, Garges S, et al. The NIH Human Microbiome Project. Genome Res. 2009;19(12):2317–23.
79. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65.
80. Long SW, Williams D, Valson C, et al. A genomic day in the life of a clinical microbiology laboratory. J Clin Microbiol. 2013;51(4):1272–7.
81. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet. 2012;13(9):601–12.
82. Jenkins SG, Schuetz AN. Current concepts in laboratory testing to guide antimicrobial therapy. Mayo Clin Proc. 2012;87(3):290–308.
83. Koser CU, Ellington MJ, Peacock SJ. Whole-genome sequencing to control antimicrobial resistance. Trends Genet. 2014;30(9):401–7.
84. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640–4.
85. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212–20.
86. Warner DF, Mizrahi V. Complex genetics of drug resistance in Mycobacterium tuberculosis. Nat Genet. 2013;45(10):1107–8.
87. Heysell SK, Houpt ER. The future of molecular diagnostics for drug-resistant tuberculosis. Expert Rev Mol Diagn. 2012;12(4):395–405.
88. MacCannell D. Bacterial strain typing. Clin Lab Med. 2013;33(3):629–50.
89. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15(9):733–47.
90. College of American Pathologists. Molecular pathology checklist. Next generation sequencing. Northfield: College of American Pathologists; 2014. p. 8.
91. Schlaberg R, Chiu CY, Miller S, et al. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Arch Pathol Lab Med. 2017;141(6):776–86.

92. Food and Drug Administration. High-throughput sequencing technologies for microbial identification and detection of antimicrobial resistance markers. 2014.
93. Doan T, Acharya NR, Pinsky BA, et al. Metagenomic DNA sequencing for the diagnosis of intraocular infections. Ophthalmology. 2017;124(8):1247–8.
94. Graf EH, Simmon KE, Tardif KD, et al. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. J Clin Microbiol. 2016;54(4):1000–7.
95. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and management: a report of the Association for Molecular Pathology. J Mol Diagn. 2015;17(6):623–34.
96. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 2008;319(5866):1096–100.
97. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med. 2008;358(10):991–8.
98. Briese T, Paweska JT, McMullan LK, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS Pathog. 2009;5(5):e1000455.
99. Yozwiak NL, Skewes-Cox P, Gordon A, et al. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. J Virol. 2010;84(18):9047–58.
100. Yongfeng H, Fan Y, Jie D, et al. Direct pathogen detection from swab samples using a new high-throughput sequencing technology. Clin Microbiol Infect. 2011;17(2):241–4.
101. McMullan LK, Frace M, Sammons SA, et al. Using next generation sequencing to identify yellow fever virus in Uganda. Virology. 2012;422(1):1–5.
102. Wamala JF, Malimbo M, Okot CL, et al. Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010-January 2011. Int J Infect Dis. 2012;16(7):e536–42.
103. Wilson MR, Zimmermann LL, Crawford ED, et al. Acute West Nile virus meningoencephalitis diagnosed via metagenomic deep sequencing of cerebrospinal fluid in a renal transplant patient. Am J Transplant. 2017;17(3):803–8.
104. Wilson MR, Suan D, Duggins A, et al. A novel cause of chronic viral meningoencephalitis: Cache Valley virus. Ann Neurol. 2017;82(1):105–14.
105. Kuroda M, Sekizuka T, Shinya F, et al. Detection of a possible bioterrorism agent, Francisella sp., in a clinical specimen by use of next-generation direct DNA sequencing. J Clin Microbiol. 2012;50(5):1810–2.
106. Mongkolrattanothai K, Naccache SN, Bender JM, et al. Neurobrucellosis: unexpected answer from metagenomic next-generation sequencing. J Pediatric Infect Dis Soc. 2017;6(4):393–8.
107. Wilson MR, Shanbhag NM, Reid MJ, et al. Diagnosing Balamuthia mandrillaris encephalitis with metagenomic deep sequencing. Ann Neurol. 2015;78(5):722–30.
108. Frey KG, Herrera-Galeano JE, Redden CL, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics. 2014;15:96.

Pharmacogenomics: Success and Challenges

38

Mohammad Omar Hussaini and Howard L. McLeod

Introduction

The central dogma of personalized medicine is to tailor therapy based on the genetic makeup of the patient with the principal aim of maximizing benefit (drug efficacy) while minimizing harm (side effects). This can be done by choosing the right drug and administering it at the right dose. Rather than operating, as we have historically done, based on statistical difference between large groups of individuals, personalized medicine tries to then layer in the unique biology of each patient to create an informed therapeutic strategy. Pharmacogenomics undergirds the enterprise of personalized medicine by telling us how information from the genetic code may affect response to a drug. To be more precise, pharmacogenomics elucidates how individual genetic variation affects both drug disposition and effect. It includes the impact of genetic variation on:

1. Pharmacokinetics: how the body handles the drug (absorption, distribution, metabolism, and elimination).
2. Pharmacodynamics: how the drug affects the body. This will vary based on individual differences in the drug target, ion channels, and intracellular regulatory proteins.

Pharmacogenomics has been greatly facilitated by the advent of robust sequencing technologies, namely, next-generation sequencing, which have made it relatively straightforward to interrogate genomes to unearth genetic variations that variably explain drug response phenotypes [1].

M. O. Hussaini
Department of Hematopathology and Laboratory Medicine, Moffitt Cancer Center, Tampa, FL, USA

H. L. McLeod (✉)
Department of Cancer Epidemiology, Individualized Cancer Medicine, Moffitt Cancer Center, Tampa, FL, USA

Department of Individualized Cancer Medicine, Moffitt Cancer Center, Tampa, FL, USA
e-mail: howard.mcleod@moffitt.org

The underlying genetic alterations responsible for pharmacogenomic phenotypes may be monogenic, oligogenic, or complex (numerous small defect variants with environmental and epigenetic modifiers) with most falling in the last category [2]. Usually monogenic variants (e.g., *NAT2, TPMT, CYP2D6*) tend to have larger phenotypic effects versus complex polygenic variants where each gene exerts a smaller effect on phenotype. However, incomplete penetrance, allelic heterogeneity, and other genomic/environmental modifiers complicate the interpretation of these monogenic variants which typically explain some, but not all, of the observed phenotypic variation [3]. Oligogenic variants may have appreciable effects resulting in several modal pharmacogenomic subgroups, a distinction that again is blurred clinically due to various environmental influences. An example of oligogenic variation is warfarin dosing based on variants in *VCORC1* and *CYP2CP*, each of which partially contributes to the degree of anticoagulation achieved in patients with these variants [2]. Given the polygenic contribution to pharmacogenomics, there is an argument to be made for preemptive multigene assays that provide a more comprehensive picture of the patient's clinically relevant genome to inform current and future drugability and dosing [1].

Evolution of Pharmacogenomics

Pythagoras in 510 BC identified a subset of individuals who experienced fatal hemolytic anemia after ingesting broad beans, a phenomenon that nearly 2000 years later was found to be attributable to glucose-6-phosphate dehydrogenase (G6PD) deficiency. Patients with this deficiency may experience hemolysis in response to various medications such as primaquine, an antimalarial drug [1]. Modern cancer pharmacogenomics has its roots in the 1950s when inherited differences in drug effects due to variations in metabolism were first documented [4, 5] giving rise to the term "pharmacogenetics." No longer limited to drug metabolism, the field has now encompasses all aspects of drug disposition, including

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_38

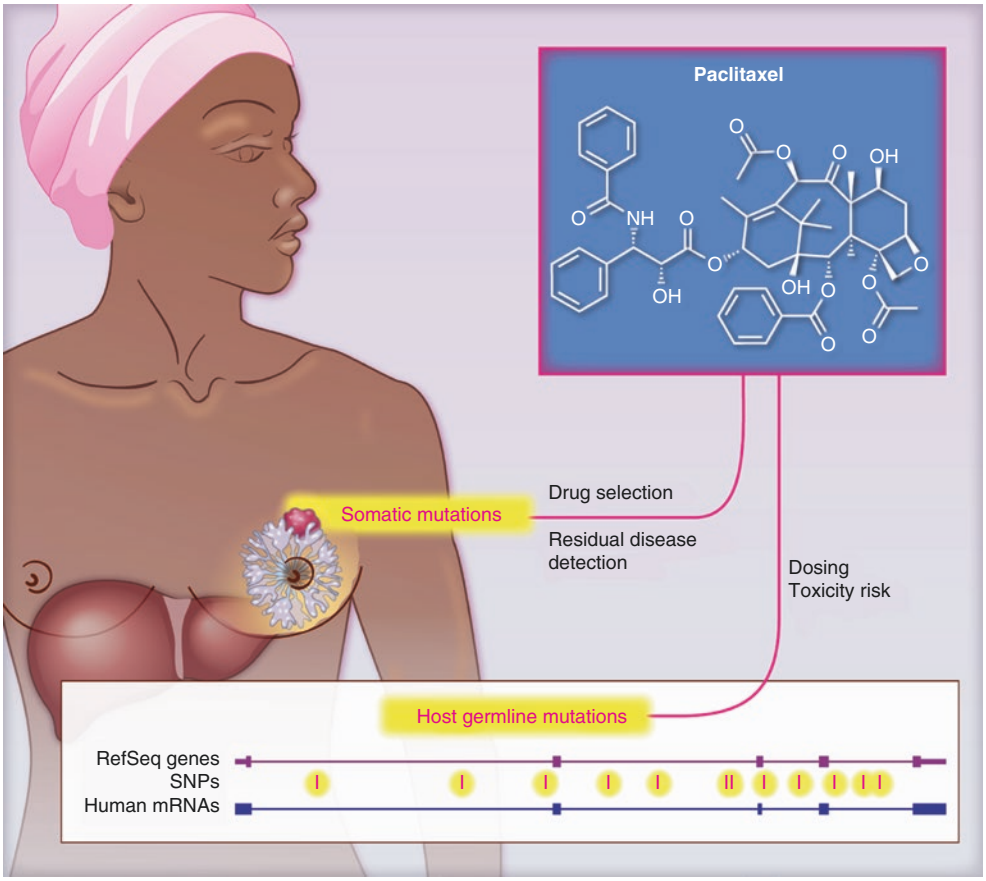


Fig. 38.1 Attention must be paid to both tumor and host: cancer pharmacogenomic variation in both the tumor (somatic changes) and normal tissues (germline variants) influence the treatment of cancer patients. (From McLeod [79] with permission the American Association for the Advancement of Science)

absorption, distribution, and excretion [6], as well as drug targets and downstream mediators of drug effect. In 1987, *CYP2D6* became the first drug metabolizing gene clone to be cloned. Subsequently, the 1990s were host to delineation of pharmacogenomic utility for several genes [1, 7, 8].

Over the past few decades, deeper mechanistic understanding of tumor biology, tumor heterogeneity, and the biologic pathways regulating neoplastic cells and their normal counterparts has catalyzed efforts to develop therapies precisely designed to target critical tumor gene(s) or pathways [9–12]. Today, more than 230 FDA-approved drug agents carry pharmacogenomic information in their label inserts which has nearly doubled since 2013 – all in all, indicating that, at varying levels, pharmacogenomics is already becoming well incorporated into the modern-day practice of health care [13, 14]. Drugs with pharmacogenomic labeling include neoplastic agents, psychotropic drugs, anti-infectious agents, and central nervous system agents. Pharmacogenetic labeling may indicate genetic testing to be required and recommended, actionable pharmacogenomics, or informative pharmacogenomics, the last category signaling possible involvement of the gene with drug action [2, 15].

Disease outcome, drug effect, and therapy response can be affected by both somatic mutations (acquired tumor-specific genetic variation) and germline mutations (heritable variations found in all cells of the individual) (Fig. 38.1). These mutations can alter the structure of the target protein or alter the quantity of the protein via gene regulation or epigenetic modulation [16].

Cancer Pharmacogenomics and Tumor Profiling: From Discovery to Patient Management

In cancer, acquired somatic mutations (biomarkers) can be broadly classified as being prognostic or predictive, forecasting how a patient will fare with his or her cancer or how amenable their cancer is to a given therapy, respectively. Of course, this is an inexact science at the time based largely on cohort studies demonstrating associations between these biomarkers and any of several given outcome variables (overall survival, disease-free survival, etc.). The FDA defines a valid biomarker as one with (1) established pharmacologic or clinical significance and (2) one for which a validated assay exists [3]. Table 38.1 outlines some current examples where genotype is used for the selection of cancer chemotherapy.

Pathway-based therapeutics have transformed certain cancers, such as chronic myelogenous leukemia and gastrointestinal stromal tumors, into essentially chronic diseases

Table 38.1 Pharmacogenomic DNA markers in clinical use for chemotherapy or supportive care of cancer patients

Germline	Somatic	Drugs	Effect	Disease
Thiopurine methyltransferase (*TPMT*)	--------	Mercaptopurine, thioguanine	Neutropenia risk	Acute lymphoblastic leukemia
UDP-glucuronosyltransferase 1A1 (*UGT1A1*)	--------	Irinotecan, nilotinib	Neutropenia risk; underdosing risk	Lung, colorectal
Glucose-6-phosphate dehydrogenase	--------	Rasburicase	Anemia	Glucose-6-phosphate dehydrogenase (G6PD) deficiency
Cytochrome P450 2D6 (*CYP2D6*)	--------	Codeine, oxycodone; tamoxifen	Altered pain control; Altered tumor control	Estrogen receptor or progesterone receptor-positive breast cancer
Solute carrier organic anion transporter family, member 1B1 (*SLCO1B1*)	--------	Methotrexate	Overdosing risk and GI toxicity [88]	
--------	Janus kinase 2 (JAK2)	Ruxolitinib	Altered drug activity	Primary myelofibrosis
--------	Human epidermal growth factor receptor 1 (EGFR)	Cetuximab Erlotinib Gefitinib Panitumumab	Altered drug activity	NLSCLC, CRC
--------	Kirsten rat sarcoma viral oncogene homolog (KRAS)	Cetuximab Panitumumab	Lack of drug activity	NLSCLC, CRC
	Abelson murine leukemia viral oncogene homolog 1 (ABL)	Imatinib, dasatinib, nilotinib	Altered drug activity	Chronic myelogenous leukemia
--------	v-KIT hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT)	Imatinib	Altered drug activity	Gastrointestinal stromal tumor
--------	Human epidermal growth factor receptor 2 (HER2)	Lapatinib Trastuzumab	Enhanced drug activity	Breast cancer
--------	v-Raf murine sarcoma viral oncogene homolog B1 (BRAF)	Vemurafenib	Enhanced drug activity	Melanoma, CRC
--------	Anaplastic lymphoma receptor tyrosine kinase (ALK)	Crizotinib	Altered drug activity	NSCLC
	ESR1, ESR2, PGR	Exemestane, anastrazole, letrozole, tamoxifen, toremifene, fulvestrant	Enhanced drug activity	Breast cancer [89]
Dihydropyrimidine dehydrogenase (*DPD*)		Fluorouracil, capecitabine	Myelosuppression, GI, and cardiac toxicity	CRC [90, 91]
	PML/RARA fusion	All-transretinoic acid	Altered drug activity	Acute promyelocytic leukemia
	VEGF	Bevacizumab	Altered drug activity	Metastatic CRC, recurrent glioblastoma multiforme [92, 93]
	BRCAness	Cisplatin, PARP inhibitors	Altered drug activity	Ovarian cancer [94]
	FLT3	Midostaurin	Multikinase inhibitor	Acute myeloid leukemia

(barring relapse or progression) requiring lifelong treatment akin to diabetes or hypertension [17]. This has proven to be a two-edged sword. With targeted molecular therapy carrying a hefty bill of $100,000/year or more in some cases, the medical community is forced to consider the sustainability of managing these cancers as chronic diseases, particularly in the context of an already burdened healthcare system. Additionally, prolonged and novel therapies come with associated cumulative toxicities that can negatively affect patient quality of life harkening us to recall "primum non nocere."

The value of assaying for biomarkers lies in providing us the ability to triage patients and therapy based on genetic information. Patients with a given disease who may do poorly may be selected for more aggressive therapy up front (e.g., allogeneic stem cell transplant in normal karyotype *FLT3*-mutated acute myeloid leukemia) [18]. The detection of key mutations also aids in the selection of effective targeted therapy (e.g., vemurafenib, a B-Raf enzyme inhibitor, in *BRAF*-mutated melanoma) and the avoidance of treatments with an unacceptable risk of adverse drug reactions (ADR) or ones to which the cancer is likely to be resistant

(e.g., *ABL* kinase domain mutations in CML) [19, 20]. Finally, detection of particular mutations also qualifies patients to any one of a number of ongoing clinical trials with gene-based targeted therapies (e.g., *IDH1* inhibitor trials for *IDH-mutated* AML) [21].

Analysis of tumor DNA to guide patient treatment has been in place for over 20 years. For example, acute lymphoblastic leukemia patients with t(9:22) are offered imatinib (or other second-generation *BCR/ABL1* tyrosine kinase inhibitor) along with chemotherapy, inducing molecular remission in nearly half of these patients in what was historically a uniformly fatal disease [22]. In solid tumors, a breast cancer patient with HER2 amplification (marking more aggressive disease) might be treated with the anti-HER2 monoclonal antibody trastuzumab or the HER2 tyrosine kinase inhibitor, lapatinib, resulting in increased complete response and possibly allowing us to sidestep the use of more cytotoxic anthracycline drugs [23–25]. Accordingly, targeted genetic profiling has become part and parcel of the routine patient management for select cancers (Table 38.1).

Initial identification of these therapy-related biomarkers has come from focused studies of biologically plausible candidate genes or genome-wide studies, with the former having proved more successful [13]. Rapidly declining sequencing costs have further spurred discovery of diagnostic, prognostic, and potentially therapeutic cancer biomarkers [26, 27].

Once discovered, biomarkers also need to be tested for as a part of patient care in a clinically reliable and cost-effective manner. Traditionally, clinical mutation testing was performed via "one-off" test assay formats by pyrosequencing, PCR-based methods, etc. Fortuitously, the same technology that has catapulted the rate of biomarker discovery, namely, next-generation sequencing has made simultaneous assessment of numerous genes clinically feasible. This is important because testing for multiple genes simultaneously by traditional methods would have been otherwise prohibitive. As such, CAP/CLIA-grade targeted gene mutation panels are already in place at academic cancer centers to facilitate the identification of somatic mutations associated with prognosis and/or to predict specific benefits (or lack thereof) from targeted therapies. Panels can be large pan-cancer panels (>400 genes) or focused (solid tumor vs. myeloid). Organ-specific panels (lung, CNS, etc.) also exist. For example, gene signature identified by a 70-gene or 21-gene panel test can be used to determine whether adjunct chemotherapy is needed in estrogen receptor-positive patients as well as recurrence risk [28].

With maturation of sequencing strategies with concomitant lowering of costs, there has been an increase in the application of these technologies to tumor profiling [29, 30]. While the current clinical focus is principally directed toward the identification of somatic DNA mutations of the coding regions of the genome, epigenetic traits including specific miRNAs, variations in RNA expression, methylation patterns, and chromatin markers may prove to be informative in the future, especially given the laudable efforts of the Encyclopedia of DNA Elements (ENCODE) project which has assigned biochemical functions for 80% of the genome, especially areas outside protein-coding regions [31]. Currently the most common type of genetic screening involves performing targeted DNA panels focused on a limited number of relevant candidate genes (isolated usually by PCR or target-capture beads, etc.) followed by sequencing [32]. NGS technology is capable of capturing the full range of genetic variation to include non-synonymous single nucleotide variants, translocations, insertion and deletions, and copy number change. Identified genetic variants are cross-referenced against the corpus of the medical literature to generate an annotated clinical report that may direct treatment to a particular biomarker or signaling pathway – information that could generally not be gleaned from tumor histology or immunohistochemistry alone. However, the large start-up and running costs, highly specialized expertise, and need for extensive bioinformatics support are all obstacles to routine application of NGS testing in clinical practice.

Expectedly, genomic interrogation of tumors has proven fruitful, in both research and the clinic. An initial deep sequencing of 145 genes in colorectal and non-small cell lung cancers found somatic mutations in 39/40 (98%) and 20/24 (83%) of tumors, respectively. More than half (52.5%) of colorectal cancers and 72% of non-small cell lung cancers contained at least one mutation that has been linked to a specific chemotherapy approach [32]. Similar data has come from the NCI/NHGRI Cancer Genome Atlas efforts across tumors from diverse anatomical locations [29, 30]. Clinical application of targeted NGS panels has been shown to frequently yield clinically relevant and potentially actionable genetic information in the hands of various groups [33–38].

Sequencing of therapy-resistant and/or responsive tumors has led to the discovery of several theranostic-specific biomarkers. Sequencing of non-small cell lung cancer which displayed sensitivity and subsequent resistance to EGFR tyrosine kinase inhibitors led to the routine use of *EGFR* sequencing to guide therapeutic choices [39, 40]. Whole exome sequencing of patients treated with everolimus for advanced bladder cancer revealed that a specific *TSC1* mutation correlated with everolimus sensitivity. Patients with *TSC1* mutation had a longer time until recurrence of tumor (4.1 versus 1.8 months) [41]. This loss of function mutation in *TSC1* was subsequently found in 5/96 (5.2%) of advanced bladder cancers, suggesting that there is a subgroup of patients with this disease for whom everolimus treatment might offer substantial benefit. However,

further clinical pharmacogenomic efforts to apply deep sequencing and other basic science methodologies to unveil the underlying mechanisms of sensitivity or resistance to drug therapy are needed given that the mechanism of clinical resistance for most anticancer drugs remains unclear.

Overall, the data show that somatic DNA mutation assessment frequently and positively impacts patient care for a wide gamut of cancers. The identification of KRAS mutations in codons 12 or 13 in ~30% of patients with colon cancer suggests that there is no tumor control benefit, but instead some toxicity risk, when patients are treated with expensive antibodies targeting EGFR [42]. Lung cancer, melanoma, and myeloproliferative disorders tend to be sensitive (or resistant in the case of most exon 20 EGFR-mutated lung cancers) to tyrosine kinase inhibitors with mutations in the *EGFR*, *BRAF*, and *JAK2*, respectively. Overall FDA-approved, gene-based targeted therapy comprise approximately 120 agents across cancer types [43]. On balance, currently there are molecular predictors of efficacy for less than 10% of the FDA-approved cancer drugs. Nonetheless, in an effort to leverage any advantage possible in the war against cancer, academic and commercial efforts focusing on disease-specific gene targets, such as the National Comprehensive Cancer Network, are actively developing guidelines for application in patient management [44]. As the costs decrease and the interpretation ability increases, somatic DNA assessment has quickly become a routine part of the management of cancer.

While the future of somatic profiling looks promising, the application of targeted therapy in clinical practice faces several challenges. First, it must be appreciated that cancer is not a static disease. Cancer cells are rapidly dividing and thus evolve. Reminiscent of antimicrobial drug resistance, cancer cell progeny may acquire mutations that confer resistance to specific drug treatments, or subclones may come to the fore as dominant clones are eradicated resulting in the proliferation of drug resistant tumors [45–47]. Moreover, beyond the initial patient treatment setting, there is a lack of personalized cancer medicine trial data to guide patient management decisions. Treatment choices often revert to the use of population average data, where it is difficult to ascertain the exact value of therapy for an individual patient. While we understand to a certain extent how specific genes may impact therapy response, our understanding of the manner in which other gene mutations and epigenetic modifiers alter and modify the clinical effect of any other mutation is still nascent. In this vein, it would be safe to say that no gene is an island [48]. Lastly, we still need more definitive discovery and validation efforts for anticancer drugs (old and new). Big data from NGS of large patient cohorts such as that from TCGA will serve this end [49]. However, scientific and financial biases may need to be overcome particularly for the older cytotoxic agents, which benefit a meaningful subset of patients, but do not have the diverse advocacy to assure that genomic knowledge is being procured and deployed in a clinically relevant manner.

Beyond biological and scientific considerations, there are also practical issues to attend to. For example, there are limits to how much tissue can be acquired from clinical biopsies and subsequently how much high-quality tumor DNA can be derived from such tissue for sequencing. Fortunately, NGS usually requires only a small amount of input DNA (on the order of a few hundred nanograms) and can be successfully performed even on cytology slides which usually do not contain many tumor cells [50]. Also, quality control issues and the complex bioinformatic algorithms underlying NGS analysis may result in uncertain or erroneous identification of mutations challenging interpretations of molecular diagnostic results from other clinical laboratories while adding confusion to the clinician's cauldron. Thankfully, this is greatly mitigated by the rigid validation and QC parameters for CAP/CLIA-grade NGS. However, the element of subjectivity and variation in analysis of complex data sets cannot be entirely avoided and is part of the art, not science, of pathology. Furthermore, predictive analyses are needed for the 25–80% of cases where variants of unknown significance are identified in genes that are of interest to a particular tumor. As increasing numbers of tumors are sequenced clinically (the numbers of which are likely to quickly outstrip research-grade resequencing studies), it is hoped that patterns of clinical relevance will emerge for these variants prospectively. This will also require laboratory and clinical consensus methods to decide which variants in which genes merit clinical actionability. This has been evident in BRCA1/2 testing, associated with breast cancers [51].

Tumor profiling is also changing the way in which we think about cancer. For example, clinical trial inclusion criteria are starting to focus less on the anatomical origins (breast cancer or lung cancer) and more on the somatic mutations identified within a tumor (e.g., *KRAS-mutated* cancer regardless of site), so-called basket trials. The focus on "driver" mutations spurring tumor invasiveness and therapeutic response requires screening many patients to find the few that are eligible for targeted therapy trials [41]. This may be advantageous as those unlikely to respond to a particular therapy may be excluded up front – a desirable selection bias. Currently a disproportionate number of patients are triaged for experimental therapy without a previously ascertained molecular profile introducing additional time and unsupported expense before assignment to a treatment trial of relevance.

Nonneoplastic Applications of Pharmacogenomics

While the most robust use of pharmacogenomics has been in the oncology space, there a several applications that exist outside of this area as well, most notably in psychiatry/neurology and cardiovascular disease.

In psychiatry, there is a wide variability in patient's response to psychotropic drugs with only 50–60% of patients showing adequate response [52]. The need for effective pharmacogenomics is amplified by the long process of drug assessment, titration, and potentially avoidable side effects that come with many psychiatric medications. It is thought that up to 50% of antidepressant response can be attributable to genetic factors although meta-analysis of GWAS studies has not shown any reliable predictors of treatment outcome [53–55]. Nonetheless over the past two decades, the growing body of knowledge with regard to pharmacokinetics psychiatric drugs based on genetic variation has been incorporated into dosing or drug selection guidelines by professional groups such as the Clinical Pharmacogenetics Implementation Consortium (CPIC). There are more than 30 psychiatric drugs with FDA drug labeling that includes pharmacogenomic biomarkers, mostly involving CYP2D6 (followed by CYP2C19) where poor metabolizers may demonstrate increased plasma concentrations of drug [56]. As an example, a 50% dose reduction in tricyclic antidepressants amitriptyline and nortriptyline is suggested for CYP2D6 or CYP2C19 poor metabolizers [57]. In bipolar disease and epilepsy, *HLA-B*1502* genotyping is warranted in Asian patients due to the risk of potentially fatal dermatologic ADRs due to use of carbamazepine or phenytoin [56, 58]. While there has been significant advances made, the lack of clear data showing clinical benefit to routine pharmacogenomic testing in psychiatric patients remains an obstacle to widespread implementation and reimbursement for genetic screening.

Cardiovascular disease is the leading cause of death worldwide and, like psychiatric disease, involves a complex interplay of both genetic and environmental factors. The application of pharmacogenomics to cardiovascular disease largely involves the impact of polymorphisms on patient's response to statins and anticoagulating agents [59]. Pharmacogenetic-guided dosing or an alternate drug usage may be considered for warfarin (*CYP2C9* and *VKORC1*) and clopidogrel (*CYP2C19* intermediate or poor metabolizers). No recommendations have been made for beta-blockers, but individuals with intermediate or low function of SLCO1B may require consideration of an alternative to simvastatin to avoid statin-induced myopathy [60]. However, due to the relative dearth of randomized control trials delineating clinical utility and few options when it comes to alternative drugs, there remains controversy regarding routine testing resulting in varying society guidelines and recommendations.

In infectious disease, there is an additional layer of complexity to implementing pharmacogenomics given that the genomes of both the host and pathogen must be considered [61]. Genetic variation in the pathogen may help predict antimicrobial resistance, and polymorphisms in the host may affect susceptibility to infection, response to treatment, and development of ADRs. For example, screening for HLA-B5701 was shown to reduce the incidence of abacavir-related hypersensitivity [62]. Patients with IL28B variants show variable response to pegylated interferon-alpha and ribavirin therapy for hepatitis C, while HCV genotype also significantly impacts response to these therapies (sustained virologic response 40–50% for genotype 1 virus versus 70–80% for genotypes 2 and 3) [63].

Impact of Germline DNA Variation on Dose Optimization and Toxicity Risk

The other head of the pharmacogenomics dragon is that of germline genetic variation, which often plays second fiddle to somatically mutated genes in cancer genomic research but can nonetheless significantly impact cancer treatment. Indeed, given that most of oncology supportive care is targeted toward mollifying the adverse effects of cytotoxic therapy, genetic variations that predict toxicity can potentially play a vital role in the selection and administration of cancer drugs. The uneven focus on achieving tumor control from a specific drug may be due in part to the lack of objective data elucidating the risk of developing severe adverse drug effects.

ADR affect millions of individuals at a cost of >$100 billion/year and accounting for more than 100,000 deaths [64]. ADRs can be the result of overmedication or can be idiosyncratic. It is thought that the latter may be, at least in part, explained on the basis of individual genetic variation [65].

In this vein, the role of gastrointestinal drug transport and hepatic drug metabolism on the dose, administration schedule, and route of administration of a drug is often undervalued. Along with C*YP2D6*, highly polymorphic human cytochrome P450 family of enzymes (including *CYP2C9*, *CYP2C19*, *CYP3A4*, and *CYP3A5*) are responsible for metabolism of most clinically used drugs [66]. In fact, C*YP2D6* alone is responsible for the metabolism of approximately a quarter of marketed drugs [16]. As an example, highly effective tyrosine kinase inhibition therapy for chronic myelogenous leukemia can be rendered variably impotent if the drug is metabolized or removed before it encounters tumor. Moreover, such underdosing is a setup for development of drug resistance [67]. More than 70 variant alleles have been found in the gene encoding CYP2D6, a liver

enzyme responsible for the metabolism of certain antidepressants, antipsychotic, and antiarrhythmic drugs. Using DNA chip microarrays, most patients can be classified as ultrarapid, extensive, intermediate, and poor metabolizers which could potentially inform dosing and drug choice [68].

The role of pharmacogenomic variation in drug metabolism and the efficacy of certain anticancer therapies are well illustrated by tamoxifen, an effective antiestrogen used in the treatment of hormone receptor-positive breast cancer. Bioconversion of tamoxifen to active metabolites, including endoxifen (most abundant active metabolite), is mostly dependent on the highly polymorphic cytochrome CYP2D6 enzyme. In the past, patients on tamoxifen also received antidepressants to manage antiestrogen-induced hot flashes. However, it was subsequently discovered that the antidepressants blocked CYP2D6 and the production of endoxifen [7]. Co-administration was discontinued as it made little sense to mitigate side effects of the primary drug if such mitigation entailed abrogation of its efficacy [8]. Clinical trials of endoxifen levels as a biomarker for tamoxifen response are now underway [69].

Approximately 7% of the US population has a genetic polymorphism or deletion in *CYP2D6*, resulting in diminished protein levels and/or function. Over 20 published studies have reported an association between *CYP2D6* polymorphisms and breast cancer outcomes after tamoxifen treatment, and several recent studies suggest that homozygote patients with reduced to absent CYP2D6 function have the poorest outcome [70]. Studies to determine the appropriate dose of tamoxifen for wild-type patients (called extensive metabolizers), and the heterozygous patients (intermediate metabolizers), suggest that the doubling of tamoxifen dose in intermediate metabolizers normalized plasma endoxifen levels to that observed in extensive metabolizers [71, 72]. Such data is usually required for the FDA prescribing recommendations for dose adjustment after organ dysfunction, drug interaction, or age and suggests a relevance to CYP2D6-guided tamoxifen dosing to avoid underdosing patients with breast cancer.

Thiopurine drugs that are used in the treatment of acute lymphoblastic leukemia as well as inflammatory bowel disease offer another example of pharmacogenomic effect on drug metabolism. These drugs have a narrow therapeutic window and can result in life-threatening bone marrow suppression. Thiopurines are inactivated by thiopurine S-methyltransferase (TPMT). Patients with the *TPMT*3A* phenotype produce a rapidly degraded gene product resulting in little or no detectable TPMT and consequently increased risk for life-threatening myelosuppression with standard dosing (such individuals need approximately 1/10th the standard dose). TPMT enzyme activity can also be directly measured in red blood cells. The advantage of ascertainment at the phenotype level (enzyme activity) is that rare genetic variants are not missed [64]. However, even when dosing is adjusted for TPMT, genetic variation in other genes such as *ITPA* becomes important [1]. Further complicating matters, there are also ethnic differences with mutations in *NUDT15* being the major determinant of thiopurine intolerance in patients of Asian or Native American ancestry rather than *TPMT*. This supports therapeutic prediction strategies that allow for contribution of multiple relevant genes, not just a "one gene at a time" approach.

Acknowledgment of the need for and value of genetically based predictive models for drug toxicity in cancer care has prompted the launch of genomic discovery programs that focus on adverse effects of cancer drugs such as sensory peripheral neuropathy, cardiotoxicity, hearing loss, and other toxicities [73–76]. For example, paclitaxel-induced neuropathy is a common adverse event often resulting in discontinuation of therapy and patient discomfort [77]. Till recently, prospective identification of patients at increased risk for neuropathy for whom choices of drug, administration schedule, and quality of life could be informed was not possible. Fortunately, both candidate gene and genome-wide association studies have begun to offer some inroads. A genome-wide association study was performed on 855 subjects of European ancestry who received paclitaxel for lymph node-negative breast cancer. This study identified a single nucleotide polymorphism in the *FGD4*, (FGD1-related F-actin-binding protein) that was associated with the onset of sensory peripheral neuropathy in the discovery cohort [Hazard ratio = 1.57; 95% CI 1.30–1.91]. This polymorphism was also observed in a European and African-American replication cohort [73]. As a congenital peripheral neuropathy gene, *FGD4* provides biologic plausibility to further assess the contribution of genetic variation to the development of peripheral neuropathy.

Another study identified breast cancer patients with nearly double the risk of paclitaxel-induced neuropathy based on *CYP2C8*3* status [74]. Although less common, *CYP2C8*3* was also found in African-Americans. The association of increased risk of paclitaxel-induced neuropathy with the *CYP2C8*3* variant across racially distinct patient population suggests that both pharmacokinetic variables (such as *CYP2C8*) and biologic variation (such as *FGD4*) contribute to patient risk of neuropathy.

Avoiding chemotherapy-associated morbidity is even more critical in the context of adjuvant chemotherapy, where the goal is to kill cancer cells that may be lurking at undetectable levels. This is simply because slaying bone fide tumor cells provides a more cogent rationale to bear toxicity than the staving off the possibility of relapse. This is further magnified in the treatment of childhood malignancies where survivorship may be high forcing the patient to endure adverse drug events for many years to come.

Overall, there are only 16 genes considered to house clinically actionable pharmacogenomic germline variation. However, these have far-reaching impact affecting 7% of medications and 18% of all prescriptions [1]. Table 38.2 summarizes several pharmacogenomically actionable germline DNA variants. While encouraging, this also means that the vast majority of prescribed drugs is still not currently optimized based on genetic testing.

There are significant limitations to pharmacogenomics discovery (both germline and somatic) for anticancer therapies, including the challenges of recruiting large patient cohorts for purposes of discovery and validation. Furthermore, it often takes 7–10 years to construct, conduct, and analyze a clinical trial for pharmacogenomics discovery.

Table 38.2 Pharmacogenomically actionable germline DNA markers

Genetic variation	Medications	Comments
TPMT	Mercaptopurine, thioguanine, azathioprine	Myelosuppression
CYP2D6	Codeine, oxycodone, tramadol, tricyclic antidepressants	Ultrametabolizers at risk for codeine-associated respiratory depression
VKORC1	Warfarin	Accounts for ~30% of warfarin variance
CYP2C9	Warfarin, phenytoin	Accounts for ~10% of warfarin variance
UGT1A1	Irinotecan, atazanavir	Variable number of TA dinucleotide repeats in promoter region. Drug-induced neutropenia and grades 3–4 diarrhea, decreased metabolism of irinotecan
SLCO1B1	Simvastatin	Myopathy
NAT2	Isoniazid	Autosomal recessive; slow acetylator vs. rapid acetylator phenotypes
CYP2C19	Clopidogrel, voriconazole, proton pump inhibitors	Clopidogrel: Increased risk of cardiovascular events due to decreased platelet inhibition. Voriconazole: Increased risk of fungal infection due to ultrarapid metabolism
*HLA-B*57:01*	Abacavir; flucloxacillin	Life-threatening hypersensitivity syndrome with abacavir; flucloxacillin-associated hepatotoxicity
DPYD	Fluorouracil	DPYD*2A associated with severe, life-threatening toxicity
*HLA-B*15:02*	Carbamazepine	Stevens-Johnson syndrome and toxic epidermal necrolysis risk
ADRB2	Albuterol	Arg16/Arg16 homozygotes show more robust bronchodilation
CYP3A5	Tacrolimus	May require higher dosing
IFNL3 (IL28B)	Interferon	Variant allele associated with poor response

Adapted from [1, 2, 16]

The same is true for a validation cohort, which partially explains why there are so few validated discoveries in the literature. There is also a paucity of information on the heritability of anticancer drug effects, which are needed to justify our search for a genomic basis to explain and predict drug effect variability. One recent approach is the use of cell lines from large, multigeneration families. Work has shown a wide variation in heritability of cytotoxicity (10–70%) across 29 commonly prescribed anticancer drugs, with 66% having greater than 30% heritability [78]. This approach presents an opportunity for a more efficient prioritization of drug assessment and preemptive ex vivo discovery, conserving precious clinical material for validation studies. To this effect, the application of bar coding and robotics has allowed the scale-up of cell line phenotyping to 500–1000 cell lines per project, followed by ex vivo genome-wide association studies [78–80]. Innovative pharmacogenomic strategies will allow us to more rapidly capture the relevance of genomic information for rational drug therapy selection.

Another limitation lies in the fact that statistically significant scientific research does not always translate to clinical utility. For example, single nucleotide polymorphisms in *LPA*, *APOE*, *SLCO1B1*, *SORT1*, and *ABCG2* are associated with LDL response. Unfortunately, no consistent risk reduction for cardiovascular events has been noted [2]. Furthermore, even in cases with demonstrated clinical utility, practical limitations such as drug-drug interactions, drug dosage, contribution of epigenetic factors (which are dynamic in nature), along with physiologic and environmental factors challenge clinical implementation of pharmacogenomic discoveries. These factors likely underlie the lack a perfect genotype-phenotype correlation [3].

Moving into Clinical Practice

Bringing pharmacogenomics to the clinic faces a host of challenges. Scientific discoveries must undergo independent validation prior to introduction as a clinical diagnostic test. This is particularly true given that population-specific ancestral variation as well as the effect of multiple genetic variants in the same or other genes may impact and confound the translation of individual genetic variation into a clinically actionable pharmacogenomic trait [1]. Further complicating the issue is the fact that SNPs in noncoding regions can affect drug phenotypes. Optimally informed therapeutic decision-making will likely require integration of transcriptome, proteome, metabolome data along with their complex interplay with environmental factors [81].

While independent validation of pharmacogenomic discoveries is necessary, it is also challenging. This is due to the fact that it is often difficult to characterize, uniformly treat,

and systematically evaluate patients to objectively quantify the drug response phenotype. Ideally, we would obtain genomic DNA from all patients entered into clinical drug trials, along with consent for pharmacogenomics studies. This occurs in most large trials being conducted by pharmaceutical industry and some NCI clinical trial groups [82, 83]. However, it has not yet become standard for academic or foundation-supported trials, perhaps due to the cost and support issues.

Even if the collection of genomic information were to become the standard of care, this does not entail smooth transition of this data to clinical practice. There is a "yin-yang" between applying newly acquired data and making sure that the data we are applying is reliable and relevant. Pushing for the former is the need for more personalized care to improve patient outcomes. With regard to the latter, there has typically been a reliance on prospective, randomized, controlled trials to justify clinical implementation. However, this may not be practical as it introduces a 5–10-year lag, while such studies are completed prior to integration into clinical practice. Moreover, with the ability to perform next-generation sequencing on archived FFPE specimens and available biorepositories across the country, the value of large, robust, albeit retrospective, data should be considered. Such retrospective analyses suffer from lack of homogeneous treatment regimens and polypharmacy; although this could conceivably be surmounted by multiinstitutional efforts for target patient population when sufficient clinical data is available. With the prospect in the not too distant future of individuals having the entire germline genome sequenced, the need for efforts to develop standardized genomically informed prescribing guidelines is underscored. One such effort is the CPIC, which includes participants from >80 institutions across 4 continents [84–86]. A key element to programs such as CPIC is the realization that there are some aspects of the medical decision process, such as drug dosing, that have robust data on benefit to patients, even as the field waits for the "perfect" studies that definitively guide therapy at a broader level. Internationally, the European Pharmacogenetics Implementation Consortium and Royal Dutch Pharmacist Association are two bodies who are also looking to bring pharmacogenomic data to bear on clinical care.

Another obstacle is the hesitancy with which the medical community approaches genetic data. Given the explosive nature and rapid rate at which genomic data is being accumulated, there is an understandable cautiousness regarding the adoption of these novel data. This has led to a double standard in clinical decision-making. For example, a drug interaction may be accepted as a credentialed, clinically relevant variable and rapidly integrated into clinical practice. On the other hand, acceptance of genetic data through the exact same mechanism may be delayed due to the desire for first accumulating of large amounts of prospective data. This is occurring for *CYP2D6* and tamoxifen, *CYP3A4* and taxane chemotherapy, and related interactions for supportive care medications. Years of familiarity and low patient expense have eased the standards for drug interaction even though these carry less functional predictability than gene deletion in the same pathway. Moving forward, we will need to implement protocols through which variations in clinically credentialed pathways can more easily be moved into the clinic.

Clinically relevant pharmacogenomic testing must demonstrate analytical validity, clinical validity, and clinical utility. CAP and CLIA accreditation in the USA ensure the high-quality of molecular data; however, due to the high complexity of next-generation sequencing testing and bioinformatics, interlaboratory variability must still be reckoned with. Clinical validity and utility can be derived from retrospective, mechanistic, preclinical, and clinical studies. Actionability and thus clinical utility are influenced by penetrance, therapeutic index, severity of drug toxicity, severity of disease, as well as the presence or absence of alternative therapies.

Traditionally, the focus of pharmacogenomics has been to explain adverse events, predict response, define dosing criteria, or preempt severe drug reactions. Other, currently neglected endpoints could serve as additional drivers of early adoption of new health-care modalities. These include mitigation of 30-day readmission rates, economics of quality-driven health-care metrics and "bundled care" and the prioritization of medication access by a health system pharmacy and therapeutics committee. These endpoints are often accessible through observational cohorts or electronic health record studies. Coupled with growing familiarity with genomic medicine, this will likely drive the implementation of pharmacogenomics into practice. There are also potential cost savings with one study demonstrating savings of approximately $621 per polypharmacy patient per year based on pharmacogenomically guided prescription recommendations [87].

The translation of pharmacogenomic data into practical clinical applications remains slow. Obstacles include lack of clarity on standards for clinical utility, cost and complexity of testing, limited number of prospective pharmacogenomics studies, and paucity of clear guidelines regarding how to implement genetic information (both well-documented pharmacogenes and rare variants). Research funding, reimbursement, and liability issues bring additional challenges to integrating genetic testing routinely in the management of cancer care. However, it would appear that genomic medicine is here to stay, and it would appear that the further we move forward, the clearer the path will become.

References

1. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015;526(7573):343–50.
2. Zhang G, Nebert DW. Personalized medicine: genetic risk prediction of drug response. Pharmacol Ther. 2017;175:75–90.
3. Weinshilboum R, Wang L. Pharmacogenomics: bench to bedside. Discov Med. 2005;5(25):30–6.
4. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. N Engl J Med. 2011;364(12):1144–53.
5. Lehmann H, Ryan E. The familial incidence of low pseudocholinesterase level. Lancet. 1956;271(6934):124.
6. Meyer UA. Pharmacogenetics and adverse drug reactions. Lancet. 2000;356(9242):1667–71.
7. Stearns V, et al. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. J Natl Cancer Inst. 2003;95(23):1758–64.
8. Kelly CM, et al. Selective serotonin reuptake inhibitors and breast cancer mortality in women receiving tamoxifen: a population based cohort study. BMJ. 2010;340:c693.
9. Cancer Genome Atlas Research Network. Electronic address, w.b.e. and N. Cancer Genome Atlas Research. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 2017;169(7):1327–1341 e23.
10. Cancer Genome Atlas Research N, et al. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017;543(7645):378–84.
11. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576–82.
12. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(7216):1061–8.
13. Weng L, et al. Pharmacogenetics and pharmacogenomics: a bridge to individualized cancer therapy. Pharmacogenomics. 2013;14(3):315–24.
14. United States Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labeling.https://www.fda.gov/Drugs/ScienceResearch/ucm572698.htm./Last updated 8/3/2018.
15. M. Whirl-Carrillo, et al. "Pharmacogenomics Knowledge for Personalized Medicine" Clinical Pharmacology & Therapeutics (2012) 92(4):414–17.
16. Lee JW, et al. The emerging era of pharmacogenomics: current successes, future potential, and challenges. Clin Genet. 2014;86(1):21–8.
17. Lee SY, McLeod HL. Pharmacogenetic tests in cancer chemotherapy: what physicians should know for clinical application. J Pathol. 2011;223(1):15–27.
18. Pratz KW, Levis M. How I treat FLT3-mutated AML. Blood. 2017;129(5):565–71.
19. Chapman PB, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011;364(26):2507–16.
20. Soverini S, et al. BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. Blood. 2011;118(5):1208–15.
21. Fujii T, et al. Targeting isocitrate dehydrogenase (IDH) in cancer. Discov Med. 2016;21(117):373–80.
22. Pui CH, Evans WE. A 50-year journey to cure childhood acute lymphoblastic leukemia. Semin Hematol. 2013;50(3):185–96.
23. Vici P, et al. Outcomes of HER2-positive early breast cancer patients in the pre-trastuzumab and trastuzumab eras: a real-world multicenter observational analysis. The RETROHER study. Breast Cancer Res Treat. 2014;147(3):599–607.
24. Xu ZQ, et al. Efficacy and safety of lapatinib and trastuzumab for HER2-positive breast cancer: a systematic review and meta-analysis of randomised controlled trials. BMJ Open. 2017;7(3):e013053.
25. Denduluri N, et al. Selection of optimal adjuvant chemotherapy regimens for human epidermal growth factor receptor 2 (HER2) -negative and adjuvant targeted therapy for HER2-positive breast cancers: an American Society of Clinical Oncology Guideline Adaptation of the Cancer Care Ontario Clinical Practice Guideline. J Clin Oncol. 2016;34(20):2416–27.
26. Braggio E, et al. Lessons from next-generation sequencing analysis in hematological malignancies. Blood Cancer J. 2013;3:e127.
27. Dewitt ND, Yaffe MP, Trounson A. Building stem-cell genomics in California and beyond. Nat Biotechnol. 2012;30(1):20–5.
28. Arranz EE, et al. Gene signatures in breast cancer: current and future uses. Transl Oncol. 2012;5(6):398–403.
29. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489(7417):519–25.
30. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330–7.
31. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
32. Lipson D, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012;18(3):382–4.
33. Goodman AM, Choi M, Wieduwilt M, Mulroney C, Costello C, Frampton G, Miller V, Kurzrock R. Next-generation sequencing reveals potentially actionable alterations in the majority of patients with lymphoid malignancies. JCO Precis Oncol. 2017;1:1–13.
34. Muller KE, et al. Targeted next-generation sequencing detects a high frequency of potentially actionable mutations in metastatic breast cancers. Exp Mol Pathol. 2016;100(3):421–5.
35. Vasan N, et al. A targeted next-generation sequencing assay detects a high frequency of therapeutically targetable alterations in primary and metastatic breast cancers: implications for clinical practice. Oncologist. 2014;19(5):453–8.
36. Blumenthal DT, et al. Clinical utility and treatment outcome of comprehensive genomic profiling in high grade glioma patients. J Neuro-Oncol. 2016;130(1):211–9.
37. Rankin A, et al. Broad detection of alterations predicted to confer lack of benefit from EGFR antibodies or sensitivity to targeted therapy in advanced colorectal cancer. Oncologist. 2016;21:1306.
38. Hagemann IS, et al. Diagnostic yield of targeted next generation sequencing in various cancer types: an information-theoretic approach. Cancer Genet. 2015;208(9):441–7.
39. da Cunha Santos G, Shepherd FA, Tsao MS. EGFR mutations and lung cancer. Annu Rev Pathol. 2011;6:49–69.
40. Paez JG, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–500.
41. Kreso A, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. Science. 2013;339(6119):543–8.
42. Amado RG, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26(10):1626–34.
43. Abramson, R. 2018. Overview of Targeted Therapies for Cancer. My Cancer Genome https://www.mycancergenome.org/content/molecular-medicine/overview-of-targeted-therapies-for-cancer/ (Updated May 25).
44. Engstrom PF, et al. NCCN molecular testing white paper: effectiveness, efficiency, and reimbursement. J Natl Compr Cancer Netw. 2011;9(Suppl 6):S1–16.

45. Walter MJ, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia. 2013;27(6):1275–82.
46. Walter MJ, et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med. 2012;366(12):1090–8.
47. Jacoby MA, Duncavage EJ, Walter MJ. Implications of tumor clonal heterogeneity in the era of next-generation sequencing. Trends Cancer. 2015;1(4):231–41.
48. Hussaini M. Biomarkers in hematological malignancies: a review of molecular testing in hematopathology. Cancer Control. 2015;22(2):158–66.
49. Cancer Genome Atlas Research N, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059–74.
50. Velizheva NP, et al. Cytology smears as excellent starting material for next-generation sequencing-based molecular testing of patients with adenocarcinoma of the lung. Cancer Cytopathol. 2017;125(1):30–40.
51. Berg JS, et al. An informatics approach to analyzing the incidentalome. Genet Med. 2013;15(1):36–44.
52. Zhou SF, et al. Clinical pharmacogenetics and potential application in personalized medicine. Curr Drug Metab. 2008;9(8):738–84.
53. Investigators G, Investigators M, Investigators SD. Common genetic variation and antidepressant efficacy in major depressive disorder: a meta-analysis of three genome-wide pharmacogenetic studies. Am J Psychiatry. 2013;170(2):207–17.
54. Tansey KE, et al. Contribution of common genetic variants to antidepressant response. Biol Psychiatry. 2013;73(7):679–82.
55. Reynolds GP. The pharmacogenetics of symptom response to antipsychotic drugs. Psychiatry Investig. 2012;9(1):1–7.
56. Drozda K, Muller DJ, Bishop JR. Pharmacogenomic testing for neuropsychiatric drugs: current status of drug labeling, guidelines for using genetic information, and test options. Pharmacotherapy. 2014;34(2):166–84.
57. Hicks JK, et al. Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. Clin Pharmacol Ther. 2013;93(5):402–8.
58. Leckband SG, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. Clin Pharmacol Ther. 2013;94(3):324–8.
59. Berinstein E, Levy A. Recent developments and future directions for the use of pharmacogenomics in cardiovascular disease treatments. Expert Opin Drug Metab Toxicol. 2017;13(9):973–83.
60. Giudicessi JR, Kullo IJ, Ackerman MJ. Precision cardiovascular medicine: state of genetic testing. Mayo Clin Proc. 2017;92(4):642–62.
61. Aceti A. Pharmacogenomics for infectious diseases. J MedMicrob Diagn. 2016;5:e223.
62. Young B, et al. First large, multicenter, open-label study utilizing HLA-B*5701 screening for abacavir hypersensitivity in North America. AIDS. 2008;22(13):1673–5.
63. Soon-U LL, Amur S. Chapter 12: "Pharmacogenomics and pharmacogenetics for infectious diseases." in Pharmacogenomics: an introduction and clinical perspective. McGraw Hill, New York, NY, USA. 2013.
64. Adams JU. Pharmacogenomics and personalized medicine. Nat Educ. 2008;1:194.
65. Daly AK. Pharmacogenomics of adverse drug reactions. Genome Med. 2013;5(1):5.
66. Dong Y, et al. Analysis of genetic variations in CYP2C9, CYP2C19, CYP2D6 and CYP3A5 genes using oligonucleotide microarray. Int J Clin Exp Med. 2015;8(10):18917–26.
67. Eechoute K, et al. A long-term prospective population pharmacokinetic study on imatinib plasma concentrations in GIST patients. Clin Cancer Res. 2012;18(20):5780–7.
68. Ma Q, Lu AY. Pharmacogenetics, pharmacogenomics, and individualized medicine. Pharmacol Rev. 2011;63(2):437–59.
69. Ahmad A, et al. Endoxifen, a new cornerstone of breast cancer therapy: demonstration of safety, tolerability, and systemic bioavailability in healthy human subjects. Clin Pharmacol Ther. 2010;88(6):814–7.
70. Hertz DL, McLeod HL, Irvin WJ Jr. Tamoxifen and CYP2D6: a contradiction of data. Oncologist. 2012;17(5):620–30.
71. Irvin WJ Jr, et al. Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: a multicenter study. J Clin Oncol. 2011;29(24):3232–9.
72. Walko CM, McLeod H. Use of CYP2D6 genotyping in practice: tamoxifen dose adjustment. Pharmacogenomics. 2012;13(6):691–7.
73. Baldwin RM, et al. A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. Clin Cancer Res. 2012;18(18):5099–109.
74. Hertz DL, et al. CYP2C8*3 predicts benefit/risk profile in breast cancer patients receiving neoadjuvant paclitaxel. Breast Cancer Res Treat. 2012;134(1):401–10.
75. Ross CJ, et al. Genetic variants in TPMT and COMT are associated with hearing loss in children receiving cisplatin chemotherapy. Nat Genet. 2009;41(12):1345–9.
76. Visscher H, et al. Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. J Clin Oncol. 2012;30(13):1422–8.
77. McWhinney SR, Goldberg RM, McLeod HL. Platinum neurotoxicity pharmacogenetics. Mol Cancer Ther. 2009;8(1):10–6.
78. Peters EJ, et al. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. Pharmacogenomics. 2011;12(10):1407–15.
79. McLeod HL. Cancer pharmacogenomics: early promise, but concerted effort needed. Science. 2013;339(6127):1563–6.
80. Cox NJ, et al. Clinical translation of cell-based pharmacogenomic discovery. Clin Pharmacol Ther. 2012;92(4):425–7.
81. Weinshilboum RM, Wang L. Pharmacogenomics: precision medicine and drug response. Mayo Clin Proc. 2017;92(11):1711–22.
82. Ratain MJ, et al. The cancer and leukemia group B pharmacology and experimental therapeutics committee: a historical perspective. Clin Cancer Res. 2006;12(11 Pt 2):3612s–6s.
83. Innocenti F, et al. A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. Clin Cancer Res. 2012;18(2):577–84.
84. Relling MV, Klein TE. CPIC: clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. Clin Pharmacol Ther. 2011;89(3):464–7.
85. Hicks JK, et al. Patient decisions to receive secondary pharmacogenomic findings and development of a multidisciplinary practice model to integrate results into patient care. Clin Transl Sci. 2018;11(1):71–6.
86. Hicks JK, et al. Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. Clin Pharmacol Ther. 2016;102:37.
87. Saldivar JS, et al. Initial assessment of the benefits of implementing pharmacogenetics into the medical management of patients in a long-term care facility. Pharmgenomics Pers Med. 2016;9:1–6.
88. Lopez-Lopez E, et al. Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia. Pediatr Blood Cancer. 2011;57(4):612–9.
89. Wheeler HE, et al. Cancer pharmacogenomics: strategies and challenges. Nat Rev Genet. 2013;14(1):23–34.
90. van Staveren MC, et al. Evaluation of predictive tests for screening for dihydropyrimidine dehydrogenase deficiency. Pharmacogenomics J. 2013;13(5):389–95.
91. Kalia M. Biomarkers for personalized oncology: recent advances and future challenges. Metabolism. 2015;64(3 Suppl 1):S16–21.

92. Patil SA, et al. Novel approaches to glioma drug design and drug screening. Expert Opin Drug Discovery. 2013;8(9):1135–51.
93. Ranieri G, et al. Vascular endothelial growth factor (VEGF) as a target of bevacizumab in cancer: from the biology to the clinic. Curr Med Chem. 2006;13(16):1845–57.
94. Konstantinopoulos PA, et al. Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. J Clin Oncol. 2010;28(22):3555–61.

The Human Microbiome in Health and Disease

39

Wyatt M. Arnold, Elle Simone Hill, Na Fei, Alyson L. Yee, Mariana Salas Garcia, Lauren E. Cralle, and Jack A. Gilbert

Introduction

The bacteria, archaea, fungi, and viruses that populate every facet of our external and internal surfaces compose what is known as the human microbiome. Colonization by these microbes is so extensive that the human body consists of roughly an equal number of human and microbial cells [1]. The bacterial portion of this cohort is especially diverse, with an estimated 500–1,000 distinct species existing on or in the average person [2]. Taking into consideration the number of strains, or subspecies, of bacteria, that count could reach up to 10,000 or more unique organisms [3]. Beyond their sheer numbers, these bacterial taxa also offer substantially more genetic diversity than what is contained within the human genome. While the average human genome consists of about 27,000 genes, the collective genome of our microbiomes dwarves that, with up to seven million unique genes. Yet different people harbor radically different collections of microbes, and very little is understood about what causes and controls this variation. While we know the microbiome can regulate health and disease, the diagnostic significance of temporal or interpersonal variations in microbial community composition is still obscure. We are now at an inflection point in the study of the microbiome, as the focus shifts away from description and survey and toward more in-depth investigations and the development of applied therapies. Here we present the current state of knowledge that links the microbiome to human health and disease development, and in doing so hopefully provide a platform for future exploration into applied clinical research.

A common misconception is that our microbes outnumber our own cells by 10:1. This number stems from a 1972 article which uses a "back of the envelope calculation" to arrive at this ratio [4]. A perhaps more precise comparison is provided by Rosner [5], who asserts that the human body is made up of 724×10^{12} human cells, which are accompanied by $30–400 \times 10^{12}$ bacterial cells. More recently, a refined estimate based on experimental observation and extrapolation settles on a ratio of 1.3 bacterial cells to every one human cell [1]. Yet, while these estimates reduce the extent to which microbial cells outnumber human cells, they do not reduce the estimates associated with the diversity of the microbiome. Human-associated bacteria and other microbes—archaea, fungi, and viruses—are extremely diverse. A rough estimate of diversity, based on the assumption of 1,000 bacterial species in the gut and 2,000 genes per species, yields an estimate of 2,000,000 bacterial genes. While this calculation is for the gut alone, it still results in a number roughly 100 times greater than the number (approximately 27,000) of human genes [2]. Moreover, these numbers are in agreement with the observed size of microbial gene catalogs obtained by microbiome projects, such as MetaHIT [6] and the Human Microbiome Project [7].

Observing the Microscopic: How We Study the Microbiome

While there are many tools and techniques that can lend insight into the ecology of microbial communities, few have proven more useful than 16S rRNA sequencing. The 16S rRNA gene, found within the small ribosomal subunit of prokaryotes, is valuable as it contains both highly conserved *and* hypervariable regions. This combination of stability and

W. M. Arnold · L. E. Cralle
Department of Surgery, University of Chicago, Chicago, IL, USA

Biosciences Division (BIO), Argonne National Laboratory, Lemont, IL, USA

E. S. Hill · N. Fei · A. L. Yee · M. S. Garcia
Department of Surgery, University of Chicago, Chicago, IL, USA

Surgery, University of Chicago Medical Center, Chicago, IL, USA

J. A. Gilbert (✉)
Department of Surgery, University of Chicago, Chicago, IL, USA

Biosciences Division (BIO), Argonne National Laboratory, Lemont, IL, USA

Surgery, University of Chicago Medical Center, Chicago, IL, USA

Marine Biological Laboratory, Woods Hole, MA, USA

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8_39

stochasticity means that, in biochemistry, the fragment can easily be targeted by nonspecific primers, yet the internal variability still allows for it to be used to accurately identify bacteria down to the species level [8]. Thus, 16S rRNA amplicon sequencing provides a robust tool for identification, classification, and even discovery of bacteria [9]. Moreover, as the 16S region is only found in prokaryotes (and the organelles descended from them), primers targeting the region overlook any eukaryotic taxa present in a sample. While this property of the 16S region often proves useful for studies targeting only bacteria and archaea, it means that other regions must be targeted in order to survey eukaryotes. For most eukaryotic organisms, the 18S region—again part of the small ribosomal subunit—acts as the equivalent to the 16S region, though for fungi the ITS region is often targeted in addition to, or instead of, the 18S region.

In a typical study utilizing 16S rRNA, sequence data is used to create an ecological profile for each sample analyzed. In its most basic form, this means that 16S rRNA sequencing data will describe what bacterial species are present in the samples and how they compare in relative abundance. When differences in bacterial communities are observed between sample, the data can be explored in greater depth to understand whether correlations exist between bacterial composition and sample metadata. For example, if differences arise between the gut microbiomes of children born to obese mothers and children born to standard-weight mothers, we could draw upon statistical tools to understand if maternal weight at birth is correlated with microbiome composition [10]. Such studies, which draw extensively upon 16S rRNA sequencing, can lead to key insights into the human microbiome. While 16S rRNA sequencing is economical, informative, and powerful, the technique is ultimately limited to describing community makeup. As the field progresses, increasing value is being put on how communities function, rather than simply how they are composed.

The emergent tool that satisfies this need is shotgun metagenomics, which enables researchers to understand the functional potential of microbial consortiums through an analysis of the complete genomic repertoire of a community. The technique works by taking all the DNA extracted from a sample, fracturing it into manageable lengths, and then sequencing those fragments [11]. Following sequencing, the many disparate reads are then stitched back together (in silico) into the full genomes of the various organisms sampled, in a process called binning. As this process does not rely on the amplification of just a single marker gene, the volume of sequence data produced is much greater. From this outpouring of data, taxonomy can still be determined from signature genes (16S rRNA), but it also becomes possible to comb through genomes in search of regions of interest. For example, by focusing on specific genes, the metabolic and signaling capacity of taxa can be deduced, and from that it can be inferred how they may interact with the host environment [12]. Moreover, as metagenomic data provides in-depth information about organisms that are difficult to culture, it can characterize microbiomes of interest in ways that few other methods can [13].

Classifying the Microbiome

The development of the human microbiome follows variable trajectories depending on the organ, with different microbial communities associating with the skin, lungs, nose, mouth, and the gastrointestinal and urogenital tracts. Moreover, each section of the body develops a unique biogeography, with distinct regionalization present. Skin, for example, often displays dramatic variation in microbial composition and structure across different body sites [14]. It turns out that the physical (i.e. moisture, sebum production) and topographical characteristics of the skin are what play a significant role in determining how similar the skin microbiome is across different areas of the body [15]. Additionally, these same factors that lead to variation across skin sites also account for the individuality of the skin microbiome. Over the course of their development, each person ends up inadvertently cultivating a unique microbial signature on their skin, irrespective of the differences between skin sites [16]. While skin microbial communities can (and do) undergo changes, especially with vigorous cleaning, the original microbiota remarkably reemerges soon after disturbance [17].

Likewise, the oral microbiome appears unique to each individual [18], though prolonged oral interaction between humans can undermine that fact [19]. Longitudinal characterization of the human gut microbiome has similarly demonstrated that the microbiota of the adult gut is relatively unique among individuals. While during the first 3 years of life the human gut microbiome is in constant flux [20, 21], even day-old preterm infants have a unique microbiome at the genotype level [22]. As a result of such vast individuality, when the microbiomes of a large number of people are compared, individuals fall on a continuum of human microbial diversity [23, 24]. This variability of the microbiome within a human population makes blanket stratification difficult for particular disease states, albeit some biomarkers can be used to identify certain conditions.

Interestingly, though individuals have distinct and fairly stable microbial profiles, these communities can be easily perturbed. In the gut, changes in diet can have profound impacts on the microbial community structure. This means that describing the gut microbiota based on the relative abundance of its members provides a limited view of the microbial assemblage [25]. Factors like gut transit time are of considerable importance, as different bacterial taxa, such as those associated with biofilm formation or rapid cellular

reproduction, are selected for in rapid transit time scenarios [26].

The vaginal microbiome has a similar degree of stability to the skin microbiome, and, unlike the gut, substantial success has been achieved in classifying the vaginal microbiome into discrete states, especially during disease. During menstruation and pregnancy, a unique microbiome emerges, one that is highly similar across populations [27, 28]. Asymptomatic women tend to have substantially different vaginal microbiomes, dominated by individual species of *Lactobacillus* and a considerable diversity of other anaerobic taxa [27]. The Lactobacilli are believed to benefit the host by lowering vaginal pH through fermentation end-products, thereby reducing the likelihood of allochthonous microbial colonization or pathogen invasion. Conversely, the disease state of bacterial vaginosis, which presents itself as a disruption of the normal vaginal functional ecosystem, is characterized by a homogenization of the microbial community, providing a generalized biomarker of disease [29].

Most commonly, the gut microbiome of an individual is inferred by collecting a fecal sample, in part because it is noninvasive and easy to collect. However, there is wide variation along the GI tract in terms of bacterial density and environmental conditions. For example, the communities in the small intestine are vastly different from those that inhabit the colon. Furthermore, within the same body region, bacteria are differentiated across microniches—lumen vs. mucosa vs. crypts [30]. It is important to consider the limitations of sampling each site in pursuing a comprehensive understanding of a disease state.

Factors Affecting the Healthy Microbiome

Genetics

As noted above, many factors can impact the "health" of the microbiome. The role of host genetics has been, however, of particular interest. While early studies suggested that monozygotic twins are no more similar in terms of their overall gut microbiota than dizygotic twins [31–33] research based on larger cohort sizes showed a small but statistically significant effect, with *Christensenella* being especially heritable, for instance [34]. Consequently, while host genetics may potentially play some small role, the large differences observed among different human populations are likely due to other factors [32].

Diet

Diet has been studied at length in relation to the gut microbiome [35] but unfortunately less so with respect to other microbiomes across the body. Evidence to date suggests that dietary choices can have profound effects in the long-term [36] and in the short-term as well given an extreme enough dietary shift [25]. Interestingly, dietary changes often do not affect different individuals in the same way. For example, the effect of the same dietary ingredient on blood glucose measurements can vary across people, an effect mediated by the microbiome [37]. An important question in the field is whether the microbiome can influence dietary preferences, which could lead to feedback loops when these dietary changes in turn alter the microbiome.

Antibiotics

The effect of antibiotics on the microbiome is, as expected, immense compared to other factors and has been studied extensively [38]. Not only is the impact of antibiotics significant, it can also be compounding, as the microbiome in adults appears to have little resilience to repeated episodes of antibiotic administration [39]. Intriguingly, it appears that the same antibiotic can have different effects on the same microbe depending on the constituency of the rest of the microbiome [40], perhaps due to different growth phases or metabolic states. An especially interesting development in this area of research is the increasing evidence that antibiotics in early life can have a profound effect on the microbiome, one that can result in an increased likelihood of adult-onset obesity [41].

Lifestyle

A collection of experiences and exposures throughout life add to the diversity of human microbiomes. Pets, in particular dogs, have a statistically significant effect on the microbiome, especially that of the skin. The microbiomes of each individual in a couple more closely resemble one another if the couple has a dog, yet the same trend does not hold if a small child is present instead. Thus, couples with a child and no dog are likely no more similar than those without a child [42]. Furthermore, pet ownership and exposure to livestock, especially cows, has been associated with a decreased risk of asthma and allergic disease [43].

Dynamics of the Microbiome Within a Host

Human interaction with the environment, including with other people, creates the potential for specific microbial taxa to invade and colonize the body, sometimes leading to observable changes on and within the host. For instance, certain individual bacterial species have been found to induce

obesity by altering hormonal regulation [44]. The bacterial phylum Firmicutes comprises a signifficantly greater proportion of the microbiota in obese individuals, and while this phylum is incredibly diverse and metabolically complex, colonization by certain bacterial species in this phylum may have a significant impact on metabolic disease and host physiology [45].

The microbiome itself maintains a circadian rhythm, which is linked to host circadian cycles. Disruption of these microbial diurnal cycles can lead to disruption in host circadian rhythms, which can specifically alter hormone regulation [46]. Thus, it becomes apparent that the human microbiome can demonstrate enormous plasticity, yet at the same time it remains robust on longer time scales (Fig. 39.1). This dichotomy may at first glance seem difficult to resolve but only until the ecological dynamics of the system are considered. All ecosystems undergo variation in population density and diversity, albeit with different magnitudes depending on the temporal scale. This variation can be spurred by, among other factors, competition among microbial taxa and shifting metabolic relationships. Moreover, these shifts can be compounded and influenced by a developing immune system, changing dietary patterns, or a constant bombardment of bacteria from peers and the environment. Longitudinal characterization of the host microbiome and its sources is therefore essential to capture dynamic variance within an individual and to determine the degree to which the system demonstrates predictable successional traits [47].

This plasticity-stability dichotomy is evident even over a period of days, as is illustrated in the first fine-scale time series analysis of the human microbiome [20]. In this study, two subjects provided daily samples of their oral, skin, and fecal microbiota—one for 6 months and the other for 18 months. The results illustrate that a very small fraction of bacterial species (approximated as operational taxonomic units [OTUs] defined by 97% sequence similarity in the sequenced portion of the 16S rRNA region) are found to be consistently present across all samples in an individual host. For skin sites (the left and right palm), there were no species found in all samples, while in the gut and the mouth, about 5% of the species were defined as belonging to a stable core microbiome. Despite all this chaos, though, each person still maintains a personalized microbiome.

The degree of personalization of the human microbiome vastly exceeds the human genome, which is over 99.5% identical between individuals, suggesting that only 0.5% of the genome is unique to an individual. Conversely, two individuals can show zero overlap in microbial species in their microbiome. This degree of personalization is so great that it has become an area of great intrigue for those interested in its forensic applications [48].

While we are now accustomed to thinking about the composition of the human microbiome as being unique to individuals, more recent work shows that the rate of change of the microbiome is even further individualized [49]. In one study, 85 college-age adults donated weekly microbiome

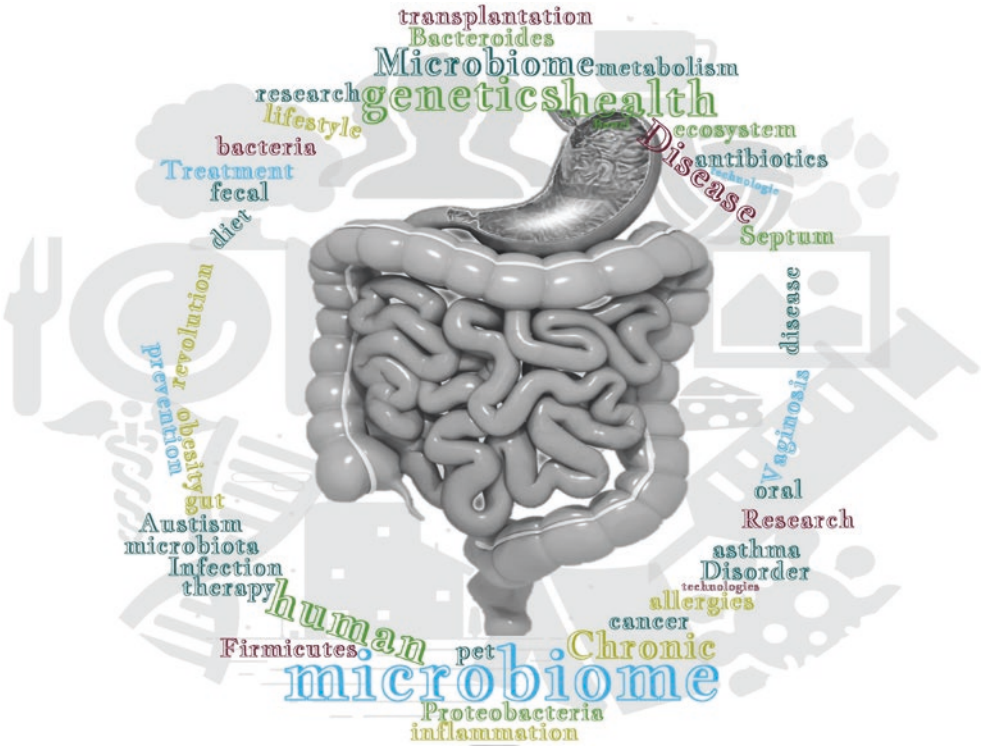


Fig. 39.1 A better understanding of how we physiologically, metabolically, and immunologically interact with, and are shaped by, our environment will revolutionize our ability to target interventions in human health and wellness

samples over a 3-month period from gut, skin, and oral sites. Over this timeframe it emerged that the microbiome of some individuals remains fairly constant, whereas others see rapid change. These differing rates of temporal variability were identified at all of the body sites profiled (the palm of the dominant hand, the forehead, the tongue, and feces), and the rate of change was not observed to be correlated across the different sites. On average, skin sites change the quickest, followed then by the gut and mouth, respectively. This pattern matches the relative sizes of the stable core microbiomes observed for each body area in the aforementioned long-term survey [20]. Of the host metadata collected, none was correlated with differing rates of microbiome variation, so it is not possible to speculate on the underlying causes of these differential rates. However, one interesting observation was that individuals who reported taking antibiotics during, or in the week preceding, the sampling period did not experience a greater rate of change in their microbiome than individuals not on antibiotics. This suggests that some individuals' microbiome rate of change may be higher than others and could reflect cumulative lifetime or more recent exposures.

Most clinically oriented studies focus on the association between the microbiome and host disease states or the likelihood of response to a treatment. However, one recent study has put more focus on the rate of change of the microbiome and suggests it may have understated clinical merit [50]. The rate of change of the vaginal microbiome differs across women with bacterial vaginosis and is predictive of the subtype of bacterial vaginosis with which the women are infected. That observation, paired with data indicating that the rate of change of the gut, skin, and oral microbiomes is unique to each individual, underscores the idea that characterizing temporal variation may be an important part of understanding and diagnosing issues with an individual's microbiome.

A better understanding of traits, such as dynamic microbiome variance in individuals, and how they relate to successional patterns will make it easier to define causative relationships and to interpret convergent cross-mapped correlations between taxonomic groups [51]. Moreover, examining what sources shape the microbiome is key to understanding the mercurial nature of it. The application of techniques like dynamic Bayesian mapping, which is used to detect significant conditional co-dependent relationships between external sources of bacteria and the human microbiome of specific body sites, can be utilized to map our interaction with the external world [16]. Bayesian statistics can further be drawn upon to map the relative contribution of a specific source to the human microbiome over time [52] or to create artificial neural networks of conditional dependencies that can then be used to capture predictive characteristics of a microbial network [53]. Using these methods, the dynamic nature of the human microbiome, both within an individual and within a population, can be captured. Once fully understood, we can harness it to provide a predictive signature or characteristic biomarker for a given disease state, or physiological, immunological, or neurological condition. The application of machine learning algorithms is a promising technology on the horizon and is already being used in microbial forensics. The technique has already been successful at identifying highly predictive characteristics of a microbial signature for the purposes of mapping the forensic relationships between people and their built environments [16].

The Human Microbiome in Early Development

As thus established, the human microbiome is unique to each individual, and the differences between individuals are large compared to the typical differences within an individual over time [16, 54]. Identical twins are barely more similar to one another compared to nonidentical twins, in terms of their microbiomes [34], suggesting that the effect of the human genome is limited and that most microbial community assembly may be stochastic or determined by environmental factors (Fig. 39.2). Yet colonizing initially germ-free mice with diverse environmental samples demonstrates that very few environmental bacteria can survive in the mouse gut and those that do are rapidly displaced by human-derived or mouse-derived bacteria upon exposure [55]. The human immune system has a complex dynamic relationship with microbiota and seems to initially recruit favorable microbial taxa [56]. During and shortly after birth, newborns are exposed to maternal and environmental microbes that initiate gut microbiome establishment [57]. Short- and long-term microbial blooms develop as food and other microbial exposures are introduced to the oral and gut cavities. Within the first year of life, 10^{14} microbes/mL, comprised of 500–1000 species, generally colonize the gastrointestinal tract [58]. After weaning, the gut microbiome often becomes firmly established, leading to a lifelong signature in healthy individuals [59]. The influence of human milk [60] and other factors like delivery mode, prenatal exposures, environment, immune activation and inflammation, disease burden, and maternal contact have dramatic influences on the composition and structure of the microbiome. Ongoing interactions between intestinal epithelial and lymphoid cells and with the developing commensal microbiome are required for normal infant gut and immune development. Apoptotic stimuli, reactive oxygen synthesis, and Toll-like receptor signaling are induced by commensal bacteria to produce a state of controlled inflammation that helps develop innate immune defense and promotes pathogen recognition by both innate and adaptive immunity [61].

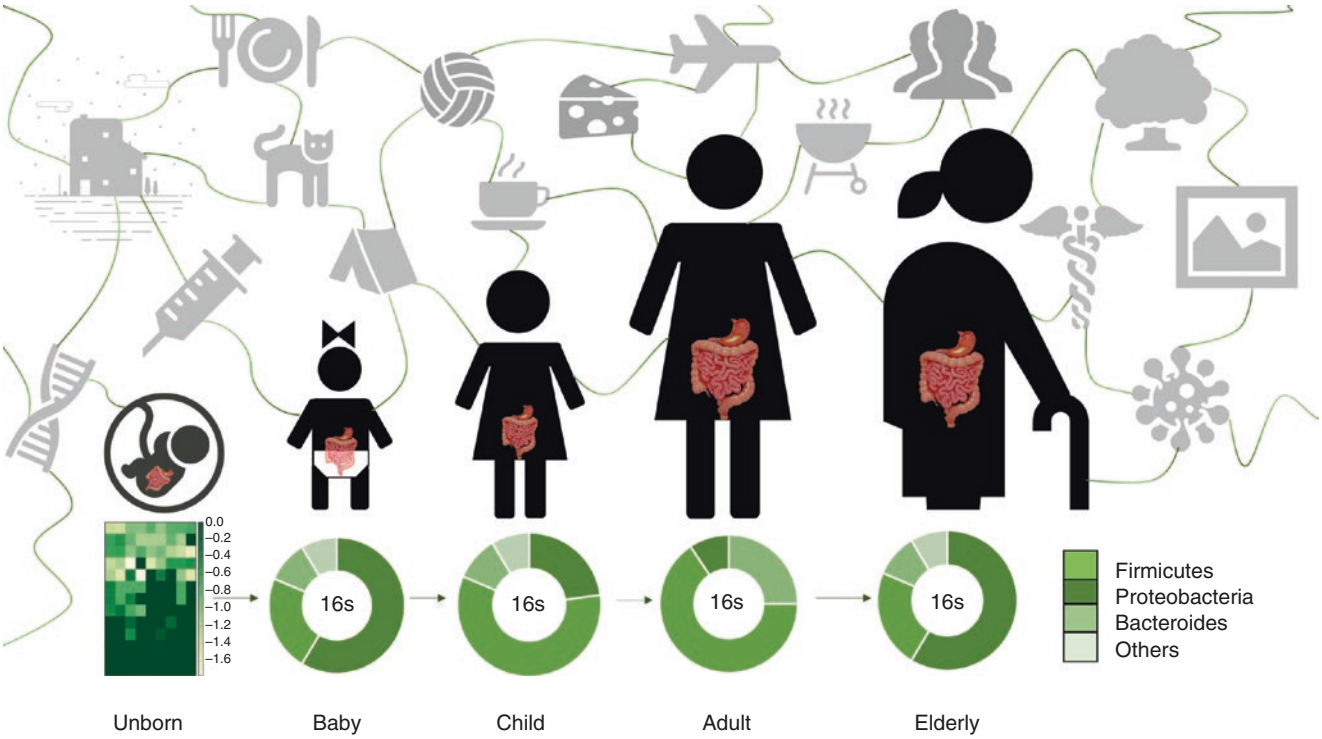


Fig. 39.2 The possible microbiome of an individual over the course of their life. From the womb onward, the human microbiome is influenced by the host's interaction with the environment, by dietary and lifestyle choices, and by physiological changes associated with aging. The intra-individual variation in the microbiome can be substantial between major life stages, however, a lack of multidecadal timeseries studies limits our ability to determine the true stability and variance in the microbiome

There is extensive evidence that the microbiome can be used to explain a substantially greater percentage of variance in a population for a given condition than human genetic factors. Perhaps the most dramatic example is the case of a recurrent *Clostridium difficile* infection (CDI), in which the aberrant stool microbiome looks nothing like that of a healthy stool but rather like a completely different body site; the restoration to the healthy state after fecal microbiota transplant is both rapid and visible at the whole-community level [62]. This effect size is much greater than any human genetic variation in the stool microbiome that has yet been observed, and it perhaps explains the high efficacy of a stool transplant compared to standard antibiotic treatments [63].

Moreover, obesity provides an example in which human genetics has failed to explain the disease epidemic. In contrast, the gut microbiome can classify individuals as lean or obese with over 90% accuracy [64]. More specifically, the abundance of particular taxonomic groups, like *Christensenella*, is found to be negatively correlated with BMI and can induce weight loss when experimentally fed to mice [34].

Autism spectrum disorder has a complex presentation of symptoms and is difficult to pin down to genetics, mainly due to the number of confounding influences and variables [65]. Yet environmental exposure, and likely the microbiome, plays a substantial role in shaping the etiology of the disease [66, 67], and animal models have been used to demonstrate the mechanistic activity of bacterial metabolites in mediating autism-like behaviors [68].

In addition to the metabolic and cognitive diseases mentioned above, the microbiome has also been implicated for its ability to impact a number of autoimmune conditions. Many of these diseases, like inflammatory bowel disease, Type 1 diabetes, and childhood-onset asthma, are further notable for becoming more and more frequent on a global scale. A growing body of evidence is linking these diseases with altered microbiota compositions, especially loss of diversity, as seen in inflammatory bowel disease patients [69] and children at risk for type 1 diabetes [70]. Altogether, the relatively simultaneous increase of these varied illnesses point toward a single factor. One hypothesis is that these illnesses are caused by an overall shift in the microbiome, rather than the acquisition or loss of specific groups. Such a microbial disturbance may be particularly impactful during early life, as it is then that immunity, metabolism, and cognition are under development. A large Canadian study of infant stool samples ($n = 319$) collected over the first 90 days of life compared the early-life (first 90 days) fecal microbiota of infants who went on to develop allergic disease and wheezing at age 2 vs. subjects who did not. It emerged that a

depletion of certain bacterial species is characteristic of atopic children and corresponds with reduced concentrations of fecal acetate and dysregulation of enterohepatic metabolites [71]. This suggests that the foundation for allergic disease development occurs in early life and is mediated, at least in part, by gut microbiome dysbiosis.

Microbial Impacts on Pathogenesis

Much of the microbiome data generated to date has focused on describing microbial community composition and taxonomic changes in order to characterize disease states and correlate changes with disease features [72–74]. A key question that remains in the field is whether these changes in human microbial composition are drivers of—or driven by—disease processes. Several independent lines of evidence have recently emerged indicating that, in many cases, microbial species and their activities are responsible for immunological and disease phenotypes. Some of the earliest research demonstrating that disease phenotypes are transmissible via the microbiome include mouse studies on body mass. In these studies, the transfer of gut microbial contents from diet-induced obese mice to previously germ-free animals is seen to cause rapid weight gain—an effect not observed in gut microbiome transplant from lean animals [75].

Thus, while various studies have begun to prove that the gut microbiome can be a driver of specific disease features, another line of inquiry has been yielding further insight into the mechanistic links between the gut microbiome and disease development. Carnitine, which is found in high concentrations in red meat, is metabolized by the gut microbiome to trimethylamine (TMA), which subsequently undergoes oxidation in the liver to trimethylamine N-oxide (TMAO). In humans, elevated serum concentrations of TMAO are associated with atherosclerosis. In one study, it was shown that while omnivorous human subjects possess a microbiome with the capacity to convert carnitine to TMA, individuals who have previously consumed a vegan or vegetarian diet have significantly reduced carnitine biotransformation capacity and lower serum levels of TMAO following consumption of red meat. The investigators also demonstrated, using a murine model, that TMAO production from L-carnitine is an inducible phenotype and, through antibiotic ablation, is critically dependent on the gut microbiome [76]. More recent studies indicate that transfer of the gut microbiome from pro-atherosclerotic mice to an atherosclerosis-resistant strain leads to increased concentrations of circulating TMAO and a choline diet-dependent increase in atherosclerotic plaque burden [77].

Other murine studies show that by inoculating mice with specific microbes, immune phenotypes associated with the development of, or protection against, disease can be induced. For example, supplementation of mice with a cocktail of 46 *Clostridium* species (belonging to clades IV and XIV) induces robust proliferation of a T-helper cell subset (IL-10 producing FoxP3$^+$ T-regulatory cells) critical to maintenance of immune homeostasis [78]. Ivanov and colleagues also demonstrate that a single species, a segmented filamentous bacterium, can induce proliferation of ileal Th17 cells—a subset of CD4$^+$ cells [79] that have been associated with a number of chronic inflammatory and autoimmune diseases [80]. In other studies utilizing murine models of allergic airway disease, manipulation of the gut microbiome via supplementation with specific bacterial species (*Lactobacillus johnsonii* or a mix of *Clostridium* species) is seen to promote downregulation of pro-inflammatory Th2 responses and to confer protection against allergic airway inflammation [78, 81].

While much has been sorted out, additional studies of the microbiome are needed to better understand its role as a driver of human disease. Furthermore, as the mucosal microbiome becomes a major target for therapeutic study, deeper investigations into it are warranted as well. Prospective cohorts, ideally birth cohorts, in which samples can be collected longitudinally and associated with expansive metadata, such as clinical events and sequelae, are critical to fully understanding the role of the microbiome in disease genesis and progression. This approach allows individuals to act as their own controls—a critical factor given the broad variation in microbiome composition across individual, temporal, geographic, and ethnic divides [32]. Such studies, coupled with in vitro investigations of the ecology of human-associated microbes, will ultimately allow for an improved understanding of microbial communities and their critical relationship with human health.

Diagnostic Potential of Microbiome Sequencing

Because microbial communities are sensitive to alterations in their environments and can turnover rapidly, they are plastic in their response to perturbations in the host. Therefore, they can be exploited as a bellwether for the state of their environment. Indeed, because the metagenome differs so much more between individuals than the human genome does, the probability of identifying important variability is much greater. As the microbiome becomes a hot topic linked to a variety of disease conditions, researchers are optimistic that it can be used as a clinical diagnostic tool and a target for personalized precision medicine. However, links between microbes and host health are multifactorial and not necessarily unidirectional. Therefore, with rare exceptions, it may be premature to implicate an altered microbiome in disease pathogenesis.

Initially, the bulk of microbiome research was focused on the gut and therefore on gastrointestinal disorders. For decades, scientists have known about the effects of *Helicobacter pylori* and antibiotic-associated diarrhea, both clear examples of altered gut microbiota. According to Quigley [82], to be able to fully utilize the microbiome for clinical applications, however, three core conditions must be met: (1) we must know what is "normal," (2) we must be able to reproducibly define what is "abnormal," and (3) we must identify a clinically meaningful relationship between a given microbiome profile and disease state.

As mentioned earlier in this chapter, there remains a lot to be learned about what constitutes a healthy and normal microbiome. Scientists are far from identifying a universal ideal microbiome, especially because the microbiome can undergo dramatic shifts and is highly driven by age, diet, and other temporal factors. Therefore, though many studies have found strong associations between given disease states and particular microbial profiles and the field is increasingly probing mechanisms, these conditions remain to be met thus far.

This is not to say that progress has not been made. There are a few examples in which abnormal microbial communities have a well-characterized relationship with disease states, and therapeutics have been developed in these areas. Perhaps best known is the case of *Clostridium difficile* infection, which usually develops when patients have been exposed to broad-spectrum antibiotics that suppress their native microbiome. Studies have elucidated a microbial signature associated with predisposition to recurrent *C. difficile* infections [83]. These signatures are identified using 16S marker gene analysis and look at diversity metrics. However, changes in the functional potential of the microbial community may be more important than simply knowing which microbes are associated with the disease, as diversity metrics alone can struggle to explain real-world phenomena. For example, a recent study on the role of the microbiome in hepatic encephalopathy shows that the antibiotic rifaximin did not change the microbial composition but did lead to shifts in bacterial metabolism after treatment [84].

The microbiome is an attractive avenue for diagnostic tools, in part because it is less invasive than many current standards of care, while the ever-decreasing cost and increasing speed of high-throughput sequencing technologies allow for rapid and inexpensive testing. A number of ventures have begun to commercialize microbiome characterization tools. So far, although 16S rRNA tests may be performed in CLIA-certified and College of American Pathologists-accredited labs, they are marketed as screening tools. The results are complex and difficult to interpret compared with a traditional culture-based lab test that may detect the presence of a single pathogen. All of the factors that affect microbiome variability affect interpretation of such screening tests. Despite not providing definitive diagnoses, these tests do have value in a healthcare setting but currently as just one part of a patient's clinical picture.

Microbiome-Based Therapeutics

Compared with the human genome, the microbiome may be much more readily manipulated. As such, it makes for an attractive therapeutic target. A number of therapies have been developed in animal models as well as clinical trials with the goal of specifically altering the relative abundances of beneficial microbial taxa.

Probiotics

Probiotics are live microorganisms intended to benefit host health. They are increasingly commercially available, manufactured as food products, supplements, and cosmetics. Since they are often regulated as supplements rather than drugs, there is a dearth of clinical trials testing probiotic efficacy. Importantly, probiotics can contain a single strain or a cocktail of different types of microbes. The most common genera of bacteria found in commercially available probiotics include *Lactobacillus spp.* and *Bifidobacterium spp.*, and yeasts like *Saccharomyces boulardii* are also commonly used. However, strains are often selected because they are easy to grow and encapsulate in an orally available form, and not necessarily because they can survive the acidity of the stomach nor colonize the gut adequately.

Various mechanisms have been proposed for the mode of action for probiotics: (1) probiotics modulate the host immune system, (2) probiotics directly affect other commensal or pathogenic microorganisms, or (3) probiotics affect microbial products like toxins or host products like bile acids [85]. All of these effects could potentially occur with the probiotic taking up residence in the gut or while transiently passing through the GI tract. Notably, the dose of bacteria contained in a probiotic supplement is a tiny fraction of the biomass of resident microbes in the gut microbiome, so even if probiotic strains manage to survive and grow, there are likely too few of them to dramatically alter the microbiome composition. Several recent studies have examined whether the microbes delivered in probiotics actually establish themselves in the gut.

A review of randomized, placebo-controlled trials evaluated whether orally ingested probiotics change the bacterial composition of patients' feces, but only one of seven studies found any significant changes in composition and did not conclude that the changes were beneficial to health [86]. Another group found that orally administered probiotics are usually transient in the human gut. In that study, one strain—

Bifidobacterium longum AH1206—stably colonized 30% of participants, yet the individuals in whom the probiotic strain persisted had lower starting levels of any *Bifidobacterium* species in their guts, implying that the probiotic strain was only able to fill a preexisting empty ecological niche [87]. Overall, the literature currently demonstrates a lack of evidence that probiotics can affect fecal microbiota composition in healthy adults. However, the idea that probiotics can be effective to recolonize empty ecological niches indicates a potential for developing personalized probiotics. With a better understanding of which interindividual factors determine the success of probiotic engraftment, it may be possible to tailor probiotics to particular patients [88].

Indeed, while there is scant evidence for probiotic efficacy in healthy adults, probiotics have been shown to be useful in certain disease cases. For example, orally ingested probiotics can alleviate some antibiotic-related or rotavirus-induced diarrhea symptoms. In these disease states, the probiotics may engraft by taking advantage of ecological niches left empty by antibiotics or the increased rate of sloughing of the microbiota.

Prebiotics and Synbiotics

A prebiotic is defined as "a selectively fermented ingredient that allows specific changes, both in the composition and/or activity in the gastrointestinal microflora that confers benefits upon host well-being and health" [89]. Most commonly, prebiotics consist of some type of indigestible fiber that can be fermented by gut microflora and, in doing so, encourages a health-positive shift in bacterial populations [90]. The definition of prebiotic has recently been expanded to include molecular substrates that lead to the selective expansion of certain beneficial bacteria. For example, oligosaccharides contained in human breast milk are selectively metabolized by *Bifidobacterium* in the infant gut, such that breastfed infants have a higher abundance of these bacteria compared to formula-fed infants [91]. Supplementing infant formula with prebiotic oligosaccharides restores some of the *Bifidobacterium* abundant in formula-fed infants [92].

Synbiotics refers to a combination of probiotics (living organisms) and prebiotics (dietary supplements designed to feed the microbiota rather than the host). This allows for the seeding of the host gut with a substrate to encourage probiotic growth and potentially increase engraftment [93].

Fecal Microbial Transplant (FMT)

Recent studies have substantiated the role of the gut microbiome in disease by demonstrating that glucose intolerance, induced via consumption of non-caloric artificial sweetener (NAS), can be transferred to germ-free mice via a fecal transplantation of gut microbiota from NAS-consuming mice [94]. While these studies demonstrate the capacity to transfer a disease phenotype via a microbial transfer, a considerable amount of research also focuses on how microbial communities might protect against disease development. An emerging treatment for both adult and pediatric patients infected with an antimicrobial-resistant *Clostridium difficile* infection is the fecal microbial transplant, which consists of an infusion of slurry generated from the feces of a healthy donor. The efficacy of this technique is in the numbers, as it has resulted in remission rates of up to 94% [95, 96]. Gut microbiome restitution, especially in regard to recurrent *C. difficile* infections, can therefore significantly improve the health of an individual.

What's Next?

Studies about the role of the human microbiome in health and disease have now arrived at somewhat of a fork in the road. With a push toward expansive clinical trials, large-scale observational studies, double-blind controlled experimental manipulation, and improved model systems for performing intervention studies, we are seeing a shift from descriptive studies able to pinpoint correlations that may infer relevant interactions toward investigations that address the need to identify causation. Yet, there is still substantial room for longitudinal studies that can yield understanding into the dynamic relationships between bacterial communities and host factors. Prospective longitudinal cohort studies offer the best opportunity to capture the inherent complexity between the myriad of factors that can influence health or disease. We are still on a voyage of discovery, and while controlled longitudinal investigation can help identify particular associations that may be responsible for observed outcomes, it can be hard to use these studies to pinpoint causation. Nonetheless, they are still essential, as they enable us to identify questions which require deeper investigation.

The advent and availability of clinical data repositories has had a significant impact on discovery. If a particular trend is observed in a longitudinal or cross-sectional study, then validating that the two factors show a similar relationship in broad-scale clinical data can be incredibly valuable. It has become essential to infer whether the observed relationships are artifacts of experimental design or representative of some broader trend. These decisions rely on excellent data curation, foresight for data collection, integration across clinical departments, and comprehensive data sharing that does not sacrifice patient data protection.

Just as experimental design has been undergoing change and rebirth, so too have there been shifts in the type of microbiome data collected. We are seeing an active marriage of

sequencing-based approaches, such as amplicon analysis, metagenomics, and metatranscriptomics, with proteomics and metabolomics technologies [97]. These non-sequence-based datasets require careful curation, integration, and quality control, which is best handled through the sharing of data and cooperative analysis. As more and more data is collected and analyzed, we must start to think about how this data might be integrated into models of the human microbiome. In general, by assembling metagenomic data into genomes, mapping transcripts onto genes, and predicting and validating the proteins and metabolites produced, fairly accurate and in-depth models of the microbiome can be produced. These flux-balance, agent-based community interaction models are also nested within human immune, neurological, and hormonal models that allow for cross-reactivity between the microbiome and human host [12]. These models will change our ability to test new drugs and develop new therapies and help us to improve studies by predicting their outcome and statistical power before they even begin.

The future of this field is bright, and the investigative potential is accelerating rapidly. More and more innovative therapies enter into development each day—a fact we believe will be truly transformational for medicine.

References

1. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell. 2016;164(3):337–40.
2. Turnbaugh P, Ley R, Hamady M, Fraser-Liggett C, Knight R, Gordon J. The human microbiome project. Nature. 2007;449(7164):804–10.
3. Locey K, Lennon J. Scaling laws predict global microbial diversity. Proc Natl Acad Sci. 2016;113(21):5970–5.
4. Luckey T. Introduction to intestinal microecology. Am J Clin Nutr. 1972;25(12):1292–4.
5. Rosner J. Ten times more microbial cells than body cells in humans? Microbe Mag. 2014;9(2):47.
6. Qin J, Li R, Raes J, Arumugam M, Burgdorf K, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65.
7. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger J, Chinwalla A, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14.
8. Clarridge J. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17(4):840–62.
9. Woo P, Lau S, Teng J, Tse H, Yuen K. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect. 2008;14(10):908–34.
10. Galley J, Bailey M, Kamp Dush C, Schoppe-Sullivan S, Christian L. Maternal obesity is associated with alterations in the gut microbiome in toddlers. PLoS ONE. 2014;9(11):e113026.
11. Sharpton T. An introduction to the analysis of shotgun metagenomic data. Front Plant Sci. 2014;5:209.
12. Cardona C, Weisenhorn P, Henry C, Gilbert J. Network-based metabolic analysis and microbial community modeling. Curr Opin Microbiol. 2016;31:124–31.
13. Sangwan N, Zarraonaindia I, Hampton-Marcell J, Ssegane H, Eshoo T, Rijal G, et al. Differential functional constraints cause strain-level endemism in polynucleobacter populations. mSystems. 2016;1(3):e00003–16.
14. Grice E, Segre J. The skin microbiome. Nat Rev Microbiol. 2011;9(4):244–53.
15. Grice E, Kong H, Conlan S, Deming C, Davis J, Young A, et al. Topographical and temporal diversity of the human skin microbiome. Science. 2009;324(5931):1190–2.
16. Lax S, Smith D, Hampton-Marcell J, Owens S, Handley K, Scott N, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345(6200):1048–52.
17. Hannigan G, Meisel J, Tyldsley A, Zheng Q, Hodkinson B, SanMiguel A, et al. The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. MBio. 2015;6(5):e01578–15.
18. Lazarevic V, Whiteson K, Hernandez D, Francois P, Schrenzel J. Study of inter- and intra-individual variations in the salivary microbiota. BMC Genomics. 2010;11(1):523.
19. Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. Microbiome. 2014;2(1):41.
20. Caporaso J, Lauber C, Costello E, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50.
21. David L, Materna A, Friedman J, Campos-Baptista M, Blackburn M, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. Genome Biol. 2014;15(7):R89.
22. Raveh-Sadka T, Thomas B, Singh A, Firek B, Brooks B, Castelle C, et al. Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. elife. 2015;4:e05477.
23. Knights D, Ward T, McKinlay C, Miller H, Gonzalez A, McDonald D, et al. Rethinking "enterotypes". Cell Host Microbe. 2014;16(4):433–7.
24. Jeffery I, Claesson M, O'Toole P, Shanahan F. Categorization of the gut microbiota: enterotypes or gradients? Nat Rev Microbiol. 2012;10(9):591–2.
25. David L, Maurice C, Carmody R, Gootenberg D, Button J, Wolfe B, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2013;505(7484):559–63.
26. Vandeputte D, Falony G, Vieira-Silva S, Tito R, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut. 2015;65(1):57–62.
27. Ma B, Forney L, Ravel J. Vaginal microbiome: rethinking health and disease. Annu Rev Microbiol. 2012;66(1):371–89.
28. Romero R, Hassan S, Gajer P, Tarca A, Fadrosh D, Nikita L, et al. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome. 2014;2(1):4.
29. Xiao B, Niu X, Han N, Wang B, Du P, Na R, et al. Predictive value of the composition of the vaginal microbiota in bacterial vaginosis, a dynamic study to identify recurrence-related flora. Sci Rep. 2016;6(1):26674.
30. Donaldson G, Lee S, Mazmanian S. Gut biogeography of the bacterial microbiota. Nat Rev Microbiol. 2015;14(1):20–32.
31. Turnbaugh P, Hamady M, Yatsunenko T, Cantarel B, Duncan A, Ley R, et al. A core gut microbiome in obese and lean twins. Nature. 2008;457(7228):480–4.
32. Yatsunenko T, Rey F, Manary M, Trehan I, Dominguez-Bello M, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222–7.
33. Stahringer S, Clemente J, Corley R, Hewitt J, Knights D, Walters W, et al. Nurture trumps nature in a longitudinal survey of salivary

bacterial communities in twins from early adolescence to early adulthood. Genome Res. 2012;22(11):2146–52.
34. Goodrich J, Waters J, Poole A, Sutter J, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789–99.
35. Albenberg L, Wu G. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology. 2014;146(6):1564–72.
36. Wu G, Chen J, Hoffmann C, Bittinger K, Chen Y, Keilbaugh S, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105–8.
37. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell. 2015;163(5):1079–94.
38. Modi S, Collins J, Relman D. Antibiotics and the gut microbiota. J Clin Investig. 2014;124(10):4212–8.
39. Dethlefsen L, Relman D. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci. 2010;108(Supplement_1):4554–61.
40. Maurice C, Haiser H, Turnbaugh P. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013;152(1–2):39–50.
41. Trasande L, Blustein J, Liu M, Corwin E, Cox L, Blaser M. Infant antibiotic exposures and early-life body mass. Int J Obes. 2012;37(1):16–23.
42. Song S, Lauber C, Costello E, Lozupone C, Humphrey G, Berg-Lyons D, et al. Cohabiting family members share microbiota with one another and with their dogs. elife. 2013;2:e00458.
43. von Mutius E. The microbial environment and its influence on asthma prevention in early life. J Allergy Clin Immunol. 2016;137(3):680–9.
44. Fei N, Zhao L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J. 2012;7(4):880–4.
45. Koliada A, Syzenko G, Moseiko V, Budovska L, Puchkov K, Perederiy V, et al. Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population. BMC Microbiol. 2017;17(1):120.
46. Thaiss C, Zeevi D, Levy M, Zilberman-Schapira G, Suez J, Tengeler A, et al. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell. 2014;159(3):514–29.
47. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman J, et al. Unlocking the potential of metagenomics through replicated experimental design. Nat Biotechnol. 2012;30(6):513–20.
48. Fierer N, Lauber C, Zhou N, McDonald D, Costello E, Knight R. Forensic identification using skin bacterial communities. Proc Natl Acad Sci. 2010;107(14):6477–81.
49. Flores G, Caporaso J, Henley J, Rideout J, Domogala D, Chase J, et al. Temporal variability is a personalized feature of the human microbiome. Genome Biol. 2014;15(12):531.
50. Gajer P, Brotman R, Bai G, Sakamoto J, Schutte U, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012;4(132):132ra52.
51. Sugihara G, May R, Ye H, Hsieh C, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. Science. 2012;338(6106):496–500.
52. Knights D, Kuczynski J, Charlson E, Zaneveld J, Mozer M, Collman R, et al. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8(9):761–3.
53. Larsen P, Field D, Gilbert J. Predicting bacterial community assemblages using an artificial neural network approach. Nat Methods. 2012;9(6):621–5.
54. iHMP. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe. 2014;16(3):276–89.
55. Seedorf H, Griffin N, Ridaura V, Reyes A, Cheng J, Rey F, et al. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell. 2014;159(2):253–66.
56. Kawamoto S, Maruya M, Kato L, Suda W, Atarashi K, Doi Y, et al. Foxp3+ T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis. Immunity. 2014;41(1):152–65.
57. Koenig J, Spor A, Scalfone N, Fricker A, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci. 2010;108(Supplement_1):4578–85.
58. Weng M, Walker W. The role of gut microbiota in programming the immune phenotype. J Dev Orig Health Dis. 2013;4(03):203–14.
59. Maynard C, Elson C, Hatton R, Weaver C. Reciprocal interactions of the intestinal microbiota and immune system. Nature. 2012;489(7415):231–41.
60. Martín R, Langa S, Reviriego C, Jimínez E, Marín M, Xaus J, et al. Human milk is a source of lactic acid bacteria for the infant gut. J Pediatr. 2003;143(6):754–8.
61. Jakaitis B, Denning P. Commensal and probiotic bacteria may prevent NEC by maturing intestinal host defenses. Pathophysiology. 2014;21(1):47–54.
62. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, et al. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. Microbiome. 2015;3(1):10.
63. Kassam Z, Lee C, Yuan Y, Hunt R. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. Am J Gastroenterol. 2013;108(4):500–8.
64. Knights D, Parfrey L, Zaneveld J, Lozupone C, Knight R. Human-associated microbial signatures: examining their predictive value. Cell Host Microbe. 2011;10(4):292–6.
65. Sahin M, Sur M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. Science. 2015;350(6263):aab3897.
66. McDonald D, Hornig M, Lozupone C, Debelius J, Gilbert J, Knight R. Towards large-cohort comparative studies to define the factors influencing the gut microbial community structure of ASD patients. Microb Ecol Health Dis. 2015;26(1):26555.
67. Kang D, Park J, Ilhan Z, Wallstrom G, LaBaer J, Adams J, et al. Reduced incidence of prevotella and other fermenters in intestinal microflora of autistic children. PLoS One. 2013;8(7):e68322.
68. Hsiao E, McBride S, Hsien S, Sharon G, Hyde E, McCue T, et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell. 2013;155(7):1451–63.
69. Knights D, Silverberg M, Weersma R, Gevers D, Dijkstra G, Huang H, et al. Complex host genetics influence the microbiome in inflammatory bowel disease. Genome Med. 2014;6(12):107.
70. Uusitalo U, Liu X, Yang J, Aronsson C, Hummel S, Butterworth M, et al. Association of early exposure of probiotics and islet autoimmunity in the TEDDY study. JAMA Pediatr. 2016;170(1):20.
71. Arrieta M, Stiemsma L, Dimitriu P, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci Transl Med. 2015;7(307):307ra152.
72. Ley R, Backhed F, Turnbaugh P, Lozupone C, Knight R, Gordon J. Obesity alters gut microbial ecology. Proc Natl Acad Sci. 2005;102(31):11070–5.
73. Abreu N, Nagalingam N, Song Y, Roediger F, Pletcher S, Goldberg A, et al. Sinus Microbiome diversity depletion and corynebacterium tuberculostearicum enrichment mediates rhinosinusitis. Sci Transl Med. 2012;4(151):151ra124.
74. Morgan X, Tickle T, Sokol H, Gevers D, Devaney K, Ward D, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.

75. Turnbaugh P, Bäckhed F, Fulton L, Gordon J. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe. 2008;3(4):213–23.
76. Koeth R, Wang Z, Levison B, Buffa J, Org E, Sheehy B, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med. 2013;19(5):576–85.
77. Gregory J, Buffa J, Org E, Wang Z, Levison B, Zhu W, et al. Transmission of atherosclerosis susceptibility with gut microbial transplantation. J Biol Chem. 2014;290(9):5647–60.
78. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of colonic regulatory T cells by indigenous Clostridium species. Science. 2010;331(6015):337–41.
79. Ivanov I, Atarashi K, Manel N, Brodie E, Shima T, Karaoz U, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell. 2009;139(3):485–98.
80. Noack M, Miossec P. Th17 and regulatory T cell balance in autoimmune and inflammatory diseases. Autoimmun Rev. 2014;13(6):668–77.
81. Fujimura K, Demoor T, Rauch M, Faruqi A, Jang S, Johnson C, et al. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. Proc Natl Acad Sci. 2013;111(2):805–10.
82. Quigley E. Gut microbiome as a clinical tool in gastrointestinal disease management: are we there yet? Nat Rev Gastroenterol Hepatol. 2017;14(5):315–20.
83. Seekatz A, Young V. Clostridium difficile and the microbiota. J Clin Investig. 2014;124(10):4182–9.
84. Bajaj J, Heuman D, Sanyal A, Hylemon P, Sterling R, Stravitz R, et al. Modulation of the metabiome by rifaximin in patients with cirrhosis and minimal hepatic encephalopathy. PLoS ONE. 2013;8(4):e60042.
85. Oelschlaeger T. Mechanisms of probiotic actions – a review. Int J Med Microbiol. 2010;300(1):57–62.
86. Kristensen N, Bryrup T, Allin K, Nielsen T, Hansen T, Pedersen O. Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials. Genome Med. 2016;8(1):52.
87. Maldonado-Gómez M, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van Sinderen D, et al. Stable engraftment of Bifidobacterium longum AH1206 in the human gut depends on individualized features of the resident microbiome. Cell Host Microbe. 2016;20(4):515–26.
88. Berry D. Making it stick: a compelling case for precision microbiome reconstitution. Cell Host Microbe. 2016;20(4):415–7.
89. Gibson G, Roberfroid M. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr. 2018;125(6):1401–12.
90. Slavin J. Fiber and Prebiotics: Mechanisms and Health Benefits. Nutrients. 2013;5(4):1417–35.
91. Subramanian S, Blanton L, Frese S, Charbonneau M, Mills D, Gordon J. Cultivating healthy growth and nutrition through the gut microbiota. Cell. 2015;161(1):36–48.
92. Haarman M, Knol J. Quantitative real-time PCR assays to identify and quantify fecal bifidobacterium species in infants receiving a prebiotic infant formula. Appl Environ Microbiol. 2005;71(5):2318–24.
93. Bezkorovainy A. Probiotics: determinants of survival and growth in the gut. Am J Clin Nutr. 2001;73(2):399s–405s.
94. Suez J, Korem T, Zeevi D, Zilberman-Schapira G, Thaiss C, Maza O, et al. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature. 2014;514(7521):181–6.
95. Russell G, Kaplan J, Youngster I, Baril-Dore M, Schindelar L, Hohmann E, et al. Fecal transplant for recurrent clostridium difficile infection in children with and without inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 2014;58(5):588–92.
96. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal E, de Vos W, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med. 2013;368(5):407–15.
97. Gilbert J, Quinn R, Debelius J, Xu Z, Morton J, Garg N, et al. Microbiome-wide association studies link dynamic microbial consortia to disease. Nature. 2016;535(7610):94–103.

Index

G. J. Netto, K. L. Kaul (eds.), *Genomic Applications in Pathology*, https://doi.org/10.1007/978-3-319-96830-8

F

T