Genomic Analysis

Sally M. Hunter, Amy E. McCart Reed, Ian G. Campbell, and Kylie L. Gorringe

Introduction

Cancer is a genetic disease, from the predisposing alleles carried in the constitutive genome to the random somatic events selected for during tumorigenesis. In the last 15 years, the analysis of cancer genomes has dramatically improved in scope and level of detail. Low resolution and low-throughput methods such as G-banded karyotyping and comparative genomic hybridization have been superseded, first by array-based and more recently by sequencing-based technologies that enable affordable genome-wide single nucleotide resolution analysis of hundreds and even thousands of tumors. In a research setting this has led to novel insights regarding the initiation and evolution of cancer, and the genetic events detected are increasingly having clinically relevant implications.

This chapter introduces the main classes of genetic events that are commonly seen in cancer genomes and discusses the contemporary methodologies with which they are detected. Applying these methods has led to a number of discover-

S.M. Hunter
Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, East Melbourne, VIC, Australia

A.E. McCart Reed
The University of Queensland, UQCCR, Herston, QLD, Australia

I.G. Campbell • K.L. Gorringe (✉)
Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, East Melbourne, VIC, Australia

Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, VIC, Australia

Department of Pathology, University of Melbourne, Parkville, VIC, Australia
e-mail: ian.campbell@petermac.org; kylie.gorringe@petermac.org

ies with implications for molecular pathology, including using genetic events to evaluate cancer risk, refine diagnoses, provide prognostic information, and most critically, determining genetic events against which molecular therapeutics can be targeted.

Classes of Genetic Events in Cancer

Cancer has long been recognized as a genetic disease, since the earliest observations of deranged chromosomes [1] and familial clustering of cases [2–5]. Predisposition to cancer, initiating events, and progression are all influenced by genetics whether they be constitutive or somatic aberrations. There are many different types of genetic alteration that can occur, each of which arise through different mechanisms and each having varying consequences. The vast majority of somatic changes that occur in a tumor are thought to have little functional effect and are consequently described as "passengers," carried along by coincidence upon selection of a co-existing "driver" in the same cell [6, 7]. Discerning driver mutations from passenger mutations remains a major challenge in translating genomic data into the clinic.

Somatic Mutation

Acquired changes in the constitutional DNA sequence are common in most cancer types and include base-pair substitutions and small (<1 kb) insertion–deletions (indels). Mutations are caused by a failure of one or more of the DNA repair pathways to recognize or accurately repair DNA following a genetic insult, which can include inherent replication errors, deamination of methylated cytosine, and mutagenic exposures such as UV light. The rate of such mutations per cancer genome varies greatly depending on the cancer type and has been estimated at 0.57/Mb of coding sequence for acute lymphoblastic leukemia [7], 0.19/Mb for breast cancer [7], ~1.8/Mb for high grade ovarian cancer [8], ~18/Mb for mutagen-exposed cancers like melanoma and non-small cell lung cancer [7, 9], and as high as 400/Mb for cancers with a DNA repair defect such as loss of mismatch repair in colorectal cancer [10]. Different cancer types often have unique mutation signatures in terms of the type of mutation, e.g. UV-exposed cancers have high rates of C:G > T:A transitions resulting in an enrichment of dipyrimidines.

The impact of a somatic mutation will vary widely depending on its location (coding/non-coding/splice site/regulatory) and type of change following translation (missense, nonsense, frameshift, etc.) (Fig. 1). Some mutations will have an immediate impact and are considered dominant while others may require loss of the remaining normal allele.
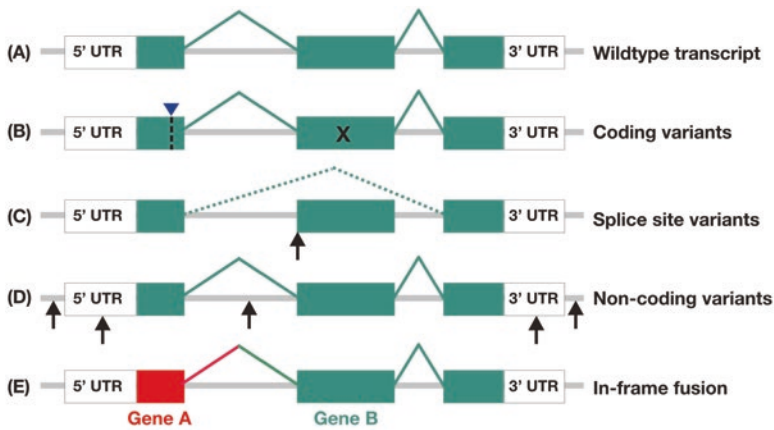
Fig. 1 Mutation types. (**a**) Wildtype transcript. *Shaded boxes* depict coding exons, *white boxes* depict the untranslated regions (UTRs) of the transcript, and intervening *grey lines* indicate intergenic and intronic regions (**b**) Coding variants: Frameshift (*triangle*) and nonsense (*X*) variants are often overtly deleterious due to protein truncation. Missense variants may be deleterious depending on the function of the specific amino acid changed and the effect on protein folding, or have no effect. Synonymous mutations do not change the amino acid identity, but may influence splice site function or binding of regulatory proteins, or have no effect. (**c**) Essential splice site variants (2 bp±intron–exon boundary, *arrow*) can result in exon skipping or cryptic exons being transcribed. (**d**) Non-coding variants (intergenic, UTRs, intronic, *arrows*) may have an effect on transcription regulatory regions, transcript splicing, and mRNA stability, or often no effect on transcript function. (**e**) Translocations resulting in an in-frame fusion can produce functional protein

Copy Number

An abnormal chromosome number (aneuploidy) is a common feature of many carcinomas. The subsequent imbalance in gene copies is thought to lead to global changes in gene expression with wide-ranging effects on cell phenotype. Aneuploidy is caused by errors in chromosome segregation during mitosis or cytokinesis, leading to gain or loss of whole chromosomes, and not uncommonly duplication of the entire chromosome complement leading to tetraploidy.

Copy number aberrations can also occur at a sub-chromosomal level through various mechanisms, often involving compromised repair of double strand (ds) DNA breaks [11, 12], a breakage-fusion-bridge cycle subsequent to dsDNA breaks or telomere dysfunction [13], and less commonly chromothripsis [14]. Copy number changes include losses of material, either hemizygous or homozygous deletions, and gains of material, which can be low level changes, such as a duplication, or high-level amplification (from five to possibly hundreds of copies) (Fig. 2), as is often observed with *ERBB2* in breast cancer.
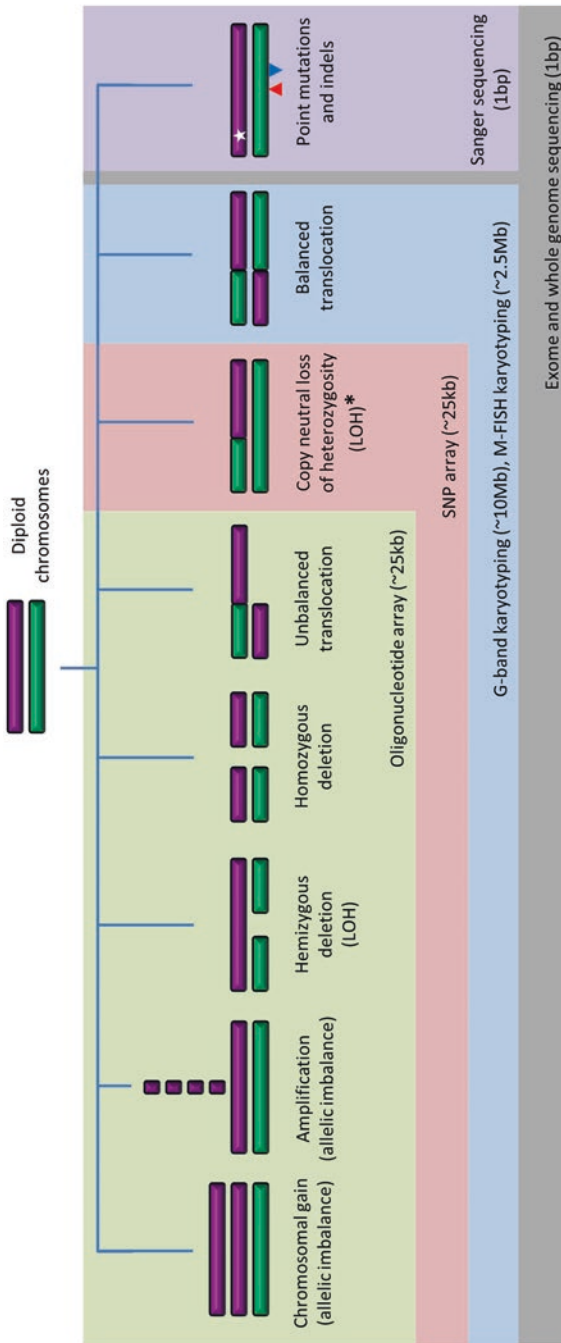
Fig. 2 Copy number aberrations and mutations. The *boxes* indicate the types of aberrations each technology is capable of detecting. The resolution of the technology indicates the smallest size of aberrations that can be robustly detected. (*Asterisk*) Detected as a copy number aberration by array methods, however, the site of the translocation fusion cannot be determined

Loss of Heterozygosity

Loss of heterozygosity (LOH) refers to the change in genotype from heterozygous to homozygous of polymorphic alleles that arises through chromosome loss, sub-chromosomal deletion or gene conversion via homologous recombination and DNA repair. LOH is often associated with copy number loss; however, gene conversion or duplication of chromosomes can lead to copy number neutral LOH (Fig. 2). LOH is distinct from allelic imbalance (AI), in which both alleles are still present but in a ratio different from 1:1 following copy number gain. The effect of LOH is to unmask recessive alleles, either inherited (e.g., *BRCA1*) or somatic (e.g., *TP53*), leading to loss of tumor suppressor gene function.

Structural Chromosome Changes

Any event involving inappropriate repair of a dsDNA break can lead to structural changes in chromosomes, including not only the aforementioned sub-chromosomal copy number changes, but also inversions and translocations. These latter events lead to the novel juxtaposition of genetic material, which can cause inappropriate gene regulation or novel protein products, such as the *BCR-ABL* translocation in chronic myelogenous leukemia [15, 16].

Germline Variation

The constitutive genetic variation carried by individuals is extensive and encompasses many of the same forms observed as somatic events. The most common class of germline aberrations most relevant to cancer are considered to be single nucleotide polymorphisms (SNPs) and small indels. It is likely that larger copy number variations and structural changes such as inversions are also important, but there are few conclusive incidences reported to date. SNPs and indels can vary widely in population frequency, from common (>10 % minor allele frequency) to rare (1–10 % frequency) to extremely rare (<1 % frequency). Most are inherited from parents, but some occur *de novo*, at a frequency estimated at 13×10^{-3}/Mb for SNPs and 0.78×10^{-3}/Mb for indels per generation [17, 18].

Methods of Genomic Analysis

High-resolution screening for genetic lesions on a genome-wide scale has only recently become feasible, i.e., on a kilobase-down to base-pair level. Prior to the invention of array and massively parallel sequencing-based methods (MPS; also

referred to as next-generation sequencing, Sanger sequencing is considered first generation sequencing), analysis resolution was limited to tens of megabase-pairs using comparative genomic hybridization (CGH) or G-banded karyotyping. In addition, these methods were time-consuming and required a high degree of individual skill to interpret the chromosome spreads, limiting the number of samples that could be studied. Other, more targeted techniques could be readily applied to multiple samples such as fluorescence in situ hybridization (FISH), Sanger sequencing, and microsatellite genotyping by PCR, but these were not easily scaled up to a genome-wide analysis. These methodologies have now been supplemented by several whole genome techniques that have delivered myriad research findings and are increasingly been applied in clinical settings (Table 1).

Karyotyping

The advent of fluorescence-based chromosome painting in the 1990s enabled a more automated karyotyping procedure compared to traditional G-banding. Variously known as Spectral karyotyping (SKY), M-FISH, and 24-color FISH, this technique uses paints made from individual flow-sorted chromosomes each labeled with a different mix of fluorophores. This paint is applied to metaphase chromosome spreads usually generated from primary tumors after short term cell culture. SKY is able to resolve complex marker chromosomes and is the only method discussed here that can measure exact ploidy (Fig. 3a). It can also give some indication of tumor genetic heterogeneity, as each nucleus is individually analyzed. However, it is still a low-resolution method (~10 Mb) and relies heavily on good quality metaphase spreads. Thus its use is limited to fresh tumor material and to laboratories with a cell culture facility.

Array-Based

All array-based methods for genomic analysis operate on the same basic principle: hybridization of a labeled DNA sample to complementary probe sequences that are immobilized to a solid substrate in known locations. The strength of the signal is proportional to the amount of each target sequence in the sample, however, with most platforms, the dynamic range tends to become compressed at high copy number and generally cannot accurately distinguish, for example, 8 copies from 12 copies.

Array-based systems all require some kind of normalization strategy to calculate copy number, where the signal intensity of the tumor sample is converted to a ratio using normal diploid samples. Normalization can be performed against matched normal DNA, which is useful for discriminating constitutional copy number variants, or against the average of multiple normal DNA samples. All data for a sample

Table 1 Methods of genomic analysis

Platform	Detects	Advantages	Disadvantages
Spectral karyotyping (SKY), M-FISH	Ploidy; translocations	Single-cell analysis; only technology not requiring comparison to normal baseline	Requires cell culture; low resolution
Multiplex ligation probe amplification (MLPA, e.g., MRC-Holland)	Ploidy; LOH	PCR-based; fast; inexpensive; simple analysis	Low resolution; each fragment requires a unique primer pair
BAC aCGH	CN	Internal control from two-color analysis	Low resolution; irregular spacing
SNP arrays (e.g., Affymetrix, Ilumina)	CN; LOH; SNP genotyping	Detect copy number neutral LOH; genotyping	Irregular probe spacing
Oligonucleotide arrays (e.g., Agilent, NimbleGen)	CN	Relatively cheap	Cannot detect copy number neutral LOH events
Molecular inversion probe arrays (e.g., Affymetrix)	CN; LOH; SNP genotyping; limited mutations	Applicable to low-quality and limited DNA; good dynamic range	Lower-density arrays
Exome sequencing (e.g., Illumina, NimbleGen)	SNP; somatic mutation data; LOH; CN	Comprehensive analysis	Complex bioinformatics to extract CN; expensive; need high quality DNA; need matching normal DNA; irregular spacing; risk of incidental findings
Whole genome sequencing (low read depth)	CN; translocations	Good dynamic range	Expensive; low sensitivity for SNP and mutation data
Whole genome sequencing (high read depth)	SNP; somatic mutation data; LOH; CN; translocation	Comprehensive analysis; base-pair resolution of breakpoints	Very expensive; complex bioinformatics; risk of incidental findings

M-FISH multi-color fluorescence in-situ hybridization, *BAC* bacterial artificial chromosome, *aCGH* array comparative genomic hybridization, *CN* copy number, *LOH* loss of heterozygosity, *SNP* single nucleotide polymorphism

are median- or mean-centered, which means that it is not possible to distinguish between perfectly tetraploid and diploid samples, and exact ploidy cannot be determined unless genotype information is also available (see below).

The level of genomic resolution of array-based platforms is inherently limited by the number and type of probe sequences selected. Initially, detection of copy number was done using cDNA arrays produced for expression analysis but these were quickly superseded by superior platforms including bacterial artificial chromosome (BAC) arrays, oligonucleotide arrays, and SNP arrays.
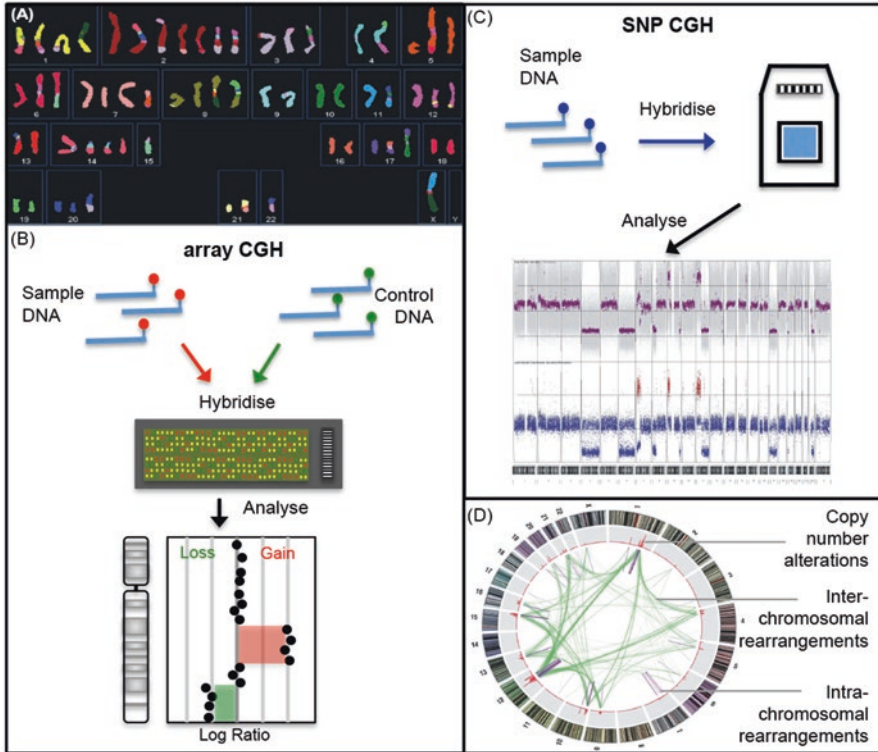
Fig. 3 Genomic analysis data. (**a**) Spectral karyotyping of the breast cancer cell line VP229, demonstrating aneuploidy and extensive translocations and structural rearrangements. (**b**) Array CGH method overview, with data analysis of one representative chromosome (log ratio data plotted), indicating where regions of chromosome have been gained (*red*) or lost (*green*). (**c**) SNP CGH method overview, with Partek® plot of Affymetrix® SNP6™ data from an ovarian tumor, showing chromosomes linearly mapped from 1 through X showing total copy number (*upper*) and allele-specific copy number (*lower*). (**d**) *Circos plot* demonstrating typical genomic copy number aberrations and structural rearrangements in the 94778 cell line derived from a retroperitoneal relapse of a well-differentiated liposarcoma [19] 94778 cells were provided by Florence Pedeutour (Laboratory of Solid Tumors Genetics, Nice University Hospital, Nice, France). The data was analyzed and figure generated by Anthony Papenfuss (Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia, courtesy of D. Thomas and D. Garsed. *CGH* comparative genomic hybridization, *SNP* single nucleotide polymorphism

BACs are large plasmid vectors with inserts of tens to hundreds of kilobases but because of the way they are constructed, they tend to be irregularly spaced across the genome, although whole genome tiling arrays with better uniformity have been produced. The resolution of BAC arrays is also limited by the large insert size. However, the signal-to-noise ratio and dynamic range are generally good.
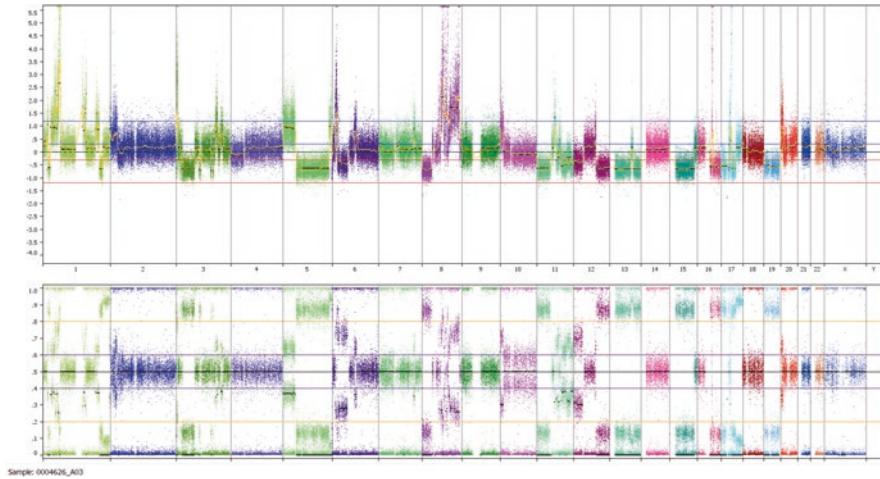
Fig. 4 Copy number analysis of FFPE DNA. DNA extracted from ductal carcinoma in situ assayed using OncoScan™ (Affymetrix®). The *upper panel* illustrates total copy number, while the *bottom panel* illustrates allele ratios. Detectable copy number aberrations include whole chromosome and chromosome arm gains and losses, focal deletions, and high-level amplifications

As a consequence of the completion of the sequence of the entire human genome in the early 2000s, synthetic oligonucleotide-based arrays have become possible. These arrays utilize relatively short (25–60 bp) synthesized oligonucleotides, overcoming issues with appropriately spacing probes throughout the genome, and increasing the resolution of breakpoint detection. Array comparative genomic hybridization (aCGH; Fig. 3b) typically refers to the use of non-polymorphic oligonucleotide arrays, which can detect changes in total copy number but not changes in allelic ratios and therefore cannot identify copy number neutral LOH (Fig. 2). Single nucleotide polymorphism (SNP) arrays, a type of oligonucleotide array, allow the detection of both total copy number and LOH by designing probes spanning known common polymorphisms in the human genome (Fig. 3c).

Oligonucleotide-based arrays perform well on good quality DNA from fresh or frozen tissues and lymphocyte DNA; however, they are less useful for degraded DNA such as that obtained from formalin-fixed, paraffin-embedded (FFPE) tissue.

DNA from FFPE sources is predominantly highly fragmented and alternative approaches have been successfully developed for its analysis (Fig. 4). Molecular inversion probe (MIP) technology has been incorporated into the Affymetrix® OncoScan™ assay. This approach involves circularizable "padlock" probes with two terminal sequences that bind to homologous sequences either side of an SNP, followed by highly specific closing of the padlock through incorporation of a nucleotide complementary to the SNP. This approach circumvents the issue of having to directly digest, ligate, and amplify the fragmented DNA. An alternative approach, taken by Illumina®, is to "restore" the FFPE DNA by ligating the fragmented DNA together to generate fragments large enough for whole genome amplification prior to labeling and hybridization. This DNA can then be hybridized to a standard bead array.

Sequencing-Based

Sanger Sequencing

In the mid-1970s, Fredrick Sanger first described "Sanger Sequencing," and since this time huge advances in the technology and its applications have been made, with the technology notably underpinning the Human Genome Project. The current iteration of the technology is based on sequencing by synthesis, with the products resolved using capillary electrophoresis and laser optics. Selective incorporation of fluorescently labeled, chain terminating dideoxynucleotides occurs during modified PCR amplification, and following purification of the products, they are resolved based on size using a capillary sequencer. Lasers excite the fluorescent dyes, and sequences of up to 700–900 bases can be determined, which are exported as a chromatogram (Fig. 5a).

Massively Parallel Sequencing

The advent of massively parallel sequencing (MPS), or the so-called next-generation sequencing, represents an enormous step forward in the genomics field. While continuing to "sequence by synthesis," these technologies vastly increase throughput by simultaneously sequencing multiple DNA strands (in "parallel"). Unlike Sanger sequencing, MPS generates millions of random short reads (35–700 bp) that must then be mapped to a reference genome (Fig. 5b). The most commonly employed technologies in cancer research are ion semi-conductor sequencing and optics-based dye sequencing. Semi-conductor sequencing (Ion Torrent Systems Inc.) measures the hydrogen release following incorporation of a nucleotide as determined by the template strand of DNA. Conversely, dye sequencing (Illumina®) relies on the immobilization of template DNA clusters onto a solid surface, upon which fluorescently labeled nucleotides competitively bind; a laser excites the label, and images of incorporated bases are recorded. Prior to sequencing, DNA samples are prepared into libraries, representing all of the desired DNA target sequences, be they the entire genome (whole genome sequencing, WGS), only the exons (whole exome sequencing, WES), or a targeted panel (Fig. 6). A targeted panel allows for specific enrichment of certain sequences; the enrichment can be performed using either a nucleic acid bait ("capture") or through PCR-based amplification. It is at the library preparation stage that samples can be barcoded and pooled to increase throughput.

Using MPS, it is possible to simultaneously detect single nucleotide mutations and SNPs, as well as small indels, large copy number aberrations, LOH, and structural rearrangements, depending on the type of sequencing performed (Fig. 2). Variants are identified ("called") by programs that assess the sequence evidence for the particular variant (read depth, base quality, etc.) and are exported into a mutation annotation format (MAF) file. This calling procedure is typically performed against a matched normal for the detection of somatic variants as the number of non-reference
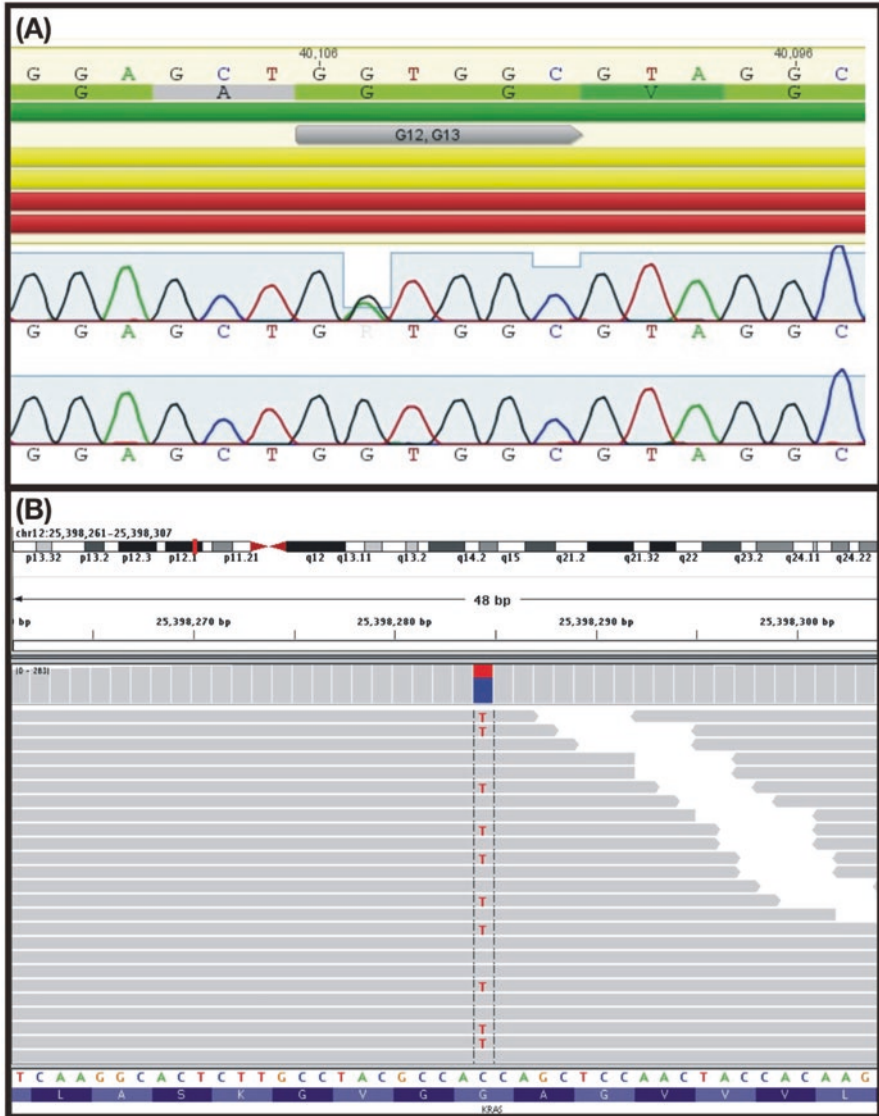
Fig. 5 Sequencing data. (**a**) Geneious (Biomatters) browser view of Sanger sequencing traces of wildtype *KRAS* (*lower*) and *KRAS* c.34G>A, p.G12D mutant (*upper*). (**b**) IGV browser view of MPS reads mapped to *KRAS* demonstrating wildtype and mutant (T) reads (*KRAS* c.34G>A, p.G12D mutant)

germline variants in any given individual is substantial. Copy number and LOH events are assessed by comparing read depths and allele frequencies from the tumor to those from the matched normal sample. Structural rearrangements (translocations) and large indels are identified by assessing paired reads that did not map
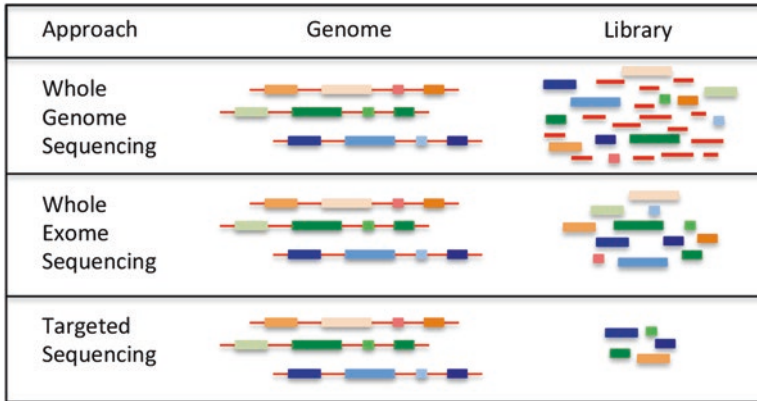
Fig. 6 MPS sequencing. Schematic of the input genome (intronic DNA in *red*, exons in shades of *blue*, *green* and *orange*) and the sequences represented in libraries generated for the three major types of sequencing; whole genome, whole exome, and targeted

within the expected distance from each other (determined by the average fragment length of the DNA library prepared for sequencing) or that mapped to different chromosomes (Fig. 3c).

Limitations of Genomic Analyses

All genomic analyses of tumors using dissociated and homogenized tissues suffer potential dilution of tumor-derived genomic events by the genomes of surrounding non-neoplastic cells. Estimating the percentage of tumor cells in a sample and enriching for tumor cells using laser or needle microdissection or selective enrichment using a tumor-specific cell surface marker (e.g., EpCam) is an important process upstream of genomic analysis. Heterogeneity of genomic events within the tumor cell population can also contribute to dilution of signals, resulting in subpopulations not being discernible at any great resolution.

Sanger sequencing and array-based copy number outputs are an average of the genomes of all of the cells from which DNA has been isolated (Figs. 3b and 5a) and therefore have limited sensitivity to detect events occurring only in a subpopulation of tumor cells. MPS, particularly at very high read depth, offers greater scope to resolve events occurring in a small subpopulation of cells and gives a digital count of variant reads (Fig. 5b). Paired-end sequencing also allows the mapping of translocation events, which cannot be resolved using array technology where chromosomes are linearly mapped (Fig. 3b, c). MPS is not without its own issues; a problematic area of MPS is the accurate mapping of reads to a reference genome and the calling of variants. Reads become difficult to correctly map to areas of the genome that are highly homologous to other regions, have repetitive sequence, or when there is an indel in the read or region relative to the reference.

Whole genome and exome sequencing provide the possibility of simultaneously detecting all variants in the genome or all coding variants. Along with real variants, PCR and sequencing artifacts are also detected, resulting in a huge number of potential variants to analyze for validity, recurrence, and functional impact. Differentiating between mutations that are driving tumorigenesis ("drivers") from those that are not anticipated to have any involvement in the development of a tumor ("passengers") is difficult and is made more onerous because passenger mutations are predicted to far outnumber driver mutations. The bioinformatic analysis burden of MPS should not be underestimated, although efforts to improve software design and usability are ongoing (see Chapter "Bioinformatics Analysis of Sequence Data").

Applications of Genomic Analysis in Cancer

Genome-Wide Association Studies

One of the most common uses of SNP arrays has been to the application of genome-wide association studies (GWAS), where linkage of SNPs to an increased (or decreased) risk of disease is assessed across the genome. Due to cost constraints, these studies are most often performed in a staged process, where a few hundred or thousand individuals with features suggestive of a genetic predisposition to cancer such as family history or early age of onset are first compared to age- and ethnicity-matched controls. Significant hits from this analysis are then validated in tens of thousands of cases and controls. This strategy has been applied to all common cancer types, with multiple predisposing SNPs identified in breast, colorectal, lung, and ovarian cancers. The risks associated with individual SNPs are usually low (1.2–1.5-fold above the general population), however, the polygenic risk when multiple low-risk SNPs are inherited together can reach much greater significance. In addition, because the resolution of the studies is limited by the array density, the SNPs with the highest risk association may not be the causative variant, but only closely linked. Thus, fine mapping is required for more precise information on the gene affected and the possible mechanism of the increase in risk. Nonetheless, some risk alleles, which have been validated in multiple independent cohorts, are now being utilized for testing in familial cancer clinics [20].

More recently, as the cost of MPS continues to fall, exome and genome sequencing of large cohorts are being undertaken. For example, large databases have been established with a focus on thoroughly characterizing common cancers [International Cancer Genome Consortium (ICGC), The Cancer Genome Atlas (TCGA)] and providing a population baseline for common and rare variants [Exome Variant Server (EVS), 1000 Genomes, dbSNP, Exome Aggregation Consortium (ExAC)]. These studies will have the advantage of being able to detect rare variants and causative alleles; however, it may be some years before they are powerful enough to identify rare, low to moderate risk alleles.

Mapping of Oncogenes and Tumor Suppressor Genes

A major goal of genomic analyses in the research setting is the discovery of the full repertoire of genes with a role in tumorigenesis. These genes can be tumor promoting when their activity is deregulated (oncogenes) or when they are inactivated (tumor suppressors [TSG]). Some genes can act as either oncogene or TSG depending on the cellular context and the pathways driving tumorigenesis; for example, *NOTCH1* is targeted by activating mutations in hematopoietic malignancies [21] and inactivating mutations in solid tumors such as head and neck squamous cell carcinoma [22].

Oncogenes commonly act in a dominant fashion, with the genetic aberration ranging from copy number increase (e.g. *ERBB2*), recurrent activating point mutation (e.g. *KRAS*, *BRAF*) (Fig. 7a, b), translocation (e.g. *BCR-ABL*), or other structural chromosomal changes leading to loss of transcriptional (e.g. *MYC*) or post-translational control (e.g. *EGFR*). These types of recurrent activating events typically make oncogenes easier to design clinical tests for (compared to TSGs) because there is a limited number of functionally relevant mutational events.

Methods for discovering new oncogenes include mapping regions of copy number gain, exome sequencing for somatic mutations, and karyotyping or genome sequencing for structural chromosome changes. Regardless of methodology, a common challenge is distinguishing the driving genetic events from benign passenger events.

Identifying genes affected by copy number alterations has been most effectively achieved using array technologies. The increases in copy number are mapped in multiple samples, and those regions of the genome that most often display increases in copy number are short-listed as potential sites of oncogenes. However, this is complicated by the degree of copy number change—should any increase be investigated even if only a single copy, or should only high-level amplifications be considered? Both methods have been applied, and bioinformatic techniques that balance both possibilities have been developed (e.g. GISTIC [23, 24]). The list of genes in minimal regions of copy number change can still be long, and expression and functional analyses are then required to identify putative drivers. For example, integrated copy number and expression analysis identified novel growth promoting genes in ovarian carcinomas [25] and candidate oncogenes driving ovarian cancer were functionally investigated using RNA interference [26].

Full genome sequencing is the most comprehensive and sensitive method to identify structural chromosome changes, although to date fusion genes have also been detected using the much cheaper approach of RNAseq to short-cut to those translocations with an expressed gene moiety. For example, RNAseq analysis identified the MHC II transactivator CIITA as a recurrent fusion partner in lymphoid cancers [27].

Tumor suppressor genes are characterized by loss-of-function genetic events (Fig. 7c, d). Apart from a few examples where dominant negative mutations can be selected for (e.g. *TP53*), it is usually expected that both copies of a tumor suppressor gene must be inactivated, either through bi-allelic point mutation, homozygous deletion, methylation, or a combination of mutation and LOH. Thus, mapping of
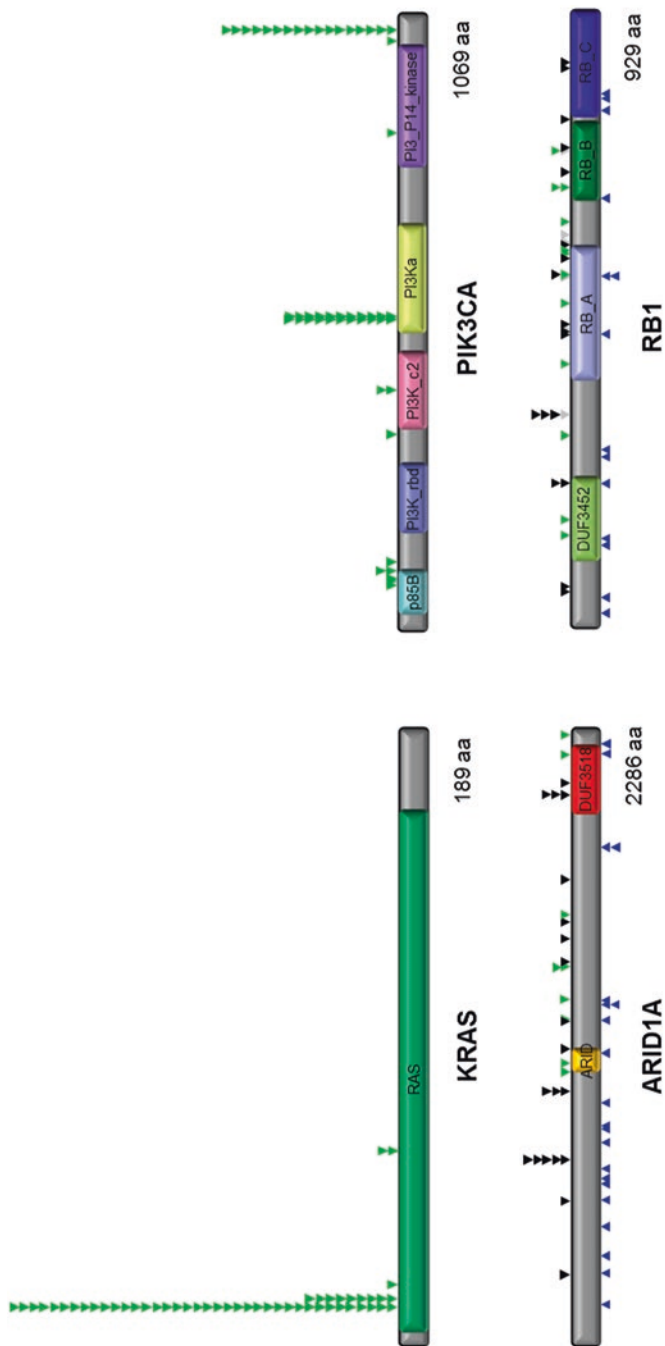
Fig. 7 Oncogene and tumor suppressor gene mutation patterns. Distribution of mutations in the oncogenes (**a**) *KRAS* and (**b**) *PIK3CA* and TSGs (**c**) *ARID1A* and (**d**) *RB1*. *Green arrowheads* indicate missense mutations, *black arrowheads* indicate nonsense mutations, *grey arrowheads* synonymous mutations, while *blue arrowheads* indicate truncating indels. Mutation patterns are based on 50 randomly selected mutations from the COSMIC database (COSMIC http://cancer.sanger.ac.uk/cosmic)

copy number loss and LOH has been applied to try and identify new TSGs. While early successes included genes where the initial mutation event was inherited (e.g. *RB1* [28]), in the genomic age there have been very few genes identified through this method [29].

Exome sequencing studies have been applied to multiple cancer types to identify both oncogenes and tumor suppressor genes. Initially, only a small number of samples were investigated, with candidates followed up in larger cohorts (e.g., CANgenes [30]). More recently, as sequencing has become relatively cheap, cohorts of hundreds of samples have been analyzed. Interestingly, apart from a few histologically defined tumor types (e.g., granulosa cell tumors) there have been very few genes identified that are mutated at high frequency. It seems that for solid tumors each tissue type has 1–5 commonly mutated genes (>10 % frequency), and a long tail of genes each with a mutation frequency of just a few percent. Thus, the issue of identifying drivers versus passengers is again a problem. One strategy used to enrich for potential driver mutations is the employment of algorithms to predict the deleteriousness of an SNV given the nature of the amino acid change, the position within the protein sequence and the level of conservation of the protein sequence compared to other species [31–33]. Another strategy is the use of statistical methods to assess the mutation rate for a given gene relative to the background mutation rate and gene size [34, 35]. Increasingly, gene discovery studies are applying algorithms to identify common pathways that are affected which can assist in identifying the likely driver genes. For example, pathway analysis identified axonal guidance pathway aberrations in pancreatic cancer, revealing novel tumorigenic roles for these proteins [36].

Association of Genetic Events with Clinical Features

Genetic events are intrinsic to the development of malignant characteristics, thus, it is logical to assume that differences in clinical behavior may be attributed to specific genetic aberrations. Many studies have investigated the association of clinical with genetic features on a genome-wide scale. Associated features may then assist in prediction and risk management, diagnosis, prognosis, or treatment.

Germline Predisposition to Cancer

Many of the well-known cancer predisposition genes, such as *APC* (familial adenomatous polyposis), *BRCA1* and *BRCA2* (hereditary breast and ovarian cancer), and *MLH1* (hereditary non-polyposis colon cancer), were identified through linkage analysis and candidate gene approaches [37–40]. This was possible because of their relative commonness and high penetrance in these hereditary conditions. Identifying additional candidates is now primarily undertaken through large-scale exome and genome sequencing of multiple members of high-risk families. However, the task of

identifying these genes remains difficult since pathogenic mutations are often vanishingly rare, as encountered with *RAD51C* mutations in *BRCA1/2* mutation-negative breast/ovarian cancer families [41–43], and definitive classification as a cancer predisposition requires very large case and control cohorts to achieve sufficient power.

Genome-wide association studies (GWAS) have identified many more common genomic variants with much lower individual effect on cancer risk. Although the functionally relevant genes may not be identified, the SNPs from these studies can prove useful for risk prediction. Common risk alleles may act in concert to produce a multiplicative polygenic risk or act as risk modifiers [20].

In order to effectively incorporate new risk alleles into the clinic, current practices for genetic testing are undergoing a shift towards gene panels, where all known cancer susceptibility genes and SNPs can be sequenced simultaneously using MPS. This approach substantially decreases the time and cost per gene tested.

Molecular Subtyping and Diagnostics

Most subtyping studies have used expression microarrays to determine classes of tumors with distinct characteristics; however, it is becoming clear that these expression subtypes are often correlated with specific underlying genetic profiles. For example, a number of prognostic tests have been developed for breast cancer subtyping (e.g., OncotypeDX, MammaPrint, and PAM50) and these reflect both histological and genetic differences [44–46]. Targeted sequencing panels are increasingly being used to inform clinical decision-making by matching patients with appropriate conventional therapies or to direct patients to relevant clinical trials. Many biotechnology companies offer companion diagnostic cancer gene panels, enriched for the so-called druggable mutations and those associated with prognostication, including, for example, Illumina® (TruSight® Cancer/Tumor; TruSeq® Cancer Amplicon [47]); Foundation One™ [48]; and Ion Torrent (IonAmpliSeq™ Comprehensive Cancer/Cancer HotSpot [49]).

Prognostic Markers

Treatment of cancer tends to be aggressive, with side effects that can have a severe impact on the quality of life both in the short and long term, including radical surgery leading to scarring and loss/reduction of organ function, radiotherapy-induced burns and increased risk of subsequent malignancy, systemic cytotoxics leading to hair loss, nausea, etc. The consequences of disease progression or recurrence are sufficiently severe that these outcomes are accepted as a necessary evil. However, not all patients are at the same risk of progression, even after controlling for known prognostic factors such as stage, grade, and histological subtype. Genomic analysis

has been applied to attempt to identify robust prognostic markers that may indicate that an aggressive treatment regime may not be necessary. For example, the presence of microsatellite instability in colorectal cancer has been shown to be a good prognostic indicator, identifying a proportion of colorectal tumors that do not respond to 5-fluorouracil (5-FU) systemic treatment, the mainstay of colorectal cancer systemic therapy [50, 51]. At the clinical level these data mean that individuals with stage II mismatch-deficient colorectal cancer are unlikely to be treated with systemic 5-FU treatment compared with mismatch repair proficient tumors as the clinical benefits do not outweigh the complications associated with this treatment.

Pharmacogenomics

Response to Conventional Therapies

In a similar manner to identifying prognostic markers of general tumor aggressiveness, studies have also tried to find markers that indicate a likely response to chemotherapies. Such a marker could be constitutional, for example, polymorphisms in cell transporter channels that affect the rate of drug efflux are strong determinants of chemotherapy toxicity and tolerable dosage [52–54]. Alternatively, deleterious germline mutations in *BRCA1* and *BRCA2* that are cancer predisposing paradoxically tend to improve the patient's response to treatment due to a heightened susceptibility to the DNA damage caused by chemotherapy. Alternatively, response to therapies could be tumor-intrinsic, for example, *CCNE1* gene amplification was determined to be an intrinsic resistance mechanism to platinum-taxol-based chemotherapy in high grade serous ovarian cancer [55].

Targeted Molecular Therapeutics

Recently, targeted molecular therapies have emerged with the potential to transform cancer treatment by personalizing drug regimens to the genetic "Achilles heels" of each tumor. Genome-wide analyses are key to identifying such targets in a research setting, and could be used clinically in the future, especially to identify the cause of therapy resistance. Obvious candidates for targeted therapies are over-active oncogenes as reducing activity is theoretically straightforward. Some prime examples of successful targeted treatments are imatinib (Glivec), which acts as an inhibitor of several tyrosine kinases including the BCR-ABL fusion, trastuzumab (Herceptin), targeted against overexpressed HER2 (first used in HER2+ breast cancers), and PLX 4032 (Vemurafenib), which targets the constitutively active form of BRAF (BRAF V600E) frequently mutated in melanoma. Targeted therapies for gene products where function has been lost tend to rely on unique weaknesses arising as a

side-effect of the loss of gene function. For example, deleterious mutations in homologous recombination repair genes *BRCA1/BRCA2/PALB2* that impede the efficient repair of DNA double-stranded breaks leave the cells susceptible to both conventional DNA damaging chemotherapies (e.g. cisplatin) and more molecularly targeted poly ADP ribose polymerase (PARP) inhibitors, which affect alternative DNA repair pathways and impede subsequent DNA replications [56].

Despite the breakthroughs in targeted molecular therapies, these almost always induce drug resistance and are often not directly transferable to other tumor types characterized by the same mutation or pathway alteration. For example, attempts to treat BRAF V600E positive colon cancers with the same BRAF inhibitors that had been successful in melanoma resulted in poor clinical response rates due to feedback activation of EGFR in response to BRAF inhibition [57]. However, in this case combination therapy with BRAF inhibitors and EGFR or PI3K inhibitors looks more promising [58].

Cancer cells can become resistant through a range of mechanisms, finding alternative ways to activate pathways or undergoing secondary mutations that reverse susceptibility. For example, initially successful treatment of colon cancer patients with EGFR inhibitors has been found to select for cancer cells with activating *KRAS* mutations, leading to bypassing of the receptor tyrosine kinase signal and resistance to EGFR inhibition [59]. Secondary mutations in *BRCA1* and *BRCA2* have been detected in chemotherapy resistant ovarian cancers, that result in restoration of protein function through reestablishment of the reading frame, mutation of deleterious nonsense codons to missense codons, and gene conversion where the mutant allele is lost [60, 61]. Detection of these resistance mechanisms is crucial for patient prognosis and identifying effective treatments for patients to progress.

Future and Near-Term Clinical Applications

New technologies are expediting the identification of cancer driver genes and potential new therapeutic targets, leading genomics to take center stage in diagnosis, prognosis, treatment planning, and the search for new treatment options. The possibility of affordable whole genome sequencing is likely to result in many current clinical tests becoming ancillary and potentially redundant.

With the advent of accessible genome-wide molecular analysis, the molecular subtyping of all cancer types using next-generation DNA and RNA sequencing, and copy number and expression arrays is currently being realized. This offers the possibility of mutation, copy number, and expression profiles superseding histological classification, particularly concerning selection of the most effective treatment options and prediction of recurrence risk.

Whole genome and whole exome sequencing of germline and tumor DNA are likely to become standard practice, both for the identification of predisposing genetic variation and to identify molecular targets for treatment and potential resistance mechanisms. Before these technologies become standard clinical techniques,

however, there are the ethical and legal hurdles of incidental findings and patents concerning certain cancer predisposition genes. In the meantime, clinical tests are being converted to these modern technologies with the creation of high-throughput panels of cancer genes.

Advancements in genomics technologies that allow very limiting amounts of DNA to be sequenced are providing future potential for real-time monitoring of treatment response and development of resistance. Isolation of circulating tumor DNA from plasma offers a non-invasive "liquid biopsy" that gives an indication of tumor burden and provides a more representative sampling of the tumor cell population than traditional core biopsies [62]. Highly sensitive monitoring of patients at the molecular level as they progress through treatment and altering treatment based on resistance mutations as they arise could drastically alter the outcome for many cancer patients. These applications are currently under investigation and offer a paradigm shift in cancer screening and treatment in the near future [62].

References

1. Boveri T (2008) Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci 121(Suppl 1):1–84
2. Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ (1966) Hereditary factors in cancer. Study of two large midwestern kindreds. Arch Intern Med 117:206–212
3. Li FP, Fraumeni JF Jr (1969) Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Ann Intern Med 71:747–752
4. Macklin MT (1959) Comparison of the number of breast-cancer deaths observed in relatives of breast-cancer patients, and the number expected on the basis of mortality rates. J Natl Cancer Inst 22:927–951
5. Macklin MT (1960) Inheritance of cancer of the stomach and large intestine in man. J Natl Cancer Inst 24:551–571
6. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA (2010) Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 107:18545–18550
7. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR (2007) Patterns of somatic mutation in human cancer genomes. Nature 446:153–8
8. Bell D, Berchuck A, Birrer MJ, Thomson E (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474:609–15

9. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 465:473–477

10. Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–7

11. Myllykangas S, Himberg J, Bohling T, Nagy B, Hollmen J, Knuutila S (2006) DNA copy number amplification profiling of human neoplasms. Oncogene 25:7324–7332

12. Weaver Z, Montagna C, Xu X, Howard T, Gadina M, Brodie SG, Deng CX, Ried T (2002) Mammary tumors in mice conditionally mutant for Brca1 exhibit gross genomic instability and centrosome amplification yet display a recurring distribution of genomic imbalances that is similar to human breast cancer. Oncogene 21:5097–5107

13. Lo AW, Sabatier L, Fouladi B, Pottier G, Ricoul M, Murnane JP (2002) DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. Neoplasia 4:531–538

14. 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 (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144:27–40

15. Heisterkamp N, Stam K, Groffen J, de Klein A, Grosveld G (1985) Structural organization of the bcr gene and its role in the Ph′ translocation. Nature 315:758–761

16. Shtivelman E, Lifshitz B, Gale RP, Canaani E (1985) Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature 315:550–554

17. Lynch M (2010) Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci U S A 107:961–968

18. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Wong WS, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471–475

19. Pedeutour F, Forus A, Coindre JM, Berner JM, Nicolo G, Michiels JF, Terrier P, Ranchere-Vince D, Collin F, Myklebost O, Turc-Carel C (1999) Structure of the supernumerary ring and giant rod chromosomes in adipose tissue tumors. Genes Chromosomes Cancer 24:30–41

20. Sawyer S, Mitchell G, McKinley J, Chenevix-Trench G, Beesley J, Chen XQ, Bowtell D, Trainer AH, Harris M, Lindeman GJ, James PA (2012) A role for common genomic variants in the assessment of familial breast cancer. J Clin Oncol 30:4330–4336

21. Kridel R, Meissner B, Rogic S, Boyle M, Telenius A, Woolcock B, Gunawardana J, Jenkins C, Cochrane C, Ben-Neriah S, Tan K, Morin RD, Opat S, Sehn LH, Connors JM, Marra MA, Weng AP, Steidl C, Gascoyne RD (2012) Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 119:1963–1971

22. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Trevino L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333:1154–1157

23. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, Du J, Kau T, Thomas RK, Shah K, Soto H, Perner S, Prensner J, Debiasi RM, Demichelis F, Hatton C, Rubin MA, Garraway LA, Nelson SF, Liau L, Mischel PS, Cloughesy TF, Meyerson M, Golub TA, Lander ES, Mellinghoff IK, Sellers WR (2007)

Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A 104:20007–20012

24. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12:R41

25. Ramakrishna M, Williams LH, Boyle SE, Bearfoot JL, Sridhar A, Speed TP, Gorringe KL, Campbell IG (2010) Identification of candidate growth promoting genes in ovarian cancer through integrated copy number and expression analysis. PLoS One 5, e9983

26. Davis SJ, Sheppard KE, Pearson RB, Campbell IG, Gorringe KL, Simpson KJ (2013) Functional analysis of genes in regions commonly amplified in high-grade serous and endometrioid ovarian cancer. Clin Cancer Res 19:1411–1421

27. Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P, Johnson NA, Zhao Y, Telenius A, Neriah SB, McPherson A, Meissner B, Okoye UC, Diepstra A, van den Berg A, Sun M, Leung G, Jones SJ, Connors JM, Huntsman DG, Savage KJ, Rimsza LM, Horsman DE, Staudt LM, Steidl U, Marra MA, Gascoyne RD (2011) MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature 471:377–381

28. Cavenee WK, Hansen MF, Nordenskjold M, Kock E, Maumenee I, Squire JA, Phillips RA, Gallie BL (1985) Genetic origin of mutations predisposing to retinoblastoma. Science 228:501–503

29. Gorringe KL, Ramakrishna M, Williams LH, Sridhar A, Boyle SE, Bearfoot JL, Li J, Anglesio MS, Campbell IG (2009) Are there any more ovarian tumor suppressor genes? A new perspective using ultra high-resolution copy number and loss of heterozygosity analysis. Genes Chromosomes Cancer 48:931–942

30. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274

31. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet, Chapter 7: Unit 7.20

32. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

33. Gonzalez-Perez A, Lopez-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88:440–449

34. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, Ding L (2012) MuSiC: identifying mutational significance in cancer genomes. Genome Res 22:1589–1598

35. Parmigiani G, Boca S, Lin J, Kinzler KW, Velculescu V, Vogelstein B (2009) Design and analysis issues in genome-wide somatic mutation studies of cancer. Genomics 93:17–21

36. 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, Lawlor RT, Capelli P, Corbo V, Scardoni M, Tortora G, Tempero MA, Mann KM, Jenkins NA, Perez-Mancera PA, Adams DJ, Largaespada DA, Wessels LF, Rust AG, Stein LD, Tuveson DA, Copeland NG, Musgrove EA, Scarpa A,

Eshleman JR, Hudson TJ, Sutherland RL, Wheeler DA, Pearson JV, McPherson JD, Gibbs RA, Grimmond SM (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491:399–405

37. Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A et al (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 368:258–261

38. Okamoto M, Sato C, Kohno Y, Mori T, Iwama T, Tonomura A, Miki Y, Utsunomiya J, Nakamura Y, White R et al (1990) Molecular nature of chromosome 5q loss in colorectal tumors and desmoids from patients with familial adenomatous polyposis. Hum Genet 85:595–599

39. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66–71

40. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789–792

41. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, Hartmann L, Schaal H, Ramser J, Honisch E, Kubisch C, Wichmann HE, Kast K, Deissler H, Engel C, Muller-Myhsok B, Neveling K, Kiechle M, Mathew CG, Schindler D, Schmutzler RK, Hanenberg H (2010) Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet 42:410–414

42. Thompson ER, Boyle SE, Johnson J, Ryland GL, Sawyer S, Choong DY, kConFab, Chenevix-Trench G, Trainer AH, Lindeman GJ, Mitchell G, James PA, Campbell IG (2012) Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. Hum Mutat 33:95–99

43. Osorio A, Endt D, Fernandez F, Eirich K, de la Hoya M, Schmutzler R, Caldes T, Meindl A, Schindler D, Benitez J (2012) Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. Hum Mol Genet 21:2889–2898

44. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99:12963–12968

45. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418–8423

46. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352

47. Jones NL, Xiu J, Reddy SK, Burke WM, Tergas AI, Wright JD, Hou JY (2015) Identification of potential therapeutic targets by molecular profiling of 628 cases of uterine serous carcinoma. Gynecol Oncol 138:620–626

48. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, An P, Sun J, Juhn F, Brennan K, Iwanik K, Maillet A, Buell J, White E, Zhao M, Balasubramanian S, Terzic S, Richards T, Banning V, Garcia L, Mahoney K, Zwirko Z, Donahue A, Beltran H, Mosquera JM, Rubin MA, Dogan S, Hedvat CV, Berger MF, Pusztai L, Lechner M, Boshoff C, Jarosz M, Vietz C, Parker A, Miller VA, Ross JS, Curran J, Cronin MT, Stephens PJ, Lipson D, Yelensky R (2013) Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 31:1023–31

49. Bartels S, Schipper E, Kreipe HH, Lehmann U (2015) Comprehensive molecular profiling of archival bone marrow trephines using a commercially available leukemia panel and semiconductor-based targeted resequencing. PLoS One 10, e0133930
50. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, French AJ, Kabat B, Foster NR, Torri V, Ribic C, Grothey A, Moore M, Zaniboni A, Seitz JF, Sinicrope F, Gallinger S (2010) Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. J Clin Oncol 28:3219–3226
51. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, Hamilton SR, Laurent-Puig P, Gryfe R, Shepherd LE, Tu D, Redston M, Gallinger S (2003) Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med 349:247–257
52. Deeken JF, Cormier T, Price DK, Sissung TM, Steinberg SM, Tran K, Liewehr DJ, Dahut WL, Miao X, Figg WD (2010) A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. Pharmacogenomics J 10:191–199
53. Tian C, Ambrosone CB, Darcy KM, Krivak TC, Armstrong DK, Bookman MA, Davis W, Zhao H, Moysich K, Gallion H, DeLoia JA (2012) Common variants in ABCB1, ABCC2 and ABCG2 genes and clinical outcomes among women with advanced stage ovarian cancer treated with platinum and taxane-based chemotherapy: a Gynecologic Oncology Group study. Gynecol Oncol 124:575–581
54. Ni LN, Li JY, Miao KR, Qiao C, Zhang SJ, Qiu HR, Qian SX (2011) Multidrug resistance gene (MDR1) polymorphisms correlate with imatinib response in chronic myeloid leukemia. Med Oncol 28:265–269
55. Etemadmoghadam D, de Fazio A, Beroukhim R, Mermel C, George J, Getz G, Tothill R, Okamoto A, Raeder MB, Harnett P, Lade S, Akslen LA, Tinker AV, Locandro B, Alsop K, Chiew YE, Traficante N, Fereday S, Johnson D, Fox S, Sellers W, Urashima M, Salvesen HB, Meyerson M, Bowtell D (2009) Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. Clin Cancer Res 15:1417–27
56. Helleday T (2011) The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. Mol Oncol 5:387–393
57. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R (2012) Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature 483:100–103
58. Mao M, Tian F, Mariadason JM, Tsao CC, Lemos R Jr, Dayyani F, Gopal YN, Jiang ZQ, Wistuba II, Tang XM, Bornman WG, Bollag G, Mills GB, Powis G, Desai J, Gallick GE, Davies MA, Kopetz S (2013) Resistance to BRAF inhibition in BRAF-mutant colon cancer can be overcome with PI3K inhibition or demethylating agents. Clin Cancer Res 19:657–67
59. Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA, Kinzler KW, Oliner KS, Vogelstein B (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 486:537–540
60. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, Villegas E, Jacquemont C, Farrugia DJ, Couch FJ, Urban N, Taniguchi T (2008) Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature 451:1116–1120
61. Swisher EM, Sakai W, Karlan BY, Wurz K, Urban N, Taniguchi T (2008) Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance. Cancer Res 68:2581–2586
62. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 368:1199–1209