Advances in Experimental Medicine and Biology 1255

Buwei Yu · Jiaqiang Zhang
Yiming Zeng · Li Li
Xiangdong Wang *Editors*

Single-cell Sequencing and Methylation

Methods and Clinical Applications

Springer

Advances in Experimental Medicine and Biology

Volume 1255

Advances in Experimental Medicine and Biology provides a platform for scientific contributions in the main disciplines of the biomedicine and the life sciences. This series publishes thematic volumes on contemporary research in the areas of microbiology, immunology, neurosciences, biochemistry, biomedical engineering, genetics, physiology, and cancer research. Covering emerging topics and techniques in basic and clinical science, it brings together clinicians and researchers from various fields.

Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 40 years, and is indexed in SCOPUS, Medline (PubMed), Journal Citation Reports/Science Edition, Science Citation Index Expanded (SciSearch, Web of Science), EMBASE, BIOSIS, Reaxys, EMBiology, the Chemical Abstracts Service (CAS), and Pathway Studio.

2019 Impact Factor: 2.450 5 Year Impact Factor: 2.324

More information about this series at http://www.springer.com/series/5584

Buwei Yu • Jiaqiang Zhang •
Yiming Zeng • Li Li • Xiangdong Wang
Editors

Single-cell Sequencing and Methylation

Methods and Clinical Applications

Springer

Editors
Buwei Yu
Ruijin Hospital
Shanghai Jiao Tong University
Shanghai, Shanghai, China

Jiaqiang Zhang
Department of Anesthesiology and
Perioperative Medicine, Center of
Clinical Single-Cell Biomedicine, Henan
Provincial People's Hospital, Zhengzhou
University People's Hospital, Henan
University People's Hospital
Zhengzhou, P.R., China

Yiming Zeng
2nd Affiliated Hospital
Fujian Medical University
Quanzhou, Fujian, China

Li Li
Centre for Single Cell Biomedicine
Henan Provincial People's Hospital
Zhengzhou, Henan, China

Xiangdong Wang
Zhongshan Hospital
Fudan University
Shanghai, Shanghai, China

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-981-15-4493-4 ISBN 978-981-15-4494-1 (eBook)
https://doi.org/10.1007/978-981-15-4494-1

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

About the Editors

Buwei Yu is a Professor at the Department of Anesthesiology, Shanghai Ruijin Hospital, Shanghai Jiaotong University School of Medicine, and President of the Chinese Medical Doctor Association and the Society of Anaesthesiologists.

Jiaqiang Zhang is a Professor and Director of the Department of Anaesthesiology and Perioperative Medicine, Henan Provincial People's Hospital, and a member of the Chinese Medical Association's Anaesthesia Branch, and the National Committee of the Anaesthesia Branch of the Chinese Medical Association. He is also Vice Chairman of the Anaesthesia Branch of the Henan Medical Association.

His main research focuses include investigating the mechanisms of postoperative cognitive dysfunction, clinical bioinformatics, anesthesia and neurodevelopment, and the physiology of pain. He is the author of more than 150 scientific publications.

Yiming Zeng is a Professor of Respiratory Medicine, Chairman of the Academic Committee of the Second Affiliated Hospital of Fujian Medical University, State Council Expert for Special Allowance, and Director of Fujian Province's Sleep Medicine Key Laboratory. He was named a middle-aged expert with outstanding contributions from the National Health and Family Planning Commission of the People's Republic of China (NHFPC). His main research focuses include clinical interventional pulmonology; sleep-breathing disorders; and non-invasive mechanical ventilation. He is the author of more than 150 scientific publications.

Li Li is the Director of Department of Science Research and Discipline Construction, and Principle Investigator of the Clinical Centre of Single-Cell Biomedicine, Henan Provincial People's Hospital. She is a member of clinical research group of the Chinese Medical Association's Scientific Research Management Branch, and of the Standing Committee of the Chinese Medical Association's Henan Research and Management Branch. Further, she is Vice-chairman of Henan Discipline Management Branch of the Chinese Hospital Management Society. She has been engaged in the management of medical scientific research for 30 years, with a focus on big data, laboratory biosafety and medical ethics. She published more than 20 scientific papers.

Xiangdong Wang is a Distinguished Professor of Medicine, Director of Shanghai Institute of Clinical Bioinformatics, Executive Director of Clinical Science Institute of Fudan University Zhongshan Hospital, Director of Fudan University Center of Clinical Bioinformatics, and Deputy Director of Shanghai Respiratory Research Institute. He is also a Visiting Professor at King's College London. His main research focuses on clinical bioinformatics, disease-specific biomarkers, chronic lung diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 300 scientific publications.

Understanding Diseases from Single-Cell Sequencing and Methylation

1

Buwei Yu, Li Li, Jiaqiang Zhang, Xiangdong Wang, and Yiming Zeng

Abstract

Clinical single-cell biomedicine has become a new emerging discipline, which integrates single-cell RNA and DNA sequencing, proteomics, and functions with clinical phenomes, therapeutic responses, and prognosis. It is of great value to discover disease-, phenome-, and therapy-specific diagnostic biomarkers and therapeutic targets on the basis of the principle of clinical single-cell biomedicine. This book reviews the roles of single-cell sequencing and methylation in diseases and explores disease-specific alterations of single-cell sequencing and methylation, especially focusing on potential applications of methodologies on human single-cell sequencing and methylation, on potential correlations between those changes with pulmonary diseases, and on potential roles of signaling pathways that cause heterogeneous cellular responses during treatment. This book also emphasizes the importance of methodologies in clinical practice and application, the potential of perspectives, challenges and solutions, and the significance of single-cell preparation standardization. Alterations of DNA and RNA methylation, demethylation in lung diseases, and a deep knowledge about the regulation and function of target gene methylation for diagnosing and treating diseases at the early stage are also provided. Importantly, this book aims to apply the measurement of single-cell sequencing and methylation for clinical diagnosis and treatment and to understand clinical values of those parameters and to headline and foresee the potential values of the application of single-cell sequencing in non-cancer diseases.

B. Yu
Ruijin Hospital, Shanghai Jiaotong University, Shanghai, Shanghai, China

L. Li
Department of Scientific Research and Discipline Construction, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China
e-mail: lili@henu.edu.cn

J. Zhang
Department of Anesthesiology and Perioperative Medicine, Center of Clinical Single-Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, P.R., China
e-mail: jqzhang@henu.edu.cn

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China
e-mail: xdwang@fuccb.com

Y. Zeng (✉)
2nd Affiliated Hospital, Fujian Medical University, Quanzhou, Fujian Province, China

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_1

Keywords

Clinical single-cell biomedicine · Single-cell sequencing and methylation · Pulmonary diseases · Methodologies · Clinical diagnosis and treatment

With the rapid development of single-cell biology and sequencing, clinical single-cell biomedicine is defined a new merging discipline to integrate single-cell RNA and DNA sequencing, proteomics, and functions with clinical phenomes, responses to therapies, and prognosis. Several hospitals start a new independent practice to perform clinical single-cell biomedicine, although there are still many challenges to be faced and solved. Clinical single-cell biomedicine is more expected to dynamically monitor cell–cell variations and communications, drug efficacy and resistances, discovery and development of therapeutic targets, and genealogic phenotypes of cells during disease progression [1–3]. Clinical single-cell biomedicine will analyze inter- and intra-cellular heterogeneity, new cell category, dysfunctional regulatory networks, microbes, and disease evolution. In addition to understanding molecular mechanisms using single-cell sequencing and measurements, it is more important to discover disease-, phenome-, and therapy-specific diagnostic biomarkers and therapeutic targets on the basis of the principle of clinical single-cell biomedicine. As the part of clinical single-cell biomedicine, we demonstrated important roles of single-cell sequencing in systems immunology in our previous book entitled "Single Cell Sequencing and Systems Immunology" [4], e.g., as a tool to deeply understanding the development and regulation of systems immunology. In this book, we furthermore overviewed the roles of single-cell sequencing and methylation in diseases and explored disease-specific alterations of single-cell sequencing and methylation. This book specially focuses on potential applications of methodologies on human single-cell sequencing and methylation, on potential correlations between those changes with pulmonary diseases, e.g., lung cancer, chronic lung diseases, and allergic lung diseases, and on potential roles of signaling pathways that cause heterogeneous cellular responses during treatment.

The first part of the book emphasizes the importance of methodologies in clinical practice and application, the potential of perspectives, challenges and solutions, and the significance of single-cell preparation standardization. Pensold and Zimmer-Bensch [5] headlined the importance of accurate and reliable cell capturing in single-cell sequencing, overviewed the current state of single-cell isolation methods, and addressed key parameters like sample compatibility, viability, purity, throughput, and isolation efficiency. Gupta et al. [6] systematically described the value of single-cell sequencing in the investigation of T cell receptors and their transcriptional profiles and firstly prospected the importance of the technological development in translational and clinical application. This is an example to apply the single-cell sequencing for special target clusters in a special cell population and illustrate the translational strategy how the single-cell sequencing is developed for clinical application. The single-cell sequencing of T cell receptors has the great value to benefit immune-therapy for cancer and autoimmune diseases.

The methylation and demethylation of cytosine in promoter regions play an important role in the control and regulation of gene expression by the modulation of translation by modifying tRNA-bases or silencing. The process of the methylation within cells can be influenced by their environment or for the development of complex organisms, especially for organs/tissues which are exposed and connected directly to the environment, e.g., lung. This book discusses alterations of DNA and RNA methylation and demethylation in lung diseases and provides the deep knowledge about the regulation and function of target gene methylation for diagnosing and treating diseases at the early stage. Zhou et al. [7] demonstrated global methylation pattern and specific gene methylation status of associated genes in the development of pulmonary fibrosis and methylation patterns and severities of the promoter regions of Thy-1, COX-2, p14ARF, and PTGER2 genes as disease-specific biomarkers to

predict the occurrence and development of the disease. Using bioinformatics, Liu et al. [8] addressed that altered methylations of inflammatory cells downregulated the gene expression of inflammatory mediators and initiated the occurrence of lung diseases. The combination of expression quantitative trait loci and genome-wide association studies was suggested as a new strategy to identify alterations of target gene methylation in chronic lung diseases, e.g., lung fibrosis [9] or chronic obstructive pulmonary diseases [10]. Of many target genes, DNA methylation of RAS-association domain family 1 was proposed as a lung cancer biomarker for new therapeutic strategies and for monitoring the reliability and sensitivity of DNA methylation [11].

One of the important issues in this book is to apply the measurement of single-cell sequencing and methylation for clinical diagnosis and treatment and to understand clinical values of those parameters. Wu et al. clearly reported the urgent need to optimize and standardize the workflow and protocol as well as standard operation performance, the comprehensive single-cell database and knowledgebase, and the design of clinical studies among various hospitals during clinical application [12]. The importance of target gene methylation and expression phenomes, e.g., Aplasia Ras homologue member I [13], P16 gene [14], and related molecular mechanisms of tumorigenesis and progression in various types of cancers, is obvious. Of those, single-cell RNA sequencing can be utilized to identify subtypes of pancreatic cancer [15] and genitourinary malignancies [16] and to improve the quality, efficiency, and specificity of cancer diagnostics [17]. In addition, new therapeutic targets and strategies can be discovered and developed with the improvement of methodologies and knowledge on single-cell sequencing and methylation. Duncan et al. offered an example of PI3K inhibitors and a frontline view of biological effects of the PI3K pathway and multiple isoforms of PI3K, mutations found in the PI3K isoforms in many different types of cancer, and new strategy of combination therapies between PI3K inhibitors and other target-driven therapies [18].

One of advances in this book is to headline and foresee the potential values of the application of single-cell sequencing in non-cancer diseases, which will be the frontline science and need more efforts to be explored. Garcia et al. provided the comprehensive understanding of single-cell RNA sequencing in human renal, pancreatic, and viral diseases [19]. This is an important and expecting review to discuss the specific application of single-cell sequencing in cellular compositions, heterogeneity and uncovering clues of viral infections and diseases of the kidney and pancreas for the development of targeted and personalized therapies. Singh specially emphasized the importance of single-cell sequencing in the discovery of the drug resistance clone, intercellular variation and communication, mutations and transcriptional profiles of a pathogen across different stages of human genital infections [20]. Rajan and Dall'Acqua addressed the potentials of those advanced technologies in the discovery and development of antibody-based humanized therapies [21]. Single B cell sequencing will provide a new approach and emerging strategy for antibody-based therapy. Chang et al. summarized the potential application and values of single-cell sequencing in the development of neurological cells and microglia as well as single-cell changes during brain injury [22].

This book is one of initiatives to deeply understand the importance and value of single-cell sequencing and methylation measurement for clinical application, although there are still many challenges and obstacles to be broken through. It is also highly expected to translate the simultaneous measurement of both single-cell sequencing and methylation in a human cell, e.g., parallel single-cell genome-wide methylome and transcriptome sequencing. There is a rapid growth in the development and improvement of single-cell methylation and sequencing, e.g., single-cell bisulfite sequencing for genome-wide base-resolution mapping of single-cell DNA methylation, random displacement amplification sequencing for the first full-length single-cell RNA-sequencing method, single-cell and single-base resolution DNA methylation analysis based on reduced-representation bisulfite sequencing,

and single-cell, locus-specific bisulfite sequencing for cell-to-cell variability and the pathogenic history. Complete DNA CpG methylomes at the single cell can be screened and compared comprehensively through whole genome bisulfite sequencing, reduced-representation bisulfite sequencing, and enrichment-based methods such as MeDIP-seq, MBD-seq, and MRE-seq. At the end, we as co-editors of this special book would like to take this special opportunity to deeply appreciate all authors and contributors for the intensive and hard works to make this book possible for publication. We are especially grateful for those experts to review and comment chapters in order to maintain the high quality and look forward to working with all of you in future.

References

1. Wang W, Gao D, Wang X (2018) Can single-cell RNA sequencing crack the mystery of cells? Cell Biol Toxicol 34(1):1–6. https://doi.org/10.1007/s10565-017-9404-y
2. Zeng Y, Chen X, Gao H, Wang X (2018) An artificial intelligent single cell is part of the cell dream world. Cell Biol Toxicol 34(4):247–249. https://doi.org/10.1007/s10565-018-9433-1
3. Busch S, Talamini M, Brenner S, Abdulazim A, Hänggi D, Neumaier M et al (2019) Circulating monocytes and tumor-associated macrophages express recombined immunoglobulins in glioblastoma patients. Clin Transl Med 8(1):18. https://doi.org/10.1186/s40169-019-0235-8
4. Wang XD (2015) Single cell sequencing and systems immunology. In: Wang X (ed) Translational bioinformatics (book series), vol 7. Springer, Singapore. ISBN: 978-94-017-9752-8 (Print) 978-94-017-9753-5 (Online)
5. Pensold D, Zimmer-Bensch G (2020) Methods for single-cell isolation and preparation. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
6. Gupta S, Witas R, Voigt A, Semenova T, Nguyen CQ (2020) Single-cell sequencing of T cell receptors: a perspective on the technological development and translational application. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
7. Zhou S, Wang X, Gao H, Zeng Y (2020) DNA methylation in pulmonary fibrosis. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
8. Liu Y, Gao H, Wang X, Zeng Y (2020) Methylation of inflammatory cells in lung diseases. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
9. Luo Q-K, Zhang H, Li L (2020) Research advances on DNA methylation in idiopathic pulmonary fibrosis. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
10. Chen X, Yan F, Lin X, Shi L, Wang X, Zeng Y (2020) DNA methylation in chronic obstructive pulmonary disease. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
11. Zhang T, Li Y, Zhang H, Wang X, Liu X, Li L (2020) The role of RASSF1 methylation in lung carcinoma. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
12. Wu F, Fan J, Fang J, Dalvi PS, Odenthal M, Fang N (2020) Single cell sequencing: a new dimension in cancer diagnosis and treatment. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
13. Liu X, Zhang T, Li Y, Zhang Y, Zhang H, Wang X, Li L (2020) The role of methylation in the CpG island of the ARHI promoter region in cancers. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
14. Li Y, Zhang T, Zhang H, Huang Q, Wang X, Liu X, Li L (2020) Clinical significance of P16 gene methylation in lung cancer. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
15. Luo Q, Fu Q, Zhang X, Zhang H, Qin T (2020) Application of single-cell RNA sequencing in pancreatic cancer and the endocrine pancreas. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
16. Murphy N, Shah P, Shih A, Khalili H, Liew A, Zhu X, Lee A (2020) Chapter 13: Single cell sequencing in GU malignancies. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
17. Probst V, Bagger FO (2020) Single cell sequencing in cancer diagnostics. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
18. Duncan L, Shay C, Teng Y (2020) PI3K Isoform-selective inhibitors in cancer. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore

19. Garcia S, Der E, Putterman C (2020) Single cell RNA sequencing in human disease: renal, pancreatic, and viral diseases. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
20. Singh R (2020) Single-cell sequencing in human genital infections. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
21. Rajan S, Dall'Acqua WF (2020) Emerging strategies for therapeutic antibody discovery from human B cells. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore
22. Chang E, Ruan X, Zhu R, Wang Y, Zhang J (2020) Single-cell mRNA sequencing technology and developing cerebral cortex. In: Single-cell sequencing and methylation: methods and clinical applications, Advances in Experimental Medicine and Biology. Springer, Singapore

Buwei Yu is a professor in the Department of Anesthesiology at Shanghai Ruijin Hospital affiliated to Shanghai Jiaotong University School of Medicine and is President of the Chinese Medical Doctor Association and the Society of Anesthesiologists.

Li Li is director of the Department of Science Research and Discipline Construction and Principal Investigator of Clinical Centre of Single-Cell Biomedicine, Henan Provincial People's Hospital. She is a member of clinical research group of Chinese Medical Association's Scientific Research Management Branch, Standing Committee member of Chinese Medical Association's Henan Research and Management Branch, and vice-chairman of Henan Discipline Management Branch of Chinese Hospital Management Society. She has engaged in the management of medical scientific research for 30 years, and her main research is focused on health management scientific research big data, laboratory biosafety, and medical ethics. She has published more than 20 scientific papers.

Jiaqiang Zhang is professor and director of the Department of Anesthesiology and Perioperative Medicine at Henan Provincial People's Hospital and is a member of Chinese Medical Association Anesthesia Branch and the National Committee of the Anesthesia Branch of the Chinese Medical Association. He is vice-chairman of the Anesthesia Branch of the Henan Medical Association. His main research is focused on investigating the mechanism of postoperative cognitive dysfunction, clinical bioinformatics, anesthesia and neurodevelopment, and physiology of pain. He is the author of more than 150 scientific publications.

Xiangdong Wang is a distinguished professor of medicine, director of Shanghai Institute of Clinical Bioinformatics, executive director of Clinical Science Institute of Fudan University Zhongshan Hospital, director of Fudan University Center of Clinical Bioinformatics, deputy director of Shanghai Respiratory Research Institute, and visiting professor of King's College London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 300 scientific publications with the impact factor about 900, citation number about 6920, h-index 48, i10-index 221, and cited journal impact factor about 8000.

Yiming Zeng is a professor of Respiratory Medicine, chairman of Academic Committee of the Second Affiliated Hospital of Fujian Medical University, State Council Expert for Special Allowance, and director of Sleep Medicine Key Laboratory of Fujian Province. He achieved the Outstanding Contribution of Middle-aged Expert of National Health and Family Planning Commission of the People's Republic of China (NHFPC). His main research is focused on clinical, interventional pulmonology, sleep-breathing disorders, and noninvasive mechanical ventilation. He is the author of more than 150 scientific publications.

Methods for Single-Cell Isolation and Preparation

2

Daniel Pensold and Geraldine Zimmer-Bensch

Abstract

Within the last decade, single-cell analysis has revolutionized our understanding of cellular processes and heterogeneity across all disciplines of life science. As the transcriptome, genome, or epigenome of individual cells can nowadays be analyzed at low cost and in high-throughput within a few days by modern techniques, tremendous improvements in disease diagnosis on the one hand and the investigation of disease-relevant mechanisms on the other were achieved so far. This relies on the parallel development of reliable cell capturing and single-cell sequencing approaches that have paved the way for comprehensive single-cell studies. Apart from single-cell isolation methods in high-throughput, a variety of methods with distinct specializations were developed, allowing for correlation of transcriptomics with cellular parameters like electrophysiology or morphology.

For all single-cell-based approaches, accurate and reliable isolation with proper quality controls is prerequisite, whereby different options exist dependent on sample type and tissue properties. Careful consideration of an appropriate method is required to avoid incorrect or biased data that may lead to misinterpretations.

In this chapter, we will provide a broad overview of the current state of the art in matters of single-cell isolation methods mostly applied for sequencing-based downstream analysis, and their respective advantages and drawbacks. Distinct technologies will be discussed in detail addressing key parameters like sample compatibility, viability, purity, throughput, and isolation efficiency.

Keywords

Single-cell isolation · Tissue dissociation · Dilution · Micromanipulation · Laser microdissection · FACS · Microfluidic

In recent years, rapid technological development and improvements have been achieved in the field of high-throughput sequencing enabling diverse applications due to significant drop of costs. Together with methodological advances especially in the field of single-cell isolation, this has paved the way for reliable single-cell-based analysis in high-throughput. Still, tissue dissociation, single-cell separation and isolation represent arguably the greatest source of technical variation, contamination, and batch effects in any single-cell study [1], and hence represent the key determinants for a successful experimental design. There are diverse approaches described to isolate material for single-cell omics with

D. Pensold · G. Zimmer-Bensch (✉)
Division of Functional Epigenetics, Institute of Zoology (Biology II), RWTH Aachen University, Aachen, Germany

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_2

different advantages and limitations for protein, RNA, or DNA analysis [2–4]. In this chapter, we will mainly discuss methods aiming at RNA and DNA isolation for sequencing-based analysis.

According to a survey about the German market carried out in 2014 [2], the most frequently used approaches for single-cell isolation were fluorescence-activated cell sorting (FACS, 33%), manual micromanipulation (17%), laser microdissection (17%), random seeding or serial limiting dilution (15%), and microfluidics/lab-on-a-chip methods (12%). Other technologies including optical tweezers, dielectrophoresis, or non-contact depending methods were less frequently reported (in total 6%). The requirements for technologies to separate and isolate single cells from specimens of different nature are as diverse as the purpose for which the cells are used in downstream processing and analysis. So far, no available method suits all demands. When only considering the isolation process in matters of efficiency and cell viability already numerous factors including cell type, sample preparation, device calibration, sorting mode, and substrate are of high relevance, in addition to other factors that are rather hard to quantify (e.g., operator skills, Fig. 2.1). The following paragraph seeks to cover the most general requirements for many of the approaches discussed in more detail in the following chapters, helping to choose the method of choice.

2.1 Parameters to Be Considered

Typically, **specimen type** and **origin** define to a great extent which technology is best applicable for sample preparation. Most methods described here require cells in suspension for separation and isolation procedures. Thus, all samples easily dissociable, or liquid materials like immunological organs such as peripheral blood, spleen, or lymph nodes called liquid biopsies, are straightforward to handle often requiring only concentration of the material. However, commonly samples originate from solid and complex tissues requiring chemical or enzymatic as well as mechanical dissociation of the cells.

Moreover, **downstream analysis** largely dictates the method of choice for sample preparation. Independent of the **targeted information level** ranging from genomics, epigenomics, transcriptomics to proteomics, **cell integrity and viability** has to be ensured, to avoid early degradation of DNA, RNA, or proteins, respectively. Stress factors like mechanical forces, radiation, chemical changes in the cellular environment, etc. may alter the intrinsic cellular states, e.g., by inducing differentiation or apoptosis [5–7]. Further relevant considerations refer to the **purity** of the isolated single cells to avoid potential **contamination** with cell fragments, free DNA or RNA molecules. For some experiments, cultivation after cell separation is intended, requesting sterile operational conditions, which can be achieved in some approaches by using disposable components (e.g., microfluidic chips).

Throughput in terms of single cells isolated per second and the **total number of single cells** are further crucial parameters. Low-throughput applications in the range of one to several hundreds of cells are usually performed manually or half-automated using systems like Fluidigm C1 (South San Francisco, CA, USA). Thereby, most manual approaches such as micromanipulation, optical or acoustic tweezers work with high precision, care, or high efficiency. Low-throughput methods apply to samples evident in low cell numbers, which require small sample volumes and low dead volumes as critical parameters.

High-throughput approaches enable capturing of cells to a range of several thousands. Such approaches may also be applied to detect underrepresented cell types like CTCs (circulating tumor cells) as a rare cell population within billions of different blood cells [8]. High-throughput methods like FACS often rely on strong sample dilution with the final sample volume being usually large, which impacts downstream processing. Another key issue is the **separation yield, cell-capture** or **isolation efficiency** in regard to the total input. Especially for low-abundant samples or rare biopsies, a maximum in capturing rate is essential [2, 9, 10]. Actually, most of the common high-throughput technologies such as FACS or droplet generation

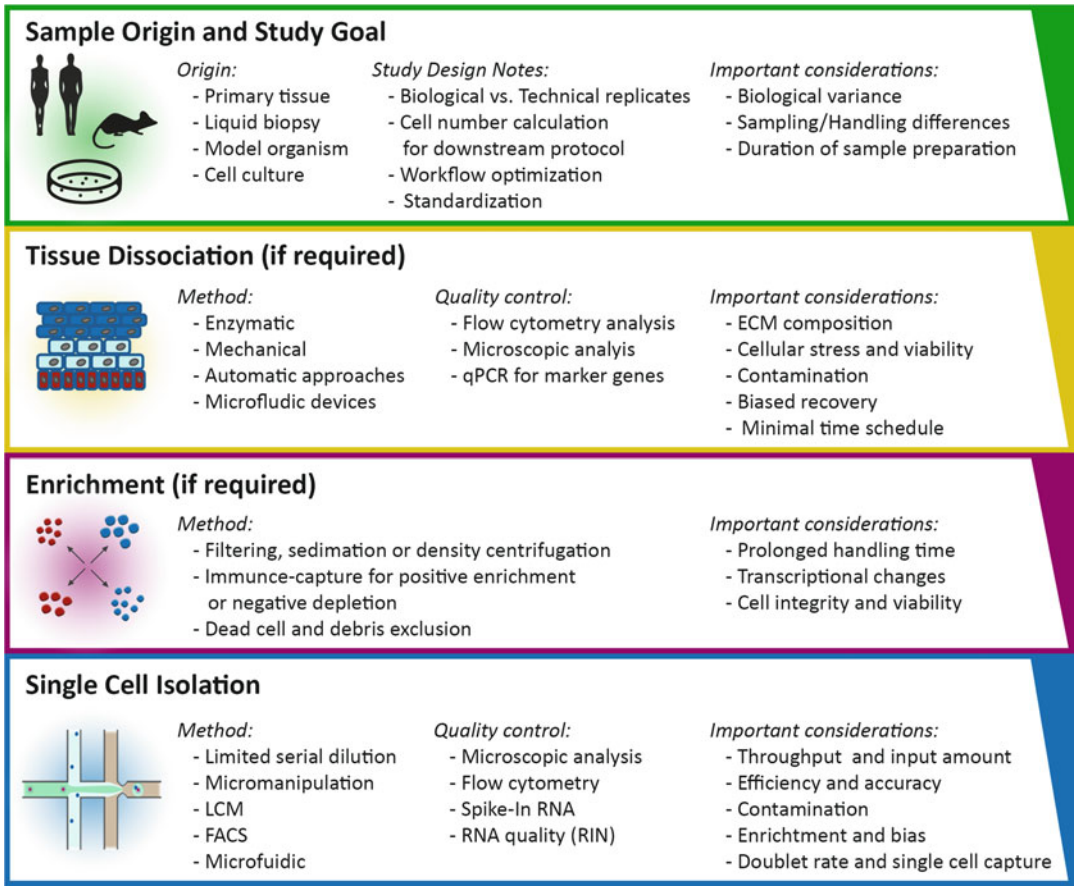


Fig. 2.1 Overview of steps associated with the isolation process. *ECM* extracellular matrix, *FACS* fluorescent-activated cell sorting, *LCM* laser capture microdissection, *RIN* RNA integrity number

have their strongest limitation in terms of precision, cell-capture efficiency, and vitality of cells after isolation.

The **economical aspect** including all costs from sample acquisition, dissociation, separation, isolation to amplification and library preparation for next-generation sequencing [11] represents another issue influencing the method of choice. Of course, analysis of low numbers of particularly selected cells will be in total cheaper than analyzing thousands of randomly chosen ones. However, costs per cell decline with throughput, which significantly dropped within the last years and which is expected to progress. The cost per cell nowadays varies depending on the approach and facility between less than $0.30 for some customized microfluidic systems up to $1 to $2 for early-indexing plate-based 3′-RNA sequencing methods. Manual methods and late-indexing full-length transcriptome profiling often require separate sample processing, consequently increasing costs (starting about $8–12 per cell) and time.

An overview of the different parameters that determine the choice of a particular approach is summarized in Table 2.1. All these and other parameters force operators to choose a certain compromise for the experimental design to match the goal of each individual study. As follows, we will discuss specific features, applications, and limitations of a broad spectrum of single-cell isolation methods.

2.2 Cell Dissociation and Enrichment

First, the protocol for dissociation and isolation needs to be carefully optimized to ensure reliable and valid data, aiming to keep original conditions to the greatest possible extent (Fig. 2.1). Tissues vary significantly in extracellular matrix (ECM) composition, cellularity, and stiffness, which influences the requirements for dissociation protocols. Dissociation is often achieved by enzymatic digestion and/or mechanical dissociation, which both can lead to activation of stress-related genes [5]. General prerequisites for a successful and mild dissociation with minimized procedure-induced intrinsic changes include a low time scale, working on ice whenever possible, keeping pipetting and centrifugation to a minimum, as well as using calcium and magnesium-free but bovine serum albumin-containing buffers and media. Cell lysis-induced extracellular DNA can be diminished by DNase I application during cell separation.

Enzymatic digestion is commonly conducted by trypsin, collagenase, protease, accutase, elastase, or dispase treatment as well as by commercial enzymatic mixtures such as TrypLE Express and Liberase Blendzyme 3, depending on the composition of the extracellular matrix and cell–cell connections in the different tissues. The application of cold active enzymes like proteases from *Bacillus licheniformis* thereby helps to reduce heat stress caused by enzymatic digestion at 37 °C [12]. The careful adjustment of enzymatic incubation duration for each sample type is crucial, as extended incubation times may induce cell lysis of fragile cell types. Short incubation in turn, bears the risk of incomplete cell separation and the exclusion of large cell clusters during subsequent filter steps, potentially leading to bias in cell composition [13]. Of note, all enzymatic treatments may affect the transcriptome of single cells through cellular uptake or by altering intercellular communications [14, 15].

For mechanical dissociation fire-polished glass capillaries with cell size and tissue input-adjusted tip sizes are usually applied. The frequency of pipetting represents another critical parameter strongly impacting cell viability and dissociation effectiveness. The avoidance of air bubbles during all pipetting steps is essential as oxidative stress is one of the major sources for cell death. The extracellular matrix and also damaged or dead cells are the largest source for contamination, e.g., by inducing free RNA and DNA, which affect subsequent downstream processing, especially for sequencing approaches. Including a density gradient centrifugation may help to reduce debris as well as free RNA and DNA molecules but may simultaneously result in biased recovery of the remaining cell populations. Another way to remove clumps and debris from suspensions is to filter with appropriately sized cell strainers. However, the produced suspension should be processed as soon as possible, ideally within 30 min after dissociation, to avoid re-aggregation of cells or induction of transcriptional changes. Otherwise, loss of cellular functions, cell–cell interactions, and the tissue architecture may lead to transcriptional adaptation [16].

Recently, innovative microfluidic cell dissociation devices have been fabricated that may drastically change the way solid tissue samples are processed into single cells, allowing for automated processes [17]. Such new technologies avoid inter-assay variation occurring from differences in handling of the tissue. Included microfluidic structures have been optimized for straightforward tissue digestion, cell dissociation, filtering, and polishing, mainly by passing the tissue sequentially through progressively smaller size scales. Tissue-specific kits can be designed to integrate differences in ECM and interconnectivity of cells, improving reproducibility and efficiency of single-cell preparations [18, 19].

Even if performed automatically, tissue dissociation accounts as a major source of variation in single-cell analysis, for which applicable methods using preserved samples [20–22] or nuclear RNA/DNA [23–26] represent an attractive alternative. Nuclear RNA enrichment based on cellular membrane disruption [23–26] has been shown to be sufficient to capture cell type information [26]. However, the overall resolution per cell is

Table 2.1 Classification and critical parameters of single-cell isolation principles

Parameter	Limiting dilution	Micromanipulation			LCM	FACS	Microfluidic			Dielectrophoresis	Optical and acoustic tweezers
		Manual total cell	Patch-clamp harvesting	Automated (partially microfluidic)			Microwells	Hydrodynamic	Droplet		
Sample type	Living	Living	Living	Living	Living or fixed	Living or fixed (e.g., INTACT)	Living	Living	Living	Living	Living
Sample requirements	D or LB	D or LB	Tissue, D or LB	D or LB, homogeneous or heterogeneous	Tissue	D or LB, heterogeneous	D or LB, homogeneous	D or LB, homogeneous or heterogeneous	D or LB, homogeneous	D or LB, homogeneous or heterogeneous	D or LB, homogeneous
Input amount	Large	Small	Small	Moderate	Small	Large	Large	Large	Large	Moderate	Moderate
Cell stress	Gentle	Gentle	Gentle to moderate	Gentle to moderate	Often impaired	Often impaired	Moderate	Diverse	Moderate	Moderate	Moderate
Enrichment/ bias	No	User selected	User selected	Diverse	User selected	Yes, user selected	Cell size and shape	Cell size and shape	Cell size and shape	Yes, user selected	Yes, user selected
Individual cell control	No	Yes	Yes	Yes	Yes	Yes	No	Yes/No	Typically not	Typically not	Yes
Throughput	Moderate	Very low	Very low	Low to high	Low	High	Low to high	Moderate	Very high	Low to moderate	Low to moderate
Precision, Accuracy	NA (Poisson)	High	High	High	High	High	Yes/No	Yes/No	NA (Poisson)	Moderate to high	High
Doublet Rate		Low	NA	Low	Low	2–3%	3–30%		4–12%		
Singlet capture rate		NA	NA	High	NA	>80%	2–40%		Up to 50%		
Separation yield, cell-capture or isolation efficiency	Low	High	High	Moderate to high	High	Low	Low to moderate	Moderate	Low to moderate	Moderate	
Purity/ Contamination risk	High false positive rate	Mechanical shearing damage	Diverse, require high user skills	Potential mechanical shearing damage	Co-isolation of surrounding material	Potential cell damage	Cross contamination by free RNA/DNA	Cross contamination by free RNA/DNA	Cross contamination by free RNA/DNA		
Automation	Yes and No	No	No	Yes	No	Yes	Yes	Yes and No	Yes	Yes	Yes and No
Commercial				CellenONE, WOLF CellSorter, CellSelector, Single-Cell Printer	Diverse	Diverse	C1, MicrowellChip, SeqWell, Nanogrid, BD Rhapsody		Chromium, Dolomite Nadia, ddSeq	DEPArray	
References	[2]	[45, 47, 49]	[45, 63]	[2, 45, 97]	[75]	[102]	[2, 97, 117]	[2, 97]	[2, 27, 116, 117]	[120]	[138]

D tissue dissociation, *FACS* fluorescent-activated cell sorting, *LCM* laser capture microdissection, *LB* liquid biopsy

reduced and the RNA is biased towards the nuclear fraction. This approach has been extensively applied for epigenetic and transcriptomic profiling of differentiated neurons, as adult neurons are difficult to dissociate reliably due to their axonal and dendritic processes, which are damaged or destroyed during the dissociation process [27, 28]. A major advantage of single nucleus RNA sequencing is that it can be applied to frozen and fixed tissue [20, 21, 29–32], enabling simultaneous processing of samples acquired at different time points, thereby minimizing technically induced batch effects.

For many applications, the enrichment of particular cell types after dissociation is beneficial or even essential for downstream analysis (Fig. 2.1). Sedimentation or density centrifugation during the dissociation process separates cells based on cell size and density. Further, mechanical filtering, hydrodynamic, electrokinetic or acoustophoresis approaches can be applied [33, 34]. Very common are immune-capture methods like magnetic bead-based immune selection (MACS) or immunofluorescence using FACS, facilitating rare cell populations to be analyzed at single-cell level without the necessity to sequence several thousands of cells. This was for example applied to enrich and analyze hematopoietic stem cells from red bone marrow [35, 36]. Still, an a priori immune selection of living cells for FACS or MACS is commonly based on cell surface receptor labeling that may induce intracellular signaling upon antibody or ligand binding [37, 38]. Most of these methods are capable of positive enrichment of cell types of interest or negative depletion of unwanted cells. Apoptosis markers and fluorescent DNA intercalating agents are applied to label and sort out damaged cells. A specific enrichment can also be achieved by induced expression of fluorescent reporter genes, which enables the identification of a particular cell type without immune-labeling. Reporter gene expression can either be driven by lineage-specific promoters, or engineered to be co-expressed with any protein of interest. A more advanced approach identifies individual cells based on the microanatomical location. For that, fluorescent reporters being either photoactivatable [39–41] or photoconvertible [42–44] are used to precisely mark cells of interest by two-photon microscopy. By linking such reporters to cell type-specific expression, cells can be precisely identified based on expression markers and additionally based on specific microanatomical locations within a tissue prior to dissociation. Recently, this method was applied to perform NICHE-seq, systematically characterizing the cellular composition of the spleen among other immune niches [40]. However, such enrichment approaches come with the disadvantage of additional steps and time, which in turn influence the intrinsic cellular state and survival.

Due to the overall cost and time-intensive process of single-cell sequencing experiments, careful quality control and monitoring is required (Fig. 2.1). The performance of the different approaches can be assessed by the use of several readouts presented in the following paragraphs. Although most of the relevant parameters for quality control are related to sequencing results and will be discussed elsewhere, quality control can and needs to be included already during isolation to minimize the proportion of low-quality cells. Most importantly, attention should be paid to the avoidance of cell duplets or even multiple cells, as well as dead cells in the cell capture step, as they remain difficult to assess in downstream analysis. Imaging-based approaches are valuable tools to control for the successful isolation and viability of single cells. Flow cytometry like FACS is particularly useful to measure several critical metrics simultaneously, such as cell viability, and rates of doublets and small cell clusters. The introduction of artificial spike-in RNAs during the isolation process (External RNA Controls Consortium (ERCC) standards [45] or Sequin standards [46]) helps to calibrate measurements and account for technical variability during subsequent sequencing data analysis.

In summary, single-cell preparation requires numerous considerations and careful optimization, to choose the most appropriate protocol (Fig. 2.1). Enrichment of certain cell types is frequently required but also comes with

drawbacks that need to be considered. In the following paragraphs, we will discuss the limitations and advantages of particular single-cell isolation methods in more detail.

2.3 Dilution-Based Single-Cell Isolation

Laboratories and companies all over the world have used serial limited dilution for decades to isolate single cells (Table 2.1). Nowadays, this method experiences a renaissance being implemented in automated pipetting robots that tremendously reduce personnel workload. In principle, many technologies, even the microfluidic traps rely on the simple fact that due to statistical distribution of cells in a suspension, the number of cells in a highly diluted sample will be at some point so low that individual cells will be present in a targeted small volume (Fig. 2.2a).

Serial limited dilutions are indeed easy to carry out with standard laboratory inventory, and with the help of automation by pipetting robots it is a simple, reproducible and relative cost-efficient method. Therefore, many pharmaceutical companies still rely on fully automated serial dilution solutions. However, due to the lack of control, it is not possible to isolate specific cell types with this method per se, but it can be combined with upstream sorting or enrichment techniques. It further has to be considered that serial limited dilutions are prone to high false positive rates and to exclusion of cells of interest. Hence, it is widely applied to microbial samples, but less applicable for isolation of cells from complex or rare specimens [47].

2.4 Isolation by Micromanipulation

In the past, manual micromanipulation was one of the mostly applied isolation approaches, nowadays still being useful for particular applications. Albeit time consuming and labor intensive, this method provides clear advantages in matters of isolation precision (Table 2.1) [48–53]. Two of the first single-cell DNA or RNA isolation methods used glass capillaries to harvest either the whole cell, which assures complete isolation and minimizes loss of cell material [54–57], or the cytoplasm by patch clamping [58, 59].

A common setup for manual cell isolation typically consists of an inverted microscope equipped with micropipettes that are coupled to a micromanipulator, controlled by motorized mechanical stages (Fig. 2.2b). Piston systems coupled to glass capillaries enable reliable and precise respiration and dispensation [54, 56]. The tip opening diameter of the glass capillaries is adjusted by electrode pullers and sharp edges are polished by heat-induced melting to avoid mechanical shearing of cells. Silanization of the micropipette prevents sticking of cells, or RNA and DNA molecules to the glass capillary [49, 51].

Cells are usually provided in suspension under a microscope allowing the operator to select for specific target cells according to optical parameters like size, shape and granularity, reporter expression or cell surface labeling (Fig. 2.2b) [49]. Such targeted isolation of a particular cell under visual control represent a crucial advantage of this approach. The target cell is aspirated into the micropipette and transferred to a new reaction tube, being released by dispensation together with the aspirated liquid volume. However, the comparatively high volume of suspension needed for cell harvesting represents a major drawback [56], as it might contain contaminants of extracellular RNA and DNA resulting from cell damage [56]. Many protocols therefore suggest several washing steps prior to cell lysis to ensure contamination-free isolation [49–51].

Although being a very flexible approach in regard to cell types and substrates, the manual process of obtaining single cells by micromanipulation limits the overall throughput (3–8 cells/h) rising the risk of intrinsic changes like transcriptomic profiles caused by the extended handling time [2, 16, 60]. In addition, the low throughput limits the applicability to studies of cellular heterogeneity requiring numerous single cells. There are some commercial technologies

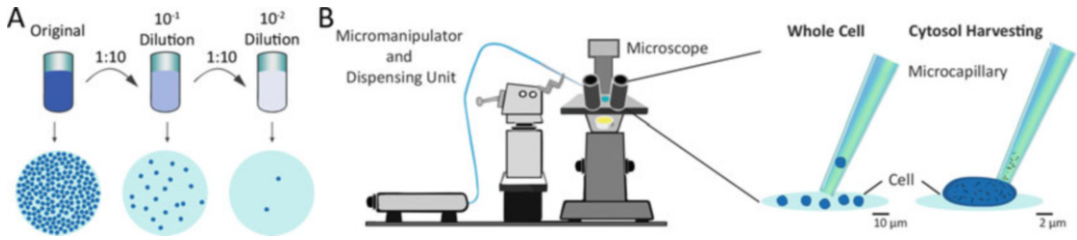


Fig. 2.2 Dilution-based approaches and micromanipulation: (**a**) Limiting serial dilution methods utilize statistical Poisson's distribution of highly diluted cells to separate and isolate individual cells. (**b**) Manual harvesting of cells or their content provides an economical and simple to use method for single-cell isolation. An experimental setup normally consists of a microscope for target identification and observation of the isolation process, a micromanipulator for three-dimensional movement of the microcapillary as well as a dispensing unit (manual or piezoelectric) connected to the microcapillary for controlled pressure driven isolation. In that way, either the whole cell out of suspension can be isolated or the cytosol of cells in culture or even in tissue enabling for instance correlation of patch-clamp electrophysiological recordings with single-cell transcriptomics

that automated the isolation to some degree and hence, increased the throughput. The ALS CellSelector is a freely configurable pipetting robot for automated transfer of single cells out of solution into a well-plate format (e.g., 96-well plate). Other automated micromanipulation systems allowing for higher throughput in single-cell isolation are based on acoustic microdispensing systems (CellenONE, Cellenion) or piezoelectric approaches (WOLF Cell Sorter, NanoCellect Biomedical or Single-Cell Printer, Cytena). The CellenONE system, integrating the microcapillary dispersion technology with constant imaging of the target cells, offers a new, innovative approach to isolate cells gently, rapidly and with high precision, applicable for downstream single-cell sequencing. Systems based on piezoelectric actuation utilize a hydrodynamic pressure pulse within a microchannel to manipulate single cells [61]. Commercialized systems like the WOLF Cell Sorter combine piezoelectric actuation with laser detection system comparable to FACS on a microfluidic scale [62]. With a response time of ~0.1 ms, more than 1000 cells can be sorted per second, being collected in 96- to 384-well plates. The Single-Cell Printer (Cytena) includes an automated image analysis system for detection and generates droplets containing a single cell using a microfluidic drop-on-demand dispenser chip [2]. The piezoelectric actuator raises the pressure inside the chip to generate a single droplet, thereby ejecting droplets through a nozzle encapsulating single cells [63], whereby unwanted droplets are sorted out enabled by image analysis. The system was shown to operate with >80% efficiency and 90% viability at optimal sample concentration of 6.2×10^5 cells/mL^2.

Alike whole cell harvesting, manual isolation of cellular material by patch-clamp approaches is time consuming and limited in throughput (Fig. 2.2b) [64–66]. It is frequently combined with electrophysiological recordings coming with the great advantage of known position and electrophysiological properties of cells. The tip of the glass capillary used for recording as well as isolation of the cell's material is much smaller and normally produces a tight sealing in a gigaohm range between the cell membrane and the tip. This ensures that even small currents across the membrane are captured while creating a barrier between the extracellular fluids and the cytoplasm. Upon finishing the electrophysiological recordings, a sub-atmospheric pressure is applied inside the glass capillary to rupture the cell membrane in the patch allowing for harvesting of the cytoplasm. The well-trained operator needs to collect as much cytoplasmic content as possible without destroying the cell. This approach comes with high sample-to-sample variation hampering quantitative analysis. Similar to whole cell harvesting, the micropipettes are often silanized to avoid loss of sample and contamination from surrounding cells and ECM [56].

For both described manual approaches, the monitoring of the successful transfer of the isolated material to the final reaction tubes is rather difficult, as this step requires leaving the focus plane with the glass capillary. Recent technologies improved on that by fully automated isolation and placement of single cells assisted by video systems and image processing algorithms [67].

Despite the mentioned limitations, applications of these methods are diverse, ranging from bacterial analysis [68, 69] to reproductive medicine [70], forensics [71], endosymbionts from termite gut analysis [72], and crenarchaeota from soil [73], and becoming especially attractive for samples with limited cell numbers or fragile cells types. Manual approaches can easily be combined with pre-enrichment methods or fluorescent labeling, like it was applied by Ramskold et al. [74] to isolate circulating tumor cells from the blood of a melanoma patient.

2.5 Laser Dissection

An alternative method of manual isolation is laser capture microdissection (LCM) or laser microbeam microdissection (LMM), an advanced technique to collect individual cells or cell compartments from usually solid tissue samples under visual control (Fig. 2.3a) [75–77]. This is especially applicable for samples or biopsies being less amenable to single-cell suspension dissociation [78]. Samples are typically provided as formalin-fixed paraffin-embedded or cryo-fixed tissues [79]. The Leica LMD7000 system with Live Cell Cutting (LCC) function even permits dissection of living cells from tissue sections [80–84].

LCM and LMM systems rely on an optical microscope coupled with a coaxial cutting laser and computer assisted control. The operator marks the targets to be cut off automatically with a cutting width of about 1 μm or less. Following laser-based sectioning, different technologies allow for precise harvesting (Fig. 2.3a; reviewed by Hodne and Weltzien [85]). The classical laser capture microdissection (LCM) exploits a contact-based extraction by employing an adhesive inert membrane to the section surface, which is melted locally after target excision by low energy infrared (IR) laser pulses [77]. Hence, samples are extracted via adhesion to the membrane or adhesive tube caps (Fig. 2.3a). For laser microbeam microdissection (LMM) [86], the laser-cut sample is either falling into a reaction tube following gravity, also called contact-free gravity-assisted microdissection (GAM), or it is catapulted against gravity. The first option requires inversely mounted substrates placed about a collection tube (Leica LMD7000, Fig. 2.3a). In contrast, contact-free laser pressure catapulting (LPC) utilizes a local plasma impulse induced beneath the cell by a short, defocused laser pulse to catapult excised samples (or compartment) into a nearby collector container (Zeiss PALM, Fig. 2.3a).

One of the main challenges in laser-assisted microdissection is to dissect only the cell or compartment of interest without contamination from neighboring cells or other unspecific fragments. Imprecise cutting or a poorly calibrated system can lead to both, false positive and false negative results. Although the newest generation of systems assist the operator with a high level of user-friendliness and automation, the selection process remains user-based and therefore strongly impacts throughput and reproducibility. Similar to the manual isolation with micromanipulation, it might remain unclear whether the cell was actually transferred and/or whether any contaminants were co-isolated, especially for contact-based cell extraction (adhesive methods) [77].

The integrity of the extracted material is important for reliable downstream analysis of biomolecules such as DNA, RNA, and proteins [87]. Depending on the applied fixation methods, or cryopreservation [88], as well as extraction methods (adhesive methods, GAM, LPC), single-cell integrity might be compromised [89]. In general, LCM/LMM based assays generate low yield of material, particularly for low abundance RNA species [89].

While most of the high-throughput methods relying on dissociated cells, LCM/LMM methods as well as cytosol harvesting with a patch-clamp

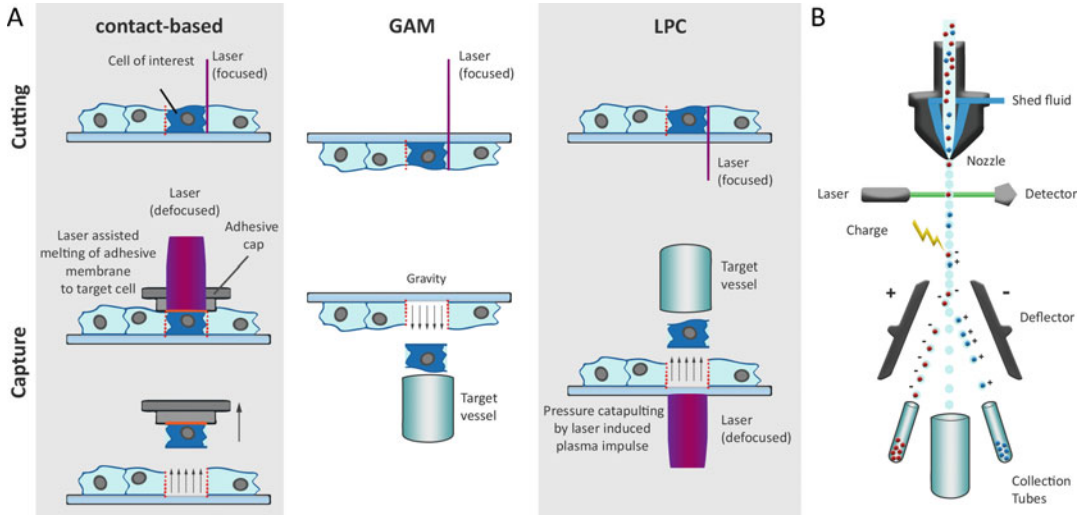


Fig. 2.3 Laser capture microdissection (LCM) and fluorescence-activated cell sorting (FACS) technology: (**a**) Schematic overview of the different technologies applied for LCM. Dissection is usually achieved after visual identification of the target cell using a focused laser. The methods mostly differ in their capture process. In contact-based methods, capturing is realized by melting an inert adhesive membrane to the target cell with a defocused laser pulse and by transferring the target to a collection tube. In gravity-assisted microdissection (GAM), the object slide is mounted inversely, and the target falls into a collection tube after dissection. Laser pressure catapulting (LPC) utilizes a defocused laser beneath the target to induce a plasma pulse hurling the sample into a collection tube. (**b**) FACS is applied to analyze suspended cells according to a variety of optical parameters ensuring single-cell separation as well as cell type specificity by fluorescent tags. Cells are lined up using a shed flow liquid. This stream is broken into individually flying droplets passing a laser-detector setup that monitors and analyzes the properties online. Droplets are then charged and separated in an electric field. The individual cells can be sorted in separated wells of well plates

pipette can be applied to intact tissue [75]. Moreover, cells are harvested under direct microscopic control providing additional information, like localization within tissue architecture, morphology, or immune-labeling.

In the past years, various single-cell analysis applications based on LCM/LMM isolated cells have been published: Single-cell RT-PCR [90], short tandem repeat analysis (STR) analysis in forensics [91], Western blot and mass spectrophotometry analysis [88]. Analysis of solid tissue sections without prior dissociation is of great interest when investigating heterogeneous specimens regarding their cellular structure as well as physiological and pathological processes [92]. For instance, in solid tumor research linking the molecular information of individual cells to their specific localization or context within a complex tissue has become an important field of research [93].

2.6 Fluorescent-Activated Cell Sorting

First applications of flow cytometry instruments reach back to the early 1970s, and since then numerous patents and methodological advancements have enabled the robust and reliable commercial flow cytometry used today [94–97]. Within the methodological spectrum of flow cytometry, FACS provides the ability to enrich and isolate particular cell types, collecting them in separate target vessels (Fig. 2.3b) [2, 13]. Nowadays, FACS is the accepted worldwide standard in analysis and sorting of cell populations [98], probably also due to the widespread distribution of devices.

By laser excitation and emission FACS devices read out multiple parameters offering various analytical options (Fig. 2.3b). Cellular

properties like relative size and granularity are registered using the forward scatter (FSC) and side scatter (SSC), respectively. This information is further relevant to discriminate between single cells and doublets, or cell cluster, which is highly important to ensure single-cell isolation. Furthermore, fluorescence signals resulting from induced reporter expression or from preceeding labeling of cell surface markers can be gathered to classify subtypes of cells [2, 99].

The suspension is pressure driven through the system and lined up using a sheath flow liquid utilizing the effect of hydrodynamic focus. To ensure single-cell separation, a relatively high liquid flow is established that consequently leads to a strong dilution of cells within the system [85]. The liquid stream is channeled through a small nozzle (typically 60–100 μm diameter) and afterwards disrupted into a continuous line of flying droplets by applying targeted ultrasound vibration. These droplets, of which some include cells, pass the different laser beams and detectors used to measure the physical, chemical, or optical properties. Analysis occurs online and cells are sorted by electrical deflection of droplets into different outlet channels according to the user-set gatings (Fig. 2.3b). In case of single-cell RNA sequencing, target cells can be sorted individually into 96- or 384-well plates. In the past, all collected cells needed to be processed manually before sequencing, causing enormous workload and thus hampering high-throughput studies. With the help of automation and individual pre-defined sequencing indices in the collection wells, more comprehensive studies are now manageable. With an advanced method called "index sort," it is even possible to retrospectively correlate scRNA-Seq data and protein expression of cell surface markers measured during isolation, which is of particular interest when cells have been stained and analyzed for multiple markers [100, 101].

Depending on the application, FACS can be carried out in different modes specialized on high-throughput, enrichment or purity, differing in the actual sorting rate (between some hundreds up to several thousands of cells/s) as well as capturing efficiency.

The major advantages of FACS are flexibility in terms of cell type, compatibility with standardized substrates and buffers, high levels of accuracy, high-throughput and last but not least the operability and cost efficiency as well as the high prevalence of operating systems (Table 2.1). Still, a proficient operator is required to consistently and accurately sort cells into the center of a well, so that cells are immersed in the lysis buffer [3]. Further, the operator needs to calibrate proper gating for the FSC and SSC to ensure low doublet rates, reported to be as low as 2.3% [102]. Consequently, the flow sheath liquid will be increased to reach such low values augmenting the sample volume for each isolated cell. This causes higher risks of co-isolation of cell debris or free RNA contaminations [51]. The high liquid volumes also limit sorting speed and hence throughput, which will be especially obvious when low-abundant cell types need to be collected from a large number of suspended cells. Nevertheless, FACS is suitable to isolate rare cell populations with less than 1% but it necessitates a comparably large amount of starting material (>10,000 cells). As FACS analysis relies on cell suspensions, dissociation of solid tissue with all the aforementioned disadvantages has to precede (see Sect. 2.2).

Due to the high-pressure flow within the system cell viability may be an issue when applying FACS [103, 104]. Moreover, the osmotic and pressure stress occuring during sorting was shown to induce changes in cell expression profiles [5–7]. Additionally, labeling of cell surface receptors mostly involves their stimulation with antigens potentially leading to intracellular signaling and hence, an altered phenotype, necessitating to keep the time of labeling and sorting procedure as short as possible [85]. As already mentioned in former chapters, FACS is routinely applied to enrich cell populations prior to other approaches of single-cell isolation. Vice versa, pre-enrichment by other approaches like negative selection or depletion of unwanted cells through magnetic-activated cell sorting (MACS) likewise improves FACS results.

Due to its broad applicability, FACS studies cover nearly every cell type ranging from

blood cells, bone marrow cells, tumor cells, neurons, plants, protoplasts, yeast, to bacteria and even viruses. Similarly, nearly all research fields have integrated FACS applications in their potential workflow resulting in methods for cell cycle analysis [105], quantification of subpopulations [5], analysis of hematopoietic stem cells [35, 36], apoptosis [106] or DNA content analysis as well as immunophenotyping [107], quantification of soluble molecules [108], microbial analysis [109], and cancer diagnostics [110, 111].

2.7 Microfluidic Approaches

In the last years, innovative microfluidic technologies have paved new ways in single-cell isolation and analysis [112] with several microfluidic or lab-on-a-chip devices [113, 114]. Microfluidic chips, mostly lithographically fabricated onto polydimethylsiloxane (PDMS), glass, or silicon, enable sample compartmentalization and the control of nanoliter reactions. Due to their low-volume reactions, they serve as an ideal method for high-throughput single-cell separation for sequencing-related downstream processing (Table 2.1) [104, 115]. Microfluidic systems are in general fully automated and closed, which reduces the risk of external contamination. Even sterile working conditions are easy to achieve as microfluidic systems often rely on disposable elements (e.g., microfluidic chips), representing a major drawback for most of the aforementioned methods. Still, internal cross contamination risk from lysed or damaged cells within the system differs tremendously, dependent on the applied approach, hence requiring monitoring. Common advantages of all microfluidic approaches are high-throughput with minimal effort, cost effectiveness, and high accuracy. On the other hand, microfluidic chips are often less applicable for heterogeneous cell sizes based on their inherent chip design and channel properties.

Nowadays, microfluidic single-cell isolation methods can be roughly categorized based on three operating principles, namely, microstructures, hydrodynamic effect-based methods [116], and droplet-based assays [117–119] (Fig. 2.4). These basic approaches can be upgraded with a multitude of technologies like optical tools [120, 121], immunomagnetophoresis [13, 99], dielectrophoresis [122], or Raman spectroscopy [123] enabling a high specificity of desired cell types. In this paragraph, we will focus on single-cell isolation technologies applicable for downstream sequencing analysis. Hence, the following discussed methods are not exhaustive for all possibilities offered for single-cell approaches and may further be applied to different non-sequencing approaches like single-cell culture or stimulation experiments.

2.7.1 Microstructure

Nanoliter well devices [116, 124] comprised of high-density arrays of wells (microstructures or compartments) that serve as miniature target vessels for single cells were probably one of the first application of microfluidics for biological purpose (Fig. 2.4a).

In their simplest versions, the microstructures are loaded with low density cell solutions ensuring one cell per microstructure, through which high throughput single-cell analysis can be achieved, capable of capturing hundreds to millions of individual cells at the same time [116, 125]. Dependent on the application, the microstructures are sealed with capping structures (e.g., glass slides) or hydrophobic liquids (e.g., oil), providing an isolated reaction volume per well. The microwells are also called mechanical or hydrodynamic traps as the cells within a suspension are passively separated by physical borders printed on the microfluidic chip and finally fixed in “traps.” Typically, doublets are attempted to be avoided by adjusting the microstructure size to the average target cell size. This results in reported doublet rates up to 30% using microwell encapsulation systems [119]. It is claimed that for the commercial platforms Clontech ICell8 and Fluidigm C1 relying on post-capture visual control for empty wells and doublets, this rate was decreased to 3%. For the

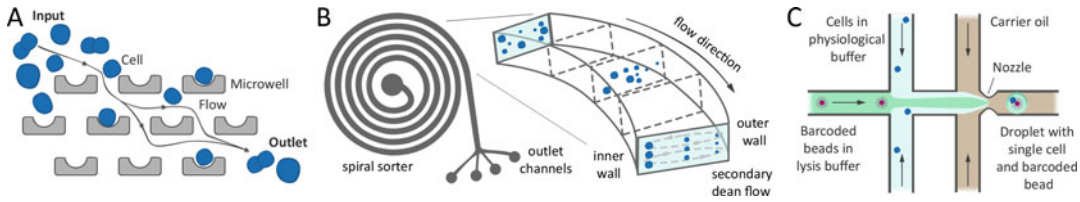


Fig. 2.4 Microfluidic approaches: (**a**) Single cells are passively trapped in microwells according to their size. Microwells can be sealed with a membrane or glass slide and inspected microscopically after isolation. Cells or components can be re-isolated from the wells for further processing. (**b**) Spiral sorter uses the dean flow effect for size/weight separation of cells and possible isolation by different outlet channels. Randomly dispersed cells at the entrance are focused on their way through the spiral at distinct positions in the channel enabling separation and isolation of cells. (**c**) Schematic overview of Drop-Seq microfluidic setup. An aqueous stream of cells (light blue) is co-flowing with an aqueous phase of lysis buffer (light green) containing individually barcoded beads for mRNA capturing. This co-flow stream is broken up by an oil stream (brown) into individual droplets containing random distribution of cells and beads. The cells are lysed inside the droplets and the mRNA binds to the barcoded beads. Afterwards all droplets are broken, and the material can be pooled and collected for downstream processing

Fluidigm C1 microfluidic robotic platform, the captured single cells can be automatically stained within the system for monitoring viability, surface markers, or reporter genes under visual control prior to cell lysis. The device automatically prepares indexed single-cell cDNA templates, applicable for downstream qPCR or sequencing analysis [126]. Since available chip designs are tuned to three size ranges (5–10, 10–17, and 17–25 μm in cell diameter based on the microstructure size), cell suspension needs to be relatively homogeneous in size. In combination with pre-enrichment steps for target cells from heterogeneous samples, these methods represent quite powerful and easy-to-handle tools to analyze large number of cells. Alternatively, recent studies in custom-designed arrays utilized immobilized antibodies on the microwell surface for subtype-specific cell capturing with a capture rate of about 95% [127], or coating with cell-recognizable aptamers to achieve cell type selectivity by encouraging strong 3D local topographic interactions of target cells with the microwell surface [128].

2.7.2 Methods Utilizing Certain Hydrodynamic Effects

Several microfluidic methods take advantage of certain hydrodynamic effects. One frequently exploited effect is the Dean flow, which is induced by fluid inertia resulting in the formation of a vortex that is perpendicular to the original flow direction. Cells with varying sizes, densities, or shapes behave differently to this inertial effect, through which they are concentrated at distinct locations within the vortex (Fig. 2.4b). This enables a passive and label-free isolation of single cells. Using Dean flow, the difficult task to capture CTCs from blood was achieved in high-throughput (1.7 mL/min) followed by fluorescence in situ hybridization detection for DNA analysis [129]. In combination with droplet encapsulation, such sorting approach was further exploited to establish a high-yield/speed single-cell isolation system reaching a yield of about 77% at a speed of 2700 cells/s [130]. In contrast to microwell-based systems, these hydrodynamic approaches establish a continuous fluid flow allowing for even higher throughput compared to hardware-limited numbers of microwells on a chip.

The development of valves on a chip, back in the early 2000 [131], likewise relying on hydrodynamic effects, helped to overcome other limitations of microwells. These devices exhibit high programmability and a relatively high-throughput by active modulation of the liquid flow inside the microfluidic chips, being routinely applied for single-cell isolation [15, 131–134]. Due to the precise fluid control exerted by

combinations of valves, a complex series of operations can be conducted on-chip, starting from single-cell isolation, reagent addition, lysis, and lysate retrieval, achieving a high degree of automation and parallelization.

Pressure-actuated valves can be applied to capture single cells in small reaction chambers with nearly 100% efficiency, which is particularly salient for rare samples and biopsies [135]. Different forms of hydrodynamic methods, such as hydrophoretic [136] and cross-flow principles [137], or pinched flow fractionation [138] have also been investigated, but are not routinely applied for single-cell isolation. Several other methods like hydrodynamic tweezers [139] or audible frequency oscillation-based tweezers [140] that induce liquid vortices for cell trapping, are currently not applicable for single-cell sequencing approaches, as proper handling of the lysed single cells is not achieved yet. Taken together, if applicable, single-cell methods using hydrodynamic effects represent a cost-efficient way to achieve high-throughput and precision in microfluidic chips. However, there are several limitations, such as reliability of device fabrication and necessity of complex optimization of the microfluidic design. Moreover, the user interface needs to be addressed for advanced applications [34]. Although valve-based systems in particular have been shown to overcome the limitations of microwells, like limited operability and downstream molecular analysis, the complex setup and required precise computer-controlled pneumatics complicate the fabrication and operation, consequently increasing the costs.

2.7.3 Droplet-Based Microfluidic Approaches

So far probably the most promising technology for controllable handling of minimal volumes of fluids in single-cell applications represent droplet-based microfluidic approaches (Fig. 2.4c) [141, 142]. Droplet-based systems produce micrometer-scaled, surfactant-stabilized droplets from an aqueous phase in an inert carrier oil [118, 119, 141, 142]. These droplets constitute individual capture and reaction volumes separated by the inert oil phase and surfactants, together preventing cross contamination by limiting the diffusion of analytes (e.g., DNA or RNA) into the oil and neighboring droplets. Droplet-based microfluidic concepts encapsulate single cells either randomly according to Poisson's distribution [117] (similar to limiting serial dilution methods discussed above) or with even higher efficiency reaching more than 80% [143]. The high-throughput character and the high frequency of droplet generation with nanoliter reaction volumes rely on a continuous flow of the aqueous phase. This minimizes the required reagent volumes, while enhancing the reaction efficiency within the individual droplets.

Based on this, the major breakthrough in regard to single-cell transcriptome sequencing was achieved in 2015 by two approaches called *inDrop* [118] and *Drop-Seq* [119], combining microfluidics and nucleotide barcoding for retrospective identification. These approaches enable parallel analysis of thousands of cells for a large variety of sequencing methods [117, 118, 144–146]. These technologies were transferred into commercialized and optimized platforms such as 10× Genomics (Pleasanton, CA, USA) [29], Dolomite Nadia (Royston, Herts, UK), and the ddSeq system (Bio-Rad, Hercules, CA, USA), increasing the flexibility and leading to dramatic cost reduction in matters of single-cell mRNA library generation.

Focusing on the initially published 3′ poly-A capture methods [118, 119], droplet-based platforms employ beads or gel spheres containing poly-dT oligonucleotide capture sequences coupled to two barcode sequences. One barcode is bead-specific that serves to identify the RNA of an individual cell. The second barcode called "unique molecular identifier" (UMI) is a short random sequence allowing for accurate quantification of each expressed transcript per cell [119]. In both approaches, *Drop-Seq* and *inDrop,* the hybridization of the mRNA to the primer sequences is performed within the droplets [119]. The droplets additionally serve as compartment for the reverse transcription reaction when applying *inDrop* [118]. Afterwards the droplets

are chemically broken and all downstream processes are accomplished in a one-tube reaction for all pooled cells [118, 119].

A great advantage of droplets-in-oil-based cell separation and sorting technologies is the tremendous throughput of up to several thousand single cells per second [143] and the low technical noise. Moreover, the possibility to handle and modify each single droplet is unique to this technology and allows for subsequent merging, incubation, reinjection, and sorting of the droplets.

2.8 Future Perspective

The field of single-cell analysis including the important steps of isolation and separation is rapidly expanding. New advances in technologies enable the interrogation of single cells at unprecedented resolution with progressively decreasing costs. This course will probably continue with novel techniques aimed at capturing even higher number of cells per experiment and the extraction of several layers of cellular information at once. Such multi-omic approaches will allow simultaneous analysis of genomic, chromatin, epigenomic, transcriptional as well as proteomic states [40, 147–150]. So far, multi-omics can be applied at single-cell level only by the use of low-throughput methods, but it is simply a matter of time until these approaches will become available for high-throughput platforms, most likely within the scope of microfluidics.

Especially, single-cell droplet microfluidic applications have increased significantly over the last years and further exploration of this research field can be expected. Due to the rapid technological progress within the microfluidic market, even more complex and multiplexing experimental designs may soon allow for automated patch-clamp recordings and intracellular calcium measurements [151, 152], followed by transcriptome and proteome analysis [153–155]. Future improvements might further focus on handling, sorting, or storage of single-cell-containing droplets as well as on integrative workflows for downstream processing, standardized approaches for barcoding and packaging, or increasing the sensitivity. The technical integration of sorting capabilities like FACS into microfluidic chip designs are ongoing and will enable fully integrated systems for the enrichment and targeted isolation of cells of interest within a microfluidic environment.

Acknowledgement *Funding*: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—368482240/GRK2416.

References

1. Tung PY et al (2017) Batch effects and the effective design of single-cell gene expression studies. Sci Rep 7:39921. https://doi.org/10.1038/srep39921
2. Gross A et al (2015) Technologies for single-cell isolation. Int J Mol Sci 16:16897–16919. https://doi.org/10.3390/ijms160816897
3. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58:610–620. https://doi.org/10.1016/j.molcel.2015.04.005
4. Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50:96. https://doi.org/10.1038/s12276-018-0071-8
5. van den Brink SC et al (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods 14:935–936. https://doi.org/10.1038/nmeth.4437
6. Xiong L, Lee H, Ishitani M, Zhu J-K (2002) Regulation of osmotic stress-responsive gene expression by the LOS6/ABA1 locus in Arabidopsis. J Biol Chem 277:8588–8596
7. Romero-Santacreu L, Moreno J, Perez-Ortin JE, Alepuz P (2009) Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. RNA 15:1110–1120. https://doi.org/10.1261/rna.1435709
8. Ilie M et al (2014) Current challenges for detection of circulating tumor cells and cell-free circulating nucleic acids, and their characterization in non-small cell lung carcinoma patients. What is the best blood substrate for personalized medicine? Ann Transl Med 2:107
9. Hu P, Zhang W, Xin H, Deng G (2016) Single cell isolation and analysis. Front Cell Dev Biol 4:116. https://doi.org/10.3389/fcell.2016.00116
10. Chen G, Ning B, Shi T (2019) Single-cell RNA-seq technologies and related computational data analysis. Front Genet 10:317. https://doi.org/10.3389/fgene.2019.00317
11. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet

24:133–141. https://doi.org/10.1016/j.tig.2007.12.007
12. Adam M, Potter AS, Potter SS (2017) Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Development 144:3625–3632. https://doi.org/10.1242/dev.151142
13. Lafzi A, Moutinho C, Picelli S, Heyn H (2018) Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. Nat Protoc 13:2742–2757. https://doi.org/10.1038/s41596-018-0073-y
14. Autengruber A, Gereke M, Hansen G, Hennig C, Bruder D (2012) Impact of enzymatic tissue disintegration on the level of surface molecule expression and immune cell function. Eur J Microbiol Immunol 2:112
15. Sanchez-Adams J, Athanasiou KA (2012) Regional effects of enzymatic digestion on knee meniscus cell yield and phenotype for tissue engineering. Tissue Eng Part C Methods 18:235–243. https://doi.org/10.1089/ten.TEC.2011.0383
16. Jahan-Tigh RR, Ryan C, Obermoser G, Schwarzenberger K (2012) Flow cytometry. J Invest Dermatol 132:1–6. https://doi.org/10.1038/jid.2012.282
17. Qiu X, De Jesus J, Pennell M, Troiani M, Haun JB (2015) Microfluidic device for mechanical dissociation of cancer cell aggregates into single cells. Lab Chip 15:339–350. https://doi.org/10.1039/c4lc01126k
18. Meeson A, Fuller A, Breault DT, Owens WA, Richardson GD (2013) Optimised protocols for the identification of the murine cardiac side population. Stem Cell Rev Rep 9:731–739. https://doi.org/10.1007/s12015-013-9440-9
19. Baldan V, Griffiths R, Hawkins RE, Gilham DE (2015) Efficient and reproducible generation of tumour-infiltrating lymphocytes for renal cell carcinoma. Br J Cancer 112:1510–1518. https://doi.org/10.1038/bjc.2015.96
20. Guillaumet-Adkins A et al (2017) Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol 18:45. https://doi.org/10.1186/s13059-017-1171-9
21. Alles J et al (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol 15:44. https://doi.org/10.1186/s12915-017-0383-5
22. Wang W, Penland L, Gokce O, Croote D, Quake SR (2018) High fidelity hypothermic preservation of primary tissues in organ transplant preservative for single cell transcriptome analysis. BMC Genomics 19:140
23. Lacar B et al (2016) Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat Commun 7:11022
24. Krishnaswami SR et al (2016) Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. Nat Protoc 11:499
25. Habib N (2016) Div-Seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. Science 353:925
26. Bakken TE et al (2018) Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One 13:e0209648. https://doi.org/10.1371/journal.pone.0209648
27. Lake BB et al (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36:70–80. https://doi.org/10.1038/nbt.4038
28. Habib N et al (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods 14:955–958. https://doi.org/10.1038/nmeth.4407
29. Zheng GX et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049. https://doi.org/10.1038/ncomms14049
30. Grindberg RV et al (2013) RNA-sequencing from single nuclei. Proc Natl Acad Sci U S A 110:19802–19807. https://doi.org/10.1073/pnas.1319700110
31. Lake BB et al (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352:1586–1590. https://doi.org/10.1126/science.aaf1204
32. Thomsen ER et al (2016) Fixed single-cell transcriptomic characterization of human radial glial diversity. Nat Methods 13:87–93. https://doi.org/10.1038/nmeth.3629
33. Cho H et al (2018) Microfluidic technologies for circulating tumor cell isolation. Analyst 143:2936–2970. https://doi.org/10.1039/c7an01979c
34. Huang Q, Mao S, Khan M, Lin JM (2019) Single-cell assay on microfluidic devices. Analyst 144:808–823. https://doi.org/10.1039/c8an01079j
35. Radbruch A, Recktenwald D (1995) Detection and isolation of rare cells. Curr Opin Immunol 7:270–273
36. Will B, Steidl U (2010) Multi-parameter fluorescence-activated cell sorting and analysis of stem and progenitor cells in myeloid malignancies. Best Pract Res Clin Haematol 23:391–401. https://doi.org/10.1016/j.beha.2010.06.006
37. Kornbluth J, Hoover RG (1989) Immunobiology of HLA. Springer, New York, NY, pp 150–152
38. Christaki E et al (2011) A monoclonal antibody against rage alters gene expression and is protective in experimental models of sepsis and pneumococcal pneumonia. Shock 35:492–498. https://doi.org/10.1097/SHK.0b013e31820b2e1c
39. Victora GD et al (2010) Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. Cell 143:592–605. https://doi.org/10.1016/j.cell.2010.10.032
40. Medaglia C et al (2017) Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science 358:1622–1626. https://doi.org/10.1126/science.aao4277

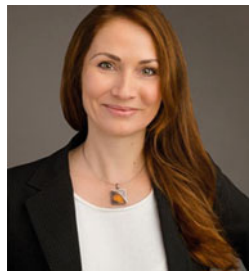
41. Patterson GH, Lippincott-Schwartz J (2002) A photoactivatable GFP for selective photolabeling of proteins and cells. Science 297:1873–1877. https://doi.org/10.1126/science.1074952
42. Tsutsui H et al (2009) The E1 mechanism in photo-induced beta-elimination reactions for green-to-red conversion of fluorescent proteins. Chem Biol 16:1140–1147. https://doi.org/10.1016/j.chembiol.2009.10.010
43. Chtanova T et al (2014) Real-time interactive two-photon photoconversion of recirculating lymphocytes for discontinuous cell tracking in live adult mice. J Biophotonics 7:425–433. https://doi.org/10.1002/jbio.201200175
44. Suan D et al (2015) T follicular helper cells have distinct modes of migration and molecular signatures in naive and memory immune responses. Immunity 42:704–718. https://doi.org/10.1016/j.immuni.2015.03.002
45. Jiang L et al (2011) Synthetic spike-in standards for RNA-seq experiments. Genome Res 21:1543–1551. https://doi.org/10.1101/gr.121095.111
46. Hardwick SA et al (2016) Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13:792–798. https://doi.org/10.1038/Nmeth.3958
47. Yilmaz S, Singh AK (2012) Single cell genome sequencing. Curr Opin Biotechnol 23:437–443. https://doi.org/10.1016/j.copbio.2011.11.018
48. Tang F et al (2009) mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. https://doi.org/10.1038/nmeth.1315
49. Pensold D et al (2017) The DNA methyltransferase 1 (DNMT1) controls the shape and dynamics of migrating POA-derived interneurons fated for the murine cerebral cortex. Cereb Cortex 27:5696–5714. https://doi.org/10.1093/cercor/bhw341
50. Gerstmann K et al (2015) Thalamic afferents influence cortical progenitors via ephrin A5-EphA4 interactions. Development 142:140–150. https://doi.org/10.1242/dev.104927
51. Hempel CM, Sugino K, Nelson SB (2007) A manual method for the purification of fluorescently labeled neurons from the mammalian brain. Nat Protoc 2:2924–2929. https://doi.org/10.1038/nprot.2007.416
52. Grindberg RV et al (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS One 6:e18565. https://doi.org/10.1371/journal.pone.0018565
53. Xue Z et al (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500:593–597. https://doi.org/10.1038/nature12364
54. Bengtsson M, Stahlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 15:1388–1392. https://doi.org/10.1101/gr.3820805
55. Guo G et al (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev Cell 18:675–685. https://doi.org/10.1016/j.devcel.2010.02.012
56. Hodne K, Haug TM, Weltzien FA (2010) Single-cell qPCR on dispersed primary pituitary cells - an optimized protocol. BMC Mol Biol 11:82. https://doi.org/10.1186/1471-2199-11-82
57. Citri A, Pang ZPP, Sudhof TC, Wernig M, Malenka RC (2012) Comprehensive qPCR profiling of gene expression in single neuronal cells. Nat Protoc 7:118–127. https://doi.org/10.1038/nprot.2011.430
58. Li HH et al (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414–417. https://doi.org/10.1038/335414a0
59. Eberwine J et al (1992) Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A 89:3010–3014. https://doi.org/10.1073/pnas.89.7.3010
60. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K (2018) Experimental considerations for single-cell RNA sequencing approaches. Front Cell Dev Biol 6:108. https://doi.org/10.3389/fcell.2018.00108
61. Schoendube J, Wright D, Zengerle R, Koltay P (2015) Single-cell printing based on impedance detection. Biomicrofluidics 9:014117. https://doi.org/10.1063/1.4907896
62. Cho SH, Chen CH, Tsai FS, Godin J, Lo YH (2010) Mammalian cell sorting using muFACS. Lab Chip 10:1567
63. Yusof A et al (2011) Inkjet-like printing of single-cells. Lab Chip 11:2447–2454. https://doi.org/10.1039/c1lc20176j
64. Liss BJ (2002) Improved quantitative real-time RT–PCR for expression profiling of individual cells. Nucleic Acids Res 30:e89
65. Monyer H, Lambolez B (1995) Molecular biology and physiology at the single-cell level. Curr Opin Neurobiol 5:382–387
66. Tsuzuki K, Lambolez B, Rossier J, Ozawa S (2001) Absolute quantification of AMPA receptor subunit mRNAs in single hippocampal neurons. J Neurochem 77:1650–1659. https://doi.org/10.1046/j.1471-4159.2001.00388.x
67. Lu Z et al (2010) 2010 IEEE International Conference on Robotics and Automation. IEEE, Washington, DC, pp 494–499
68. Frohlich J, Konig H (2000) New techniques for isolation of single prokaryotic cells. FEMS Microbiol Rev 24:567–572
69. Brehm-Stecher BF, Johnson EA (2004) Single-cell microbiology: tools, technologies, and applications. Microbiol Mol Biol Rev 68:538–559. https://doi.org/10.1128/MMBR.68.3.538-559.2004. Table of contents

70. Wright G, Tucker MJ, Morton PC, Sweitzer-Yoder CL, Smith SE (1998) Micromanipulation in assisted reproduction: a review of current technology. Curr Opin Obstet Gynecol 10:221–226
71. Li CX et al (2011) New cell separation technique for the isolation and analysis of cells from biological mixtures in forensic caseworks. Croat Med J 52:293–298. https://doi.org/10.3325/cmj.2011.52.293
72. Hongoh Y et al (2008) Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc Natl Acad Sci U S A 105:5555–5560. https://doi.org/10.1073/pnas.0801389105
73. Kvist T, Ahring BK, Lasken RS, Westermann P (2007) Specific single-cell isolation and genomic amplification of uncultured microorganisms. Appl Microbiol Biotechnol 74:926–935. https://doi.org/10.1007/s00253-006-0725-7
74. Ramskold D et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30:777–782. https://doi.org/10.1038/nbt.2282
75. Emmert-Buck MR et al (1996) Laser capture microdissection. Science 274:998–1001. https://doi.org/10.1126/science.274.5289.998
76. Espina V, Heiby M, Pierobon M, Liotta LA (2007) Laser capture microdissection technology. Expert Rev Mol Diagn 7:647–657. https://doi.org/10.1586/14737159.7.5.647
77. Fend F, Raffeld M (2000) Laser capture microdissection in pathology. J Clin Pathol 53:666–672. https://doi.org/10.1136/jcp.53.9.666
78. Foley JW et al (2019) Gene-expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ. Genome Res 29:1816
79. Esposito G (2007) Microarray technology and cancer gene profiling. Springer, New York, NY, pp 54–65
80. Nakamura N et al (2007) Single cell diagnostics. Springer, New York, NY, pp 11–18
81. Walch A et al (2001) Tissue microdissection techniques in quantitative genome and gene expression analyses. Histochem Cell Biol 115:269–276
82. Bonner RF et al (1997) Laser capture microdissection: molecular analysis of tissue. Science 278:1481–1483. https://doi.org/10.1126/science.278.5342.1481
83. Schüitze K, Lahr GJ (1998) Identification of expressed genes by laser-mediated manipulation of single cells. Nat Biotechnol 16:737
84. Podgorny OV (2013) Live cell isolation by laser microdissection with gravity transfer. J Biomed Opt 18:55002. https://doi.org/10.1117/1.JBO.18.5.055002
85. Hodne K, Weltzien FA (2015) Single-cell isolation and gene analysis: pitfalls and possibilities. Int J Mol Sci 16:26832–26849. https://doi.org/10.3390/ijms161125996
86. Bohm M, Wieland I, Schutze K, Rubben H (1997) Microbeam MOMeNT: non-contact laser microdissection of membrane-mounted native tissue. Am J Pathol 151:63–67
87. Bevilacqua C, Makhzami S, Helbling JC, Defrenaix P, Martin P (2010) Maintaining RNA integrity in a homogeneous population of mammary epithelial cells isolated by laser capture microdissection. BMC Cell Biol 11:95. https://doi.org/10.1186/1471-2121-11-95
88. DeCarlo K, Emley A, Dadzie OE, Mahalingam M (2011) Laser capture microdissection. Springer, New York, NY, pp 1–15
89. Liu A (2010) Laser capture microdissection in the tissue biorepository. J Biomol Tech 21:120–125
90. Keays KM, Owens GP, Ritchie AM, Gilden DH, Burgoon MP (2005) Laser capture microdissection and single-cell RT-PCR without RNA purification. J Immunol Methods 302:90–98. https://doi.org/10.1016/j.jim.2005.04.018
91. Vandewoestyne M, Deforce D (2010) Laser capture microdissection in forensic research: a review. Int J Legal Med 124:513–521. https://doi.org/10.1007/s00414-010-0499-4
92. Fink L, Kwapiszewska G, Wilhelm J, Bohle RM (2006) Laser-microdissection for cell type- and compartment-specific analyses on genomic and proteomic level. Exp Toxicol Pathol 57(Suppl 2):25–29. https://doi.org/10.1016/j.etp.2006.02.010
93. Fink L, Bohle RM (2005) Laser capture microdissection. Springer, New York, NY, pp 167–185
94. Moldavan A (1934) Photo-electric technique for the counting of microscopical cells. Science 80:188–189. https://doi.org/10.1126/science.80.2069.188
95. Gucker FT Jr, O'Konski CT, Pickard HB, Pitts JN Jr (1947) A photoelectronic counter for colloidal particles1. J Am Chem Soc 69:2422–2431
96. Fulwyler MJ (1965) Electronic separation of biological cells by volume. Science 150:910–911. https://doi.org/10.1126/science.150.3698.910
97. Shapiro HM (2005) Practical flow cytometry. John Wiley & Sons, New York, NY
98. Herzenberg LA et al (2002) The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. Clin Chem 48:1819–1827
99. Nguyen A, Khoo WH, Moran I, Croucher PI, Phan TG (2018) Single cell RNA sequencing of rare immune cell populations. Front Immunol 9:1553. https://doi.org/10.3389/fimmu.2018.01553
100. Wilson NK et al (2015) Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. Cell Stem Cell 16:712–724. https://doi.org/10.1016/j.stem.2015.04.004
101. Paul F et al (2015) Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 163:1663–1677. https://doi.org/10.1016/j.cell.2015.11.013

102. Jaitin DA et al (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343:776–779. https://doi.org/10.1126/science.1247651
103. Mollet M, Godoy-Silva R, Berdugo C, Chalmers JJ (2008) Computer simulations of the energy dissipation rate in a fluorescence-activated cell sorter: implications to cells. Biotechnol Bioeng 100:260–272. https://doi.org/10.1002/bit.21762
104. Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 14:618–630. https://doi.org/10.1038/nrg3542
105. Hashimshony T et al (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol 17:77. https://doi.org/10.1186/s13059-016-0938-8
106. Valet G (2003) Past and present concepts in flow cytometry: a European perspective. J Biol Regul Homeost Agents 17:213–222
107. Iriondo O, Rabano M, Vivanco MD (2015) FACS sorting mammary stem cells. Methods Mol Biol 1293:63–72. https://doi.org/10.1007/978-1-4939-2519-3_3
108. Brown M, Wittwer C (2000) Flow cytometry: principles and clinical applications in hematology. Clin Chem 46:1221–1229
109. Davey HM, Kell DB (1996) Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiol Rev 60:641–696
110. Lacombe F, Belloc F (1996) Flow cytometry study of cell cycle, apoptosis and drug resistance in acute leukemia. Hematol Cell Ther 38:495–504
111. McCoy JP Jr, Carey JL (1990) Recent advances in flow cytometric techniques for cancer detection and prognosis. Immunol Ser 53:171–187
112. Guo MT, Rotem A, Heyman JA, Weitz DA (2012) Droplet microfluidics for high-throughput biological assays. Lab Chip 12:2146–2155. https://doi.org/10.1039/c2lc21147e
113. Sims CE, Allbritton NL (2007) Analysis of single mammalian cells on-chip. Lab Chip 7:423–440. https://doi.org/10.1039/b615235j
114. Lecault V, White AK, Singhal A, Hansen CL (2012) Microfluidic single cell analysis: from promise to practice. Curr Opin Chem Biol 16:381–390. https://doi.org/10.1016/j.cbpa.2012.03.022
115. Gomez-Sjoberg R, Leyrat AA, Pirone DM, Chen CS, Quake SR (2007) Versatile, fully automated, microfluidic cell culture system. Anal Chem 79:8557–8563. https://doi.org/10.1021/ac071311w
116. Di Carlo D, Wu LY, Lee LP (2006) Dynamic single cell culture array. Lab Chip 6:1445–1449. https://doi.org/10.1039/b605937f
117. Brouzes E et al (2009) Droplet microfluidic technology for single-cell high-throughput screening. Proc Natl Acad Sci U S A 106:14195–14200. https://doi.org/10.1073/pnas.0903542106
118. Klein AM et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161:1187–1201. https://doi.org/10.1016/j.cell.2015.04.044
119. Macosko EZ et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
120. Piggee C (2009) Optical tweezers: not just for physicists anymore. Anal Chem 81:16–19. https://doi.org/10.1021/ac8023203
121. Ashkin A, Dziedzic JM, Yamane T (1987) Optical trapping and manipulation of single cells using infrared laser beams. Nature 330:769–771. https://doi.org/10.1038/330769a0
122. Di Trapani M, Manaresi N, Medoro G (2018) DEPArray system: an automatic image-based sorter for isolation of pure circulating tumor cells. Cytometry A 93:1260–1266. https://doi.org/10.1002/cyto.a.23687
123. Zhang P et al (2015) Raman-activated cell sorting based on dielectrophoretic single-cell trap and release. Anal Chem 87:2282–2289. https://doi.org/10.1021/ac503974e
124. Torres AJ, Hill AS, Love JC (2014) Nanowell-based immunoassays for measuring single-cell secretion: characterization of transport and surface binding. Anal Chem 86:11562–11569. https://doi.org/10.1021/ac4030297
125. Han X et al (2018) Mapping the mouse cell atlas by microwell-seq. Cell 172:1091–1107.e17
126. Kim KT et al (2015) Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. Genome Biol 16:127. https://doi.org/10.1186/s13059-015-0692-3
127. Revzin A, Sekine K, Sin A, Tompkins RG, Toner M (2005) Development of a microfabricated cytometry platform for characterization and sorting of individual leukocytes. Lab Chip 5:30–37. https://doi.org/10.1039/b405557h
128. Chen Q, Wu J, Zhang Y, Lin Z, Lin JM (2012) Targeted isolation and analysis of single tumor cells with aptamer-encoded microwell array on microfluidic device. Lab Chip 12:5180–5185. https://doi.org/10.1039/c2lc40858a
129. Warkiani ME et al (2016) Ultra-fast, label-free isolation of circulating tumor cells from blood using spiral microfluidics. Nat Protoc 11:134–148. https://doi.org/10.1038/nprot.2016.003
130. Kemna EW et al (2012) High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel. Lab Chip 12:2881–2887. https://doi.org/10.1039/c2lc00013j
131. Unger MA, Chou HP, Thorsen T, Scherer A, Quake SR (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. Science 288:113–116. https://doi.org/10.1126/science.288.5463.113

132. Prakadan SM, Shalek AK, Weitz DA (2017) Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. Nat Rev Genet 18:345
133. Lu Y, Yang L, Wei W, Shi Q (2017) Microchip-based single-cell functional proteomics for biomedical applications. Lab Chip 17:1250–1263. https://doi.org/10.1039/C7LC00037E
134. Murphy TW, Zhang Q, Naler LB, Ma S, Lu C (2018) Recent advances in the use of microfluidic technologies for single cell analysis. Analyst 143:60–80. https://doi.org/10.1039/C7AN01346A
135. Streets AM et al (2014) Microfluidic single-cell whole-transcriptome sequencing. Proc Natl Acad Sci U S A 111:7048–7053. https://doi.org/10.1073/pnas.1402030111
136. Choi S, Song S, Choi C, Park JK (2007) Continuous blood cell separation by hydrophoretic filtration. Lab Chip 7:1532–1538. https://doi.org/10.1039/b705203k
137. Earhart CM et al (2014) Isolation and mutational analysis of circulating tumor cells from lung cancer patients with magnetic sifters and biochips. Lab Chip 14:78–88. https://doi.org/10.1039/c3lc50580d
138. Lu X, Xuan X (2015) Continuous microfluidic particle separation via elasto-inertial pinched flow fractionation. Anal Chem 87:6389–6396. https://doi.org/10.1021/acs.analchem.5b01432
139. Karimi A, Yazdi S, Ardekani AM (2013) Hydrodynamic mechanisms of cell and particle trapping in microfluidics. Biomicrofluidics 7:21501. https://doi.org/10.1063/1.4799787
140. Lutz BR, Chen J, Schwartz DT (2006) Hydrodynamic tweezers: 1. Noncontact trapping of single cells using steady streaming microeddies. Anal Chem 78:5429–5435. https://doi.org/10.1021/ac060555y
141. Joensson HN, Svahn HA (2012) Droplet microfluidics—a tool for single-cell analysis. Angew Chem Int Ed 51:12176–12192
142. Seemann R, Brinkmann M, Pfohl T, Herminghaus S (2012) Droplet based microfluidics. Rep Prog Phys 75:016601. https://doi.org/10.1088/0034-4885/75/1/016601
143. Edd JF et al (2008) Controlled encapsulation of single-cells into monodisperse picolitre drops. Lab Chip 8:1262–1264. https://doi.org/10.1039/b805456h
144. Zilionis R et al (2017) Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc 12:44. https://doi.org/10.1038/nprot.2016.154
145. Collins DJ, Neild A, deMello A, Liu AQ, Ai Y (2015) The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation. Lab Chip 15:3439–3459. https://doi.org/10.1039/c5lc00614g
146. Koster S et al (2008) Drop-based microfluidic devices for encapsulation of single cells. Lab Chip 8:1110–1115. https://doi.org/10.1039/b802941e
147. Stoeckius M et al (2017) Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14:865–868. https://doi.org/10.1038/nmeth.4380
148. Macaulay IC et al (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 12:519. https://doi.org/10.1038/nmeth.3370
149. Angermueller C et al (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13:229–232. https://doi.org/10.1038/nmeth.3728
150. Clark SJ et al (2018) scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun 9:781. https://doi.org/10.1038/s41467-018-03149-4
151. Wheeler AR et al (2003) Microfluidic device for single-cell analysis. Anal Chem 75:3581–3586. https://doi.org/10.1021/ac0340758
152. Pantoja R et al (2004) Silicon chip-based patch-clamp electrodes integrated with PDMS microfluidics. Biosens Bioelectron 20:509–517. https://doi.org/10.1016/j.bios.2004.02.020
153. Wu AR et al (2014) Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods 11:41–46. https://doi.org/10.1038/nmeth.2694
154. Wu H, Wheeler A, Zare RN (2004) Chemical cytometry on a picoliter-scale integrated microfluidic chip. Proc Natl Acad Sci U S A 101:12809–12813. https://doi.org/10.1073/pnas.0405299101
155. Hong JW, Studer V, Hang G, Anderson WF, Quake SR (2004) A nanoliter-scale nucleic acid processor with parallel architecture. Nat Biotechnol 22:435–439. https://doi.org/10.1038/nbt951

Daniel Pensold is a post-doc and lab manager at the RWTH Aachen University, Germany in the group of Prof. Zimmer-Bensch. After finishing his PhD entitled "Single cell transcriptomics reveal regulators of progenitor cell fate and postmitotic maturation during brain development" in the research group "Functional Neuroepigenetics" at University Hospital Jena, Germany in 2017, he is involved in a post-doc project dealing with high-throughput single-cell transcriptomics. His research profile covers interneuron development and aging with focus on epigenetic and transcriptional network regulation.



Geraldine Zimmer-Bensch is a Distinguished Professor of Epigenetics at the RWTH Aachen University, Germany. Since the beginning of her academic career she was fascinated by brain and neuronal development. What are the mechanisms behind the highly orchestrated unfolding of such a complex organ and how external stimuli can influence cell-intrinsic developmental processes? In this context, the research of Geraldine Zimmer-Bensch is focused on the identification and validation of epigenetic regulatory networks, which direct discrete neurodevelopmental processes like differentiation, migration, and network formation, thereby integrating environmental information. The ultimate goal of her research is to approach causes for neurodevelopment defects and hence, the pathophysiology of associated diseases. Apart from generating and analyzing transgenic mouse models, single-cell-based approaches are an important aspect of her work.

3 Single-Cell Sequencing of T cell Receptors: A Perspective on the Technological Development and Translational Application

Shivai Gupta, Richard Witas, Alexandria Voigt, Touyana Semenova, and Cuong Q. Nguyen

Abstract

T cells recognize peptides bound to major histocompatibility complex (MHC) class I and class II molecules at the cell surface. This recognition is accomplished by the expression of T cell receptors (TCR) which are required to be diverse and adaptable in order to accommodate the various and vast number of antigens presented on the MHCs. Thus, determining TCR repertoires of effector T cells is necessary to understand the immunological process in responding to cancer progression, infection, and autoimmune development. Furthermore, understanding the TCR repertoires will provide a solid framework to predict and test the antigen which is more critical in autoimmunity. However, it has been a technical challenge to sequence the TCRs and provide a conceptual context in correlation to the vast number of TCR repertoires in the immunological system. The exploding field of single-cell sequencing has changed how the repertoires are being investigated and analyzed. In this review, we focus on the biology of TCRs, TCR signaling and its implication in autoimmunity. We discuss important methods in bulk sequencing of many cells. Lastly, we explore the most pertinent platforms in single-cell sequencing and its application in autoimmunity.

Keywords

T cells · T cell receptors · T cell receptor signaling · Single cell sequencing · Bulk sequencing · Sjogren's syndrome

3.1 Introduction

T cells are a critical component of the adaptive immune system capable of recognizing a nearly infinite variety of antigens. T cells recognize antigens through unique antigen receptors called T cell receptors (TCRs). Unlike the closely related B cell receptors (BCRs), which can secrete as an antibody, TCRs are strictly membrane bound and are not secreted [1, 2]. As a result, TCRs lack the opsonization and neutralization abilities of antibodies and are committed to

S. Gupta · A. Voigt · T. Semenova
Department of Infectious Diseases and Immunology, College of Veterinary Medicine, Gainesville, FL, USA

R. Witas
Department of Oral Biology, College of Dentistry, Gainesville, FL, USA

C. Q. Nguyen (✉)
Department of Infectious Diseases and Immunology, College of Veterinary Medicine, Gainesville, FL, USA

Department of Oral Biology, College of Dentistry, Gainesville, FL, USA

Center of Orphaned Autoimmune Diseases, University of Florida, Gainesville, FL, USA
e-mail: nguyenc@ufl.edu

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_3

participating in antigen recognition followed by intracellular signaling and subsequent T cell activation [3]. TCRs are further distinguished from BCRs by their mechanisms for antigen recognition. While BCRs can directly bind to a wide variety of molecules, TCRs are restricted to recognizing short peptides of protein antigens processed and presented by major histocompatibility complexes (MHCs) on the body's own antigen presenting cells (APCs) [4]. Fundamentally, TCRs function as an antigen recognition receptor complexed to T cell activating signaling machinery and are activated in response to short continuous amino acid sequences presented on MHC. TCRs are composed of two heterodimeric polypeptide chains linked by a disulfide bond [1]. Each chain of the TCR consists of two extracellular immunoglobulin domains, a transmembrane region and a short cytoplasmic tail [1]. The two extracellular domains are made up of the variable (V) region and constant (C) region [4]. The heterodimeric structure of the TCR is analogous to the heavy and light chain heterodimers of BCRs [5]. However, the forked structure of the BCR consists of two antigen binding sites, whereas each TCR possesses a single antigen binding site. The majority of TCRs possess an α chain and a β chain and are referred to as αβ TCRs. A subset of T cells possesses a γ chain and δ chain and are referred to as γδ TCRs. γδ TCRs are capable of directly recognizing antigens outside the context of MHC and are even capable of recognizing non-peptide antigens [6–8].

T cells possess the ability to bind to a vast array of peptide antigens through their TCRs; it has been estimated that humans can produce between 10^{15} and 10^{20} possible unique TCR chains [9]. This enormous variety is imparted by an unusual genetic mechanism, largely shared with BCR generation, that provides diversity concentrated in the antigen binding regions of the TCR [10]. The V region is the portion of the TCR that participates in antigen binding. The V region is not encoded by a single segment of DNA, but rather is composed of multiple gene segments that are rearranged through somatic DNA recombination. Combinatorial diversity afforded through recombination of the gene segments is further augmented by junctional diversity through the random addition of nucleotides at the interface between segments, thus allowing for the generation of a nearly limitless array of TCRs [11]. The DNA encoding the α chain of the TCR possesses multiple variable (V) and joining (J) segments, whereas the β chain possesses multiple V, diversity (D), and segments [11, 12] as represented in Fig. 3.1.

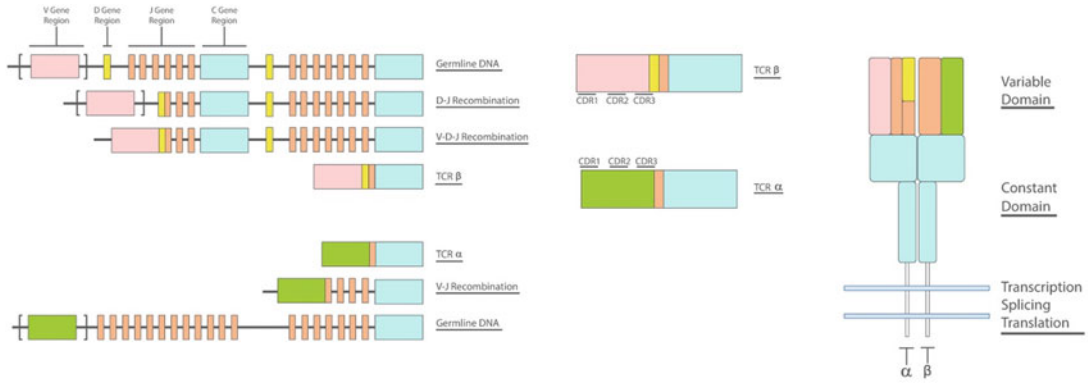


Fig. 3.1 The genetic basis of TCR formation. The figure represents the mRNAs somatic VDJ recombination to form the alpha and beta chains of TCRs. As indicated, there is an arrangement step that recombines the VDJ segment for TCR β and V and J segments for the TCR α chain. The mRNAs formed have addition and deletion of nucleotides at the junctions of these segments leading to junctional diversity that leads to variability for assessing specific antigens. There may be different combinations of genes leading to the final formation of the TCR that consists of the TCR α and β subunit organized in a constant and variable region wherein the variable region is responsible for antigen recognition

A functional TCR will consist of an α chain with a single V and J segment and a β chain with a single V, D, and J segment. The act of rearranging these segments into functional TCRs is referred to as VDJ recombination. During VDJ recombination, the gene segments for the α and β chains of an individual T cell are rearranged in order to produce a new functional TCR. VDJ recombination employs both the lymphoid-specific RAG1-RAG2 complex, so named for recombination activating genes (RAG), and enzymes of the more broadly utilized double stranded break repair (DSBR) system [12, 13]. Antigen binding within the TCR V region involves the three complementarity determining regions (CDRs) that contact the antigen MHC complex. CDR1 and CDR2 are primarily encoded in the V germline segments and therefore experience less diversity. CDR3 however includes the junctional regions and is the primary region in contact with the antigen [5, 14]. The diversity of the TCR repertoire can change or evolve at different stages of the disease. If one can grasp the repertoire diversity, one can understand the immune response under various disease conditions, specifically infectious diseases, cancer, and autoimmune diseases.

In the past decades, in an attempt to examine the TCR repertoires, we relied on quantitative polymerase chain reaction (PCR) and spectratyping techniques to capture the TCR sequences. The limitations of these processes are the low resolution, favoring of the most dominant sequences, and the inability to pair the receptor expression. These constraints limit our ability to resolve the heterogeneity of TCR expression at high resolution. Currently, with the advancement of single-cell isolation processes and sequencing technologies, we are able to decipher the paired expression of TCRs at single-cell resolution. Understanding TCRs at the individual cell level will revolutionize how to profile immune cell repertoires, antigen discovery, disease prognosis, and treatment. The body of this review will focus on both the specific techniques and strategies used in the past and current single-cell technologies used to define and sequence TCR sequences, and how these techniques provide a better understanding of the immune cascade following antigen presentation in autoimmune diseases.

3.2 TCR Repertoires in Autoimmunity

3.2.1 TCR Signaling

TCRs recognize processed peptide antigen presented on MHC on the surface of the body's own cells. The two conventional MHCs, MHC I and MHC II are both polygenic and polymorphic noncovalent protein complexes composed of two polypeptide chains [4, 15]. TCRs are specific to both peptide antigen and the MHC to which it is bound, a phenomenon known as MHC restriction [4, 16]. MHC I is on the surface of virtually all nucleated cells in the body. Peptides presented on MHC I are generally 8–10 amino acids in length and result from the processing of foreign intracellular proteins [17, 18]. For this reason, MHC I is frequently used to signal viral infection to cytotoxic CD8 T cells. MHC II is only present on the surface of antigen presenting cells of the immune system including B cells, macrophages, and dendritic cells. MHC II presents peptides of 13–17, amino acids in length that have been collected from the extracellular environment [15]. MHC II is recognized by CD4 helper T cells which upon activation stimulate the effector arms of the immune system [19]. The T cell co-receptors CD4 and CD8 increase the T cell sensitivity to antigens and serve to jumpstart TCR signaling [1, 2]. T cell activation is considered to require three discrete signals delivered to T cells [3] (Fig. 3.2). Signal 1 occurs through the interaction of TCR and antigenic peptide presented on MHC. Association of T cell co-receptors CD4 or CD8 and MHC II or MHC I respectively are also considered part of signal 1. Signal 2 consists of a co-stimulatory signal commonly transmitted by B7.1 and B7.2 interacting with CD28 on the T cell surface [19, 20]. Signal 1 and signal 2 together drive T cell survival and clonal expansion. Signal 3 is delivered through the cytokine environment acting on the T cell, and determines

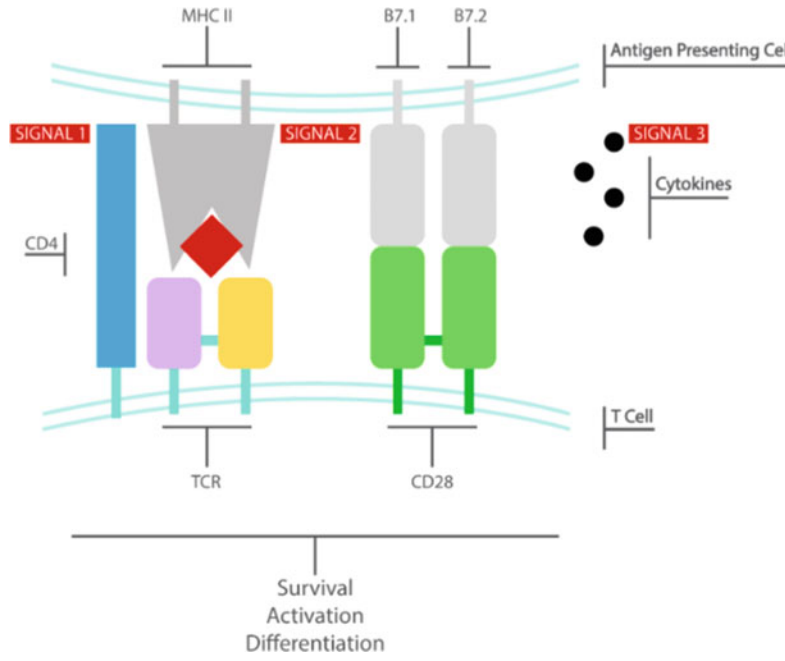


Fig. 3.2 Three discrete signals for T cell activation. Signal 1 occurs through the interaction of TCR and antigenic peptide presented on MHC. Signal 2 is indicated that consists of a co-stimulatory signal that is commonly transmitted by the B7.1 and B7.2 receptors of the antigen presenting cell that interact with CD28 on the T cell surface. Signal 3 that is delivered as a result of the cytokine environment acting that determines the differentiation of the T cell to the Th1, Th2, or Th17 subset

differentiation of CD4 T cells into T helper Th1, Th2, or Th17 subsets [19] which further in an immune response lead to different effector functions (Fig. 3.3).

3.2.2 Importance of Shaping the T cell Repertoire in Autoimmunity

While VDJ recombination allows for the development of an arsenal of TCRs capable of recognizing a nearly limitless array of foreign peptides and mounting an immune response, some of the receptors generated will be reactive against self-peptides. These self-reactive TCRs can erroneously engage the machinery of the immune system against the body's own tissues, resulting in tissue destruction and potentially an autoimmune disease [21]. T cells expressing TCRs that are strongly self-reactive are typically culled or inactivated through the mechanisms of central tolerance during T cell development in the thymus and peripheral tolerance after the T cells have exited the thymus and entered the secondary lymphoid tissues [21, 22]. However, the systems maintaining self-tolerance are imperfect and some self-reactive T cells will escape selection [23]. Additionally, some TCRs are specific for foreign peptides but also reactive against self-peptides [24]. Therefore, the immune system must balance the elimination of self-reactive T cells with the maintenance of a TCR repertoire capable of providing protection. During infection, when latently autoreactive T cells are exposed to self-peptide in an inflammatory environment, the mechanisms of self-tolerance can break down and permit the activation of autoreactive T cells [25, 26].

Autoreactive T cells are critically involved in many autoimmune diseases either as effector cells or through the promotion of autoantibody responses by providing T cell help to B cells [21]. Type I diabetes is characterized by the infiltration of leukocytes, particularly autoreactive CD8 T cells, into the pancreas [27] where the cytotoxic CD8 T cells kill the insulin producing beta cells, resulting in the development of disease [28]. Multiple sclerosis is another T cell mediated autoimmune disease where Th17 cells, subsets of

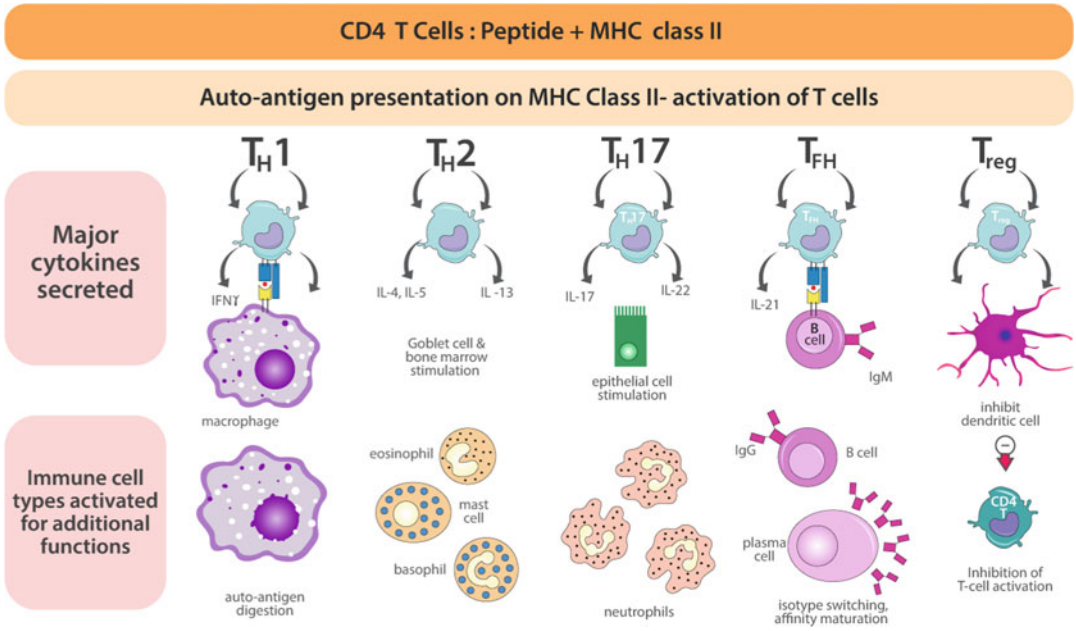


Fig. 3.3 Effector functions of CD4 T Cells. The differentiation of CD4 T cells into Th1, Th2, or Th17 subsets which further lead to an immune response and have effector functions. Different T cell subtypes are associated with different autoimmune diseases

CD4 T cells, propel an inflammatory response that results in the destruction of the myelin sheath around nerve axons [29–31]. During the pathogenesis of psoriasis, multiple autoreactive T cell subtypes including Th1 and Th17 cells are activated and contribute towards the development of inflammatory skin disease [32–34]. For many autoimmune diseases, the antigens capable of activating autoreactive T cells have yet to be defined. Discovery of the sequences of TCRs of clonally activated T cells in autoimmune diseases presents an opportunity to elucidate the antigen-specific response of these clones and gain a better understanding of the determinants of autoimmune disease.

3.3 Bulk Sequencing Technologies

Multiple techniques of immune repertoire analysis have been developed over the last decade. Pioneering studies on TCR repertoires were initially based on the cloning and sequencing of PCR-amplified cDNA collections [35–40]. The mRNA gene expression level of TCR variable gene segments in bulk is analyzed by techniques such as semi-quantitative PCR using TCR Vβ family specific primers [41–44]. Bulk sequencing and analysis do not capture the expression of single cells, but rather capture bulk samples or many cells within the samples. Despite being comprehensive, this qualitative and quantitative technique does not provide any information about CDR3 region of TCRs which forms an integral component of the receptor [4]. The following discussion will detail some of the major techniques for bulk TCR analysis.

3.3.1 CDR3 Sequence Analysis by Immunoscope

Immunoscope, also known as CDR3 spectratyping, was the first qualitative method at the molecular genetic level for analyzing CDR3 polymorphisms and sequence length diversity [9, 45–47]. This method based on electrophoretic analysis of CDR3-PCR fragment length has been

the standard technique of TCR evaluation since the 1990s [39, 48]. Musette et al. found an expansion of oligoclonal Vβ5.3 + T cells population in HLA-DR2 multiple sclerosis (MS) patients [49]. Ercolini and Miller characterized the cross-reactive anti-self response induced by a peptide mimic on a model for MS [50]. Immunoscope analysis of TCRVβ repertoire diversity in patients with autoimmune lymphoproliferative syndrome showed an expansion of TCRα/β^{+}CD4^{-}CD8^{-} double negative T cells derived extra-thymically from CD8^{+} T cells [51]. Immunoscope analysis has technical limitations and lacks the key parameters of capturing TCR diversity such as specific nucleotide sequences, codon usage, and amino acid composition [49]. In parallel to spectratyping, development of real-time PCRs opened the possibility to evaluate TCR repertoires with higher precision [39, 50–54]. Pairing immunoscope with other techniques allowed capture of quantitative and qualitative data for the repertoire of V domain and C region combinations [55–57]. Current approaches aim to quantify genomic DNA (gDNA) or mRNA sequences corresponding to CDR3 region of TCR β chain, the most variable TCR region [58, 59]. CDR3 sequences are usually unique markers and the key determinants of antigenic recognition, quantitative T cell responses, and clonal composition of the T cell repertoire [60]. However, CDR1 and CDR2 are also associated with MHC restriction [61, 62] and therefore sequencing of all regions is important to provide the complete information about TCR structure and its binding characteristics [63]. Selecting between gDNA and mRNA is an important consideration for sequencing. The gDNA presents high long-term stability, ease of isolation, and presence of a single DNA template per T cell. The disadvantages include possible errors of sequencing due to introns, cells with residues of VDJ rearrangements and interfering priming sites of the sample. Contrary to the presence of multiple TCR transcripts in gDNA, using mRNA provides the sequence of the entire V and J gene [51, 59, 63] which is the basis of specific antigen recognition. Lim et al. combined immunoscope with MHC peptides multimer-based T cell sorting; however, specific T cells could not be detected due to their low frequency and/or an insufficient amount of TCR β chain transcript [57, 64]. For the first time at the nucleotide sequence level, the TCR repertoire was analyzed using molecular cloning and Sanger sequencing [9, 65, 66]. Nonetheless these low-throughput approaches were limited due to their failure to individually sequence a large number of T cell clones. Over the years, there have been several methodologies taken into consideration for successful and correct sequencing analyses. With technological interventions, it has become possible to amplify both TCR α and β chains simultaneously. Fundamentally, the β chain is more preferred for analysis due to the unique presence of the D gene component [63].

3.3.2 High-Throughput Sequencing Techniques

3.3.2.1 Multiple-Based PCR Methods

All previous techniques presented several limitations to create an extensive analysis of TCR repertoire profiling until the development of high-throughput sequencing (HTS) techniques. These allowed for massive parallel sequencing of millions of TCR sequences. The application of HTS presented promising results in clinical needs pertaining to infectious diseases, cancer, and specially autoimmunity [60, 67, 68]. There are three principal PCR-based enrichment techniques that are used: the multiplex PCR, target enrichment PCR, and 5′RACE-switch-oligo nested PCR. The multiplex technique allows working with both starting materials: gDNA and RNA and contains a mix of primers complementary to all known V segments and a pool of reverse primers for J segment or constant region exons [69, 70]. There are several kits that are available on the market, for example: BGI, Adaptive Biotechnologies, iRepertoire [71–73] that facilitate the multiplex PCR technique. Target enrichment-based PCRs also work with both gDNA and RNA. TCR-specific baits complementary to α and β transcripts hybridize with molecules in the library, are conjugated to

magnetic beads and then undergo amplification before sequencing. Kits such as SMARTer ThruPLEX have been developed by Takara Bio. Lastly the nested PCR approach is an alternative technology that has been introduced into the bulk sequencing and is based on the 5′RACE switch-oligo approach designed only for RNA. It involves presenting an insert of an adaptor molecule at the 5′ end of the cDNA during cDNA synthesis. This technique was developed by Clontech and follows the principle of a PCR where the forward primer binds to 5′ adaptor molecule and reverse primer binds to the C region of transcript improving specificity [74]. Hendriksen et al. analyzed T cell repertoires of paired liver and gut samples from patients suffering from primary sclerosing cholangitis with inflammatory bowel disease, and demonstrated their common origin [78]. Spreafico et al. discovered for the first time circulating pathogenic-like lymphocytes in patients with juvenile and rheumatoid arthritis TCR. These cells are enriched in synovial clonotypes, circulate through the site of autoimmune reaction, and correlated with progression of juvenile and adult autoimmune arthritis [79]. Expanded findings are summarized in Table 3.1.

3.3.2.2 Commercial Platforms for Next Gen Sequencing

Next Gen Sequencing (NGS) consists of current technologies that have higher accuracy as compared to Sanger High-Throughput Sequencing (HTS) platforms. The NGS platforms have refined technologies over time and as a result there are several available currently. Technologies such as GS FLX by 454 Life Sciences/Roche diagnostics, Genome Analyzer, HiSeq, MiSeq and NextSeq by Illumina, Inc., SOLiD by ABI, Ion Torrent by Life Technologies, Oxford Nanopore, Complete Genomics by Beijing Genomics Institute, and GnuBIO are a few platforms that have different characteristics of sequencing accuracy. The advantages they provide over HTS differ based on sequencing depth, read length, error type and their frequency but include individual TCR clonotype quantification [58, 75]. Since error rate is a major consideration, there is a trade-off in choosing a sequencing platform needed to correctly manage it. Library preparation for companies such as Illumina sequencing, Roche, Life Technologies, and Pacific Biosciences are almost the same with differences in parameters such as sequencing adapters, cluster generation

Table 3.1 Application of bulk sequencing technologies in TCR repertoire analysis for autoimmune diseases

Paper	Methods	Results
Musette et al. [49]	Immunoscope	Expansion of oligoclonal Vβ5.3 + T cells population in HLA-DR2 MS patients
Ecrolini and Miller [50]	Immunoscope	Characterization of the cross-reactive anti-self response induced by a peptide mimic on the model for MS
Bristeau-Leprince et al. [51]	Immunoscope	Analysis of TCRVβ repertoire diversity in patients with autoimmune lymphoproliferative syndrome demonstrated an expansion of TCRα/β^{+}CD4^{-}CD8^{-} double negative T cells derived extra-thymically from CD8^{+} T cells
De Palma et al. [87]	Immunoscope	Patients affected by SS
Henriksen et al. [78]	Multiplex PCR, immunoSEQ assay (Adaptive Biotechnologies)	Analysis of T cell repertoires of paired liver and gut samples showed common origin in patients with PSC-IBD
Oftedal et al. [88]	Multiplex PCR, ImmunoSEQ assay	Specific TCRβ sequences with limited diversity recognize self-antigen myeloperoxidase in autoimmune Aire$^{-/-}$ mice, the utilization of V and J genes are altered due to autoimmunity and immunization
Spreafico et al. [79]	Multiplex PCR, immunoSEQ assay	TCR repertoire of CPLs is enriched in synovial clonotypes and correlated with progression of juvenile and adult autoimmune arthritis

MS multiple sclerosis, *PSC-IBD* primary sclerosing cholangitis-inflammatory bowel disease, *SS* systemic sclerosis, *AIRE* autoimmune regulator, *CPLs* circulating pathogenic-like lymphocyte

and application of technologies such as paired end sequencing. The basic process includes RNA extraction and fragmentation from the sample and conversion into cDNA by reverse transcription. DNA fragments are ligated to sequencing adapters with specific sequences [76]. These sequences are designed to interact with a specific platform either the surface of an individual flow cell in case of Illumina or the microemulsion-based platform for Ion Torrent. Library construction involves three primary steps: (1) fragmentation of starting material DNA or RNA, (2) adaptor ligation, and (3) amplification [81, 82]. Initial DNA can be sheared into short fragments of 200–400 bp (depending on the sequencing platform) by physical methods (acoustic sonication and hydrodynamic shearing), enzymatic (endonuclease and transposase) or chemical methods (heat digestion with divalent metal cations) [81, 83]. Random/unbiased fragmentation allowing to produce overlapping segments of optimal length is one of successful factors for NGS library construction. After fragmentation end repair is required, single stranded segment ends are blunted and 5′ phosphorylated, following A-tailing by Taq or a Klenow fragment. The sequencing adaptors are then ligated to the DNA fragments forming the fragment library. Several cycles of PCR are performed to enrich for the library product.. Finally, the DNA library can be evaluated for quality control and followed by sequencing through NGS [81, 84].

3.3.2.3 The Illumina Platform and Technologies

The Illumina platform presents the leading position in the NGS market and it provides sequencing by synthesis based on reversible dye-terminators. Illumina HiSeq and MiSeq use four-channel sequencing systems with individual image detection of each base. The NextSeq 500 presents a two-sequencing system wherein only two images are needed for four base calls that allow reduction in time, cost of sequencing, and number of cycles as there are two images simultaneously analyzed from both ends [77]. Illumina MiSeq is widely used for capturing the most common and expanded clonotypes, whereas Illumina HiSeq is mostly applied for increasing the depth in sequencing [59, 63]. The principal drawback of bulk sequencing is that it provides information about the frequency of single TCR chains, but no information regarding their pairing [63]. To remove this bottleneck, unique molecular identifiers (UMIs) have been introduced during cDNA synthesis that allow for reduction in errors during amplification during a PCR [74]. Using high-throughput deep TCRβ chain sequencing, Muraro et al. showed that autologous stem cell transplantation has distinctive effects on $CD4^+$ and $CD8^+$ T cell repertoires in multiple sclerosis patients [78]. Incorporating multiple-PCR and 5′ rapid-amplification of cDNA ends (RACE) to capture CDR3 regions and high-throughput sequencing platform HiSeq2000, the authors found no significant difference in TCR repertoires between pancreatic cancer patients versus healthy controls, and no differences were found between the samples of tumor tissue and the blood samples from patients. Applying a template-switch anchored RT-PCR and Illumina sequencing, diverse repertoire of clonally expanded tumor-reactive lymphocytes was identified on $CD8^+$ melanoma tumor-infiltrating lymphocytes (TILs). Additionally, sequencing the TCRβ revealed that tumor-reactive and mutation-specific clonotypes were highly expanded in the $CD8^+$ population and preferentially expanded in the PD-1+ population [79–81].

3.4 Single-Cell Sequencing Techniques

As opposed to the bulk cell analyses that provide coarse or broader resolution, the TCR repertoires can be analyzed at a second level that is more fine-tuned at higher resolution with single-cell sequencing analysis. This allows us to visualize the complex interplay between varied cell types of the immune system bringing about a systemic

response. Bulk technologies fail mostly in understanding the nuances of the complex cell responses while single cells help in resolving ambiguity as analyzing the interactions microscopically between individual cells provides details giving insight to bigger reactions [82]. As indicated, bulk sequencing procedures are only able to catch the most common and expanded clonotypes which may or may not be ideally suggestive of the immune responses during an infection or autoimmune process [83]. Single-cell analyses on the other hand provide information indicative about rare TCR subset sequence repertoires that will yield a comprehensive understanding of the biological processes of individual T cells and their dynamic interaction within the biological system.

As discussed, initially platforms focused primarily on providing sequencing information on the dominant TCR α and β chains, but current technologies can sequence the rarer and less frequent cell populations such as TCR of γδ T cells [84]. In T cell repertoire formation, it is essential to understand the development and expansion of the cells based on TCR expression. The differentiation of a progenitor T cell into a specific subset can be identified by single-cell RNA (scRNA) sequencing as there is clonal expansion of a specific prototype [84, 85]. Since single-cell sequencing helps in identification of the TCR chains at a cellular level, depth becomes an important consideration as it provides vital information regarding repertoire formation [86, 87]. Major variability can be observed between two similar cells during the analysis of single-cell-based data. This variability may be attributed to either inherent biological variations or a result of the technique for repertoire analysis. Single-cell technologies for TCR repertoire analysis include PCR that involves the reverse transcription of the mRNA to amplify the cDNA from a single cell. The goal in single-cell receptor sequencing is to obtain maximum coverage of cell types and a low cost per cell, with comprehensive algorithms and platforms for analysis that provide extensive detail [88]. The challenges include the capture of single cells, isolation and amplification of sufficient materials, and generation of vast volumes of data that require faster methods of computation and algorithmic tools that allow efficient data management.

3.4.1 Microfluidic Technologies and Platforms

One of the challenges of single-cell sequencing is the capture of the single cells. Major technological improvement in the area of microfluidics has facilitated the capturing process. Microfluidic technology involves capturing of individual or population of cells in emulsion droplets in the picoliter or nanoliter range that is pumped in by different speeds using microfluidic devices [89]. The cell droplets constitution is compatible with mammalian cells and as a result it keeps cells functional for further investigation [90]. Cells are barcoded via specific primers and cDNAs are generated from thousands of cells in parallel [91, 92]. The cDNAs formed are pooled by breakage of the microfluid droplet, amplified by transcription and then sequencing libraries are prepared [93]. Microfluidic technologies are currently being used widely in the space of TCR enrichment sequencing [94]. The advantages that the process provides consist of providing T cell activation information, pathogens targeted by cytotoxic T cells and cytokine release [95–97]. Not only does it provide real-time information about the T cell repertoire, but it also provides the functional monitoring of the individual T cell types as opposed to the conventional methods [70, 98] such as identifying cell populations via FACS, carrying out PCR and then mass population sequencing that leads to loss of diversity [98, 99]. Chokkalingam et al. have used droplet technology combined with flow cytometry to identify T cell populations that secrete IL-2, IFNγ, and TNF-α [100]. Konry et al. have characterized T cells, natural killer cells, and antigen presenting complexes by multiplexing different parameters that are found in a single droplet. McDaniel et al. have proven that the pairing of both TCR chains (αβ or γδ) can

be kept intact by single-cell sequencing by carrying out high-throughput sequencing on millions of T cell populations by obtaining individual chain sequences [101]. Further Segaliny et al. demonstrated the dynamics of TCR T cell interactions and downstream molecular analysis using single-cell sequences that confirmed their matching with target antigens with an exact specificity [102].

Droplet-based microfluidics has led to rapid discovery and has provided new tools for learning single-cell biology that include being a high-throughput technology [89, 91], providing insight to sequencing [103–108], secretome analysis [95, 96, 109], cell–cell interactions [97, 110], protein and antibody engineering [111], repertoire analysis and multiplex biological interactions [112]. Furthermore, microfluidics allows for analysis of drug interactions [113], antigen-specific T cell repertoires [97, 104, 114], CAR-T cell populations [115] and facilitates investigation of other protein secretion, e.g., antibodies [111, 116–118]. Microfluidics uses in-situ PCR and can display libraries in the same run simultaneously [119–121]. Platforms like the Chromium Single Cell Immune Profiling Solution by 10× Genomics can provide pertinent information on TCR, BCR, cell surface protein expression, and RNA expression in the same cells [122]. In addition, the small volume provides increased sensitivity, decreases time and it improves the signal to noise ratio [123]. As a result, it provides an insight into complex cellular heterogeneity instead of an average readout and provides information on molecular details that are essential for T cell activation. Not only can real-time sequencing be done with control of space time and environmental factors but it does not compromise the screening sensitivity [123]. There are multiple studies that have been conducted to understand multiplex biological responses, CDR3 CAR-T cell products, and adoptive T cells that can be used in varied therapeutic platforms, and provide real-time data for analysis by microfluidics [70].

3.4.2 Generation of TCR Sequences Using scRNA Sequencing

Single-cell TCR sequencing typically requires a distinct four step process. First, the isolation of single cells. This process is often performed using FACS sorting to isolate the target single cells. As discussed, microfluidic devices can be used to sequester single cells [86, 124]. Secondly, cells are isolated and the RNA is released into a fluid medium post lysis of the cell and Oligo dt primers recognize polyadenylated regions which are amplified as a result of selective reverse transcription. In the third step, the complimentary DNA molecule that is obtained from this process is amplified to produce multiple replicons, and then sequenced. Lastly, the output sequence reads from the reactions are analyzed and quantified [125]. The reference genome and transcriptome of interest can be traced back by comparing the sequence read subset to a reference genome or transcriptome [86, 124–126]. A shift has been observed in the single-cell sequencing landscape, as previously used capillary-based techniques are being replaced by emulsion-based PCR techniques. The methodology is improved as devices use technology where oil-in-water emulsions are pumped and high sequence read repeats are analyzed for individual cells. This provides a detailed insight on the αβ TCR chains as the mRNA coding them is released into the individual droplet that acts as an individual reaction platform. Within the droplet, the mRNA is amplified using a multiplex PCR-based approach [70] (Fig. 3.4).

3.4.3 Single-Cell Sequencing Data

Multiplex PCR systems and high-throughput systems have been used in the past for analyzing the TCR sequences. The protocol as described above uses isolation, amplification, library preparation, and sequencing for TCR α and β chains [79, 127]. There are two fundamental strategies:

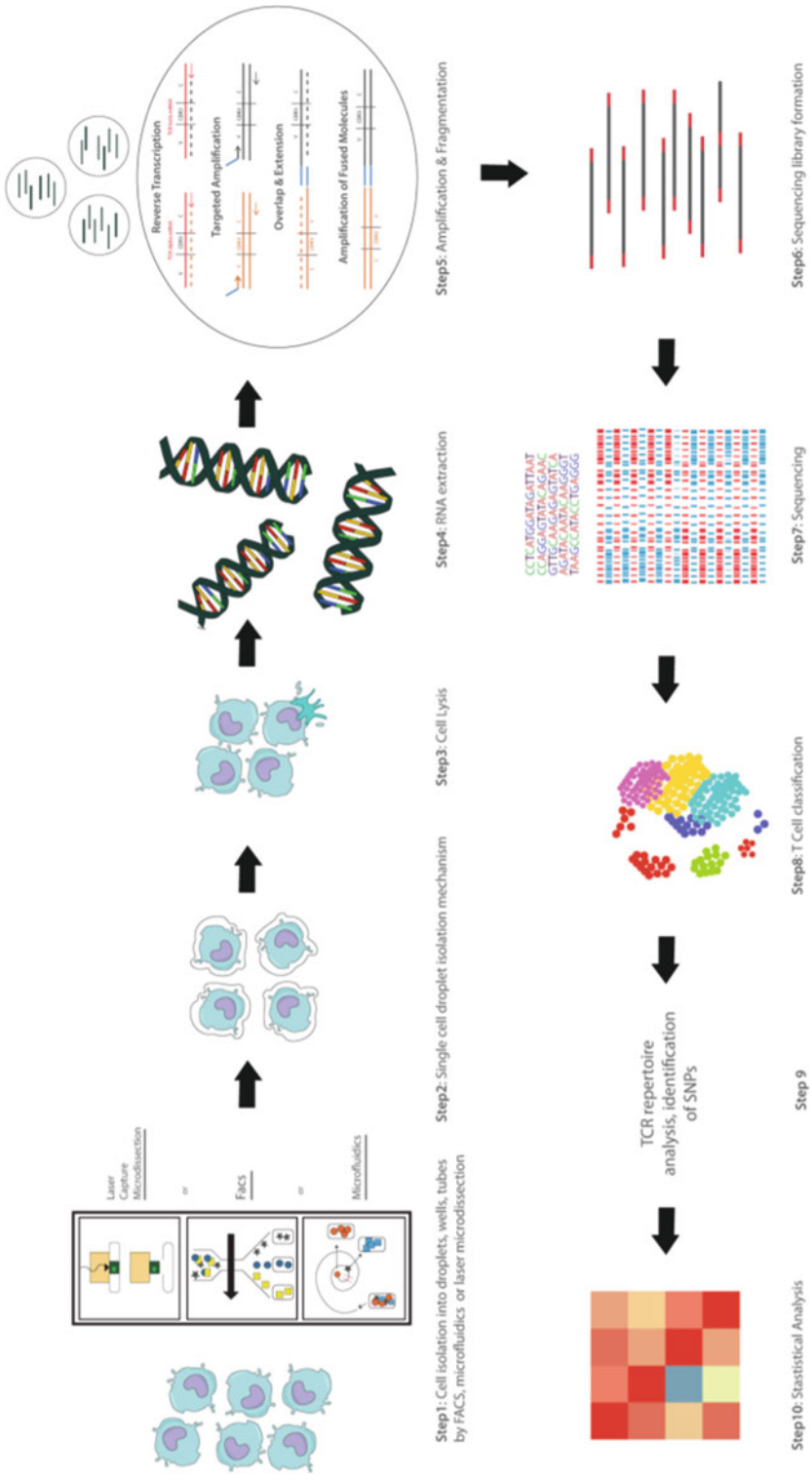


Fig. 3.4 Representative experimental process for single-cell sequencing of TCRs

barcoding and full-length complimentary DNA sequencing. PCR has been used in the process of single-cell barcoding, a method of identification of cell subtypes. A barcode is a unique nucleotide sequence obtained from a single cell that tags cell transcripts and allows tracing the RNA back to its origin. This once linked to the TCR can help in identifying individual T cell types and subsets with distinct functions [128]. Full-length sequencing on the other hand involves pooling the cDNA obtained from each cell and then making a single-cell library for comparison of the sequence. This process is time consuming and expensive but sensitive and can provide information about changes in individual base pairs, mutations in the CDR3 region, difference in isoforms, and differences in α and β chain pairing [98]. The SMART-seq protocol that was developed by Stubbington was one of the first tools developed to reconstruct paired TCR α and β chains. Named TraCer, it was used to validate the sequences of the TCR αβ chains. The same cDNA was used for sequencing libraries and this validated the experimental approach as the start was the same PCR-based approach and the result a comparison for the chain sequences. This procedure has given an impetus to the requirement of efficiency in the approach [98]. TRAPes was another tool that was created to read short-read single-cell RNA-sequence libraries [129]. Other methods such as the non-strand specific poly-A tailing described by Tang et al. showed a weak positional bias for the 3′ end of the RNA [105]. The CEL-seq and MARS-seq that are similar techniques are also based on the principle of in-vitro transcription and having the same strong 3′ positional bias [130]. Quartz-seq a platform based on Poly-A tailing is able to provide low depth sequencing accurate results as it is one of the few that use a non-specific 3′ weak positional bias [131]. Further, platforms like the VDJ puzzle allow the analysis of T cell gene expression and TCR diversity, which consist of a library formation and then comparison of the sequences to the antigen-specific circulating CD8 T cells for additional validation [132].

3.4.4 An Application of Single-Cell TCR Sequencing

It has always been the goal of TCR research to identify the T cells at the genetic and biological levels simultaneously; it is an ongoing quest to understand the individuality of pathogenic or protective T cells in diseases. The challenges often encountered include the insufficiently small size of the human patients' biopsies and the near impossibility of simultaneously examining biological functions of individual live cells isolated from the glands and their genetic profiles. Numerous attempts have been made using techniques such as laser capture micro-dissection or flow cytometric sorting to obtain single-cell populations. However, both techniques are cumbersome and neither can satisfactorily ensure single-cell analysis. Furthermore, the integrity of such isolated cells is greatly compromised due to the vigorous isolation process, which can affect subsequent micromanipulation and even functions of infiltrating T cells. Due to the fact that the purified cells are often either not viable or damaged, important information regarding the biological function(s) is difficult or impossible to obtain. To circumvent these challenges, we utilize the single-cell microengraving technology which was developed originally by the Love group at MIT as a tool for identification, profiling, and selection of single cells, specifically based on their reactivity to or secretion of proteins [118, 133–135]. The nanochip is fabricated via soft lithography, where arrays of nanowells are manufactured onto a glass slide consisting of 24 × 72 blocks containing 7 × 7 arranged nanowells within each block, yielding 84,672 50 μm nanowells per array. Additionally, these can be manufactured to produce 30 μm nanowells, where the arrays consist of 248,832 nanowells. Sylgard 184 silicone elastomer base (polydimethyl-siloxane, PDMS) is cured in a custom-built aluminum mold to produce these microchips. After cells are stained for desired cell surface markers with fluorescently conjugated antibodies and/or calcein (a live/dead

marker), the cells are distributed on the microchip. As shown in Fig. 3.5, they can be analyzed by fluorescent microscopy, where images are taken of the 7 × 7 grid of nanowells to determine the identity of the cells. Next, the microchip is hybridized with a capture slide. This is a poly-L-Lysine slide coated with a desired protein or an antibody against a desired protein. For example, for identification of T helper cells, one would coat the slide with anti-IFNγ and anti-IL-17. In combination with the data from the fluorescent microscopy, identification of Th1, Th2, and Th17 cells would be possible. Alternatively, to isolate reactive antibodies, one would coat the capture slide with proteins of interest in order to identify B cell reactivity. Notably, in the case of identifying cells based on secreted proteins or antibodies, single-cell microengraving is advantageous over comparable techniques, such as flow, because the cells remain active during the process and do not need to be permeabilized, such as in the case of flow cytometry. After hybridization, the capture slide is exposed to detection (fluorescent) antibodies and/or proteins. The capture slide can be analyzed on a microarray scanner and the array data is then aligned against the fluorescent microscopy data to give a complete profile of the cells.

One notable application of this approach is the area of Sjogren's syndrome (SjS) in which we sought to determine the TCR repertoires of pathogenic T cells that infiltrate the salivary glands, the main targeted tissue in SjS. Other groups have been trying to sequence pathogenic T cells for well over two decades utilizing a variety of techniques [136–141]. Commonly, the final technique involves nested PCRs to identify the TCR sequences. Single-cell sorting by flow is the most common technique to isolate T cells prior to this [142–145]. As previously mentioned, the pitfall of this technique is the lack of specific subset identification of the T cells. The one measure to counteract this is to permeabilize the membrane, killing the cells. Currently, there is no way to identify intracellular antigens while keeping the cells alive. Recently, single-cell microengraving has been successfully utilized in this effort [146]. As outlined in Fig. 3.6, cells presenting the correct phenotypes were selected, e.g., live Th1, Th17 cells, from the biopsies of SjS patients and sicca controls (patients who present a similar disease phenotype, but who do not meet the criteria of SjS). After the cells were lysed to retrieve RNA, RT-PCR was performed, immediately followed by nested PCR for the alpha and beta TCR chains and those products were sequenced. These sequences were then aligned against databases for identification of VDJ regions and compared between the subject and control groups as well as the subset of cells to identify unique subsets presenting specific TCRs, implying a shared antigen. Our recent study has identified that glandular Th1 and Th17 cells of control and SjS patients expressed common TCRβ variable (TRBV)3-1 and TRBV20, whereas TCRα variable (TRAV)8-2 was uniquely expressed by Th1 of SjS patients [146]. Using the SjS animal model, we have shown that salivary Th1 cells of male mice selected for TRAV8 and TRBV16 in Th1 and Th17 cells, whereas female Th1 cells selected for TRAV8, TRAV13D-2, and TRBV23 [147]. Other studies attest to our findings by ascertaining unique glandular TCRs in the human and animal models of SjS [145, 148]. Our seminal studies clearly imply that the clonal expansion of the effector T cells with the conserved TCRs is driven by SG cell antigens, and autoimmune responses to SG cell autoantigens evidence a specific loss of immunological self-tolerance.

3.4.5 Advantages and Limitations of Single-Cell Sequencing

A variety of bulk sequencing measurements give insight about large, mixed populations of cells. ELISA, PCR, and transcriptional profiling provide information about rapidly multiplying cells [149]. The bulk measurements stated above do not provide input on unique subsets of cells such as clonally expanded B cells and T cells. The parameters tested in bulk measurements average out phenotypic states of cells and give a generic outlook of the basic interactions among cells [150]. TCR profiling includes analyzing the

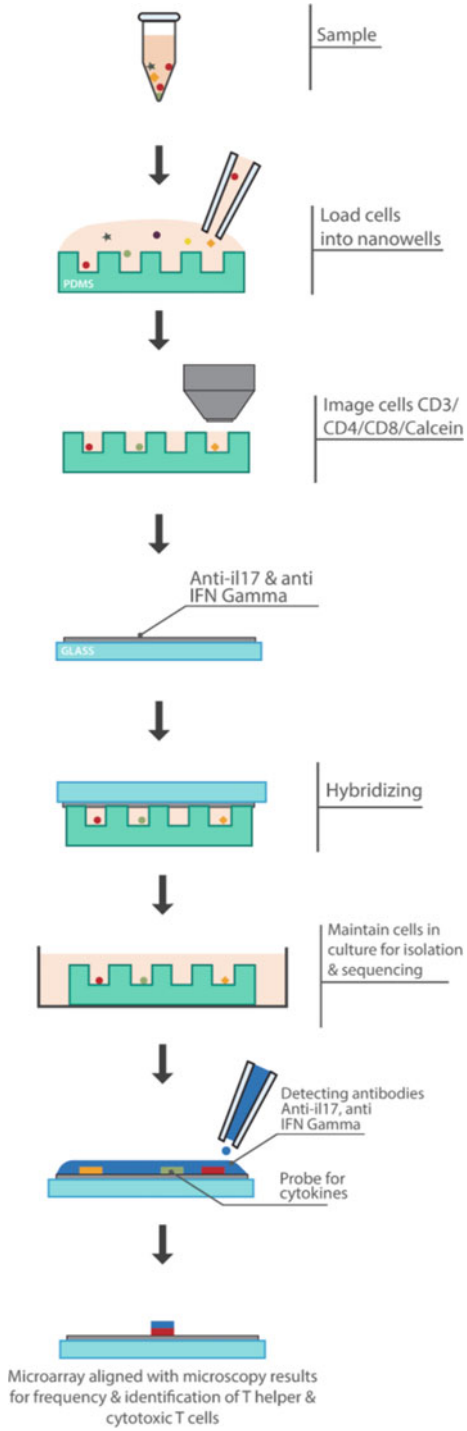


Fig. 3.5 Single-cell microengraving technology. The fabricated nanowells were loaded with lymphocytes obtained from biopsies and were imaged using an automated epifluorescence microscope to cell populations based on surface markers. Hybridized nanowells with capture slides containing anti-human immunoglobulins against IL-17 and IFNγ for Th17 and Th1 cells as an example were analyzed. Post incubation, nanowells containing intact live cells and capture slides were separated and antibody-specific fluorochrome-conjugated antibodies were used detection antibodies. Lastly, micrographs of microarrays were generated by scanning using a GenePix Autoloader 4200AL microarray scanner

transcriptome that is generated by the RNA. While offering many advantages, there are gaps in information which are unaccounted for in single-cell sequencing techniques. Some include accountability for stress conditions in the surrounding environment that affect individual cells. Cells may lie in varied cycle phases and each individual unit may display a variation in gene expression, size, and RNA content. Furthermore, every run in the single-cell sequencing process requires a large cell population harvested from fresh material, and it is expensive and covers limited cell populations [151]. Hence, despite the accuracy the variation that arises is mostly a consequence of all these factors. This aspect is not addressed clearly in most commercially available platforms available for TCR analysis [88, 92]. To combat these issues, there is a requirement of unprejudiced and reproducible gene expression patterns in single cells [152]. Seeing the varied factors that influence TCR analysis, there are different computational approaches that use variable models to account for variability in results and provide a better insight to the data obtained [92].

3.5 Conclusion

Significant technological advancement has catapulted the field of single-cell TCR sequencing. The remaining unanswered quests are how fast the field of bioinformatics will evolve to analyze and manage the large data generated. Recent interests in systemic biology of single T cells involve identification of the T cell subsets based on surface markers, gene expression at the

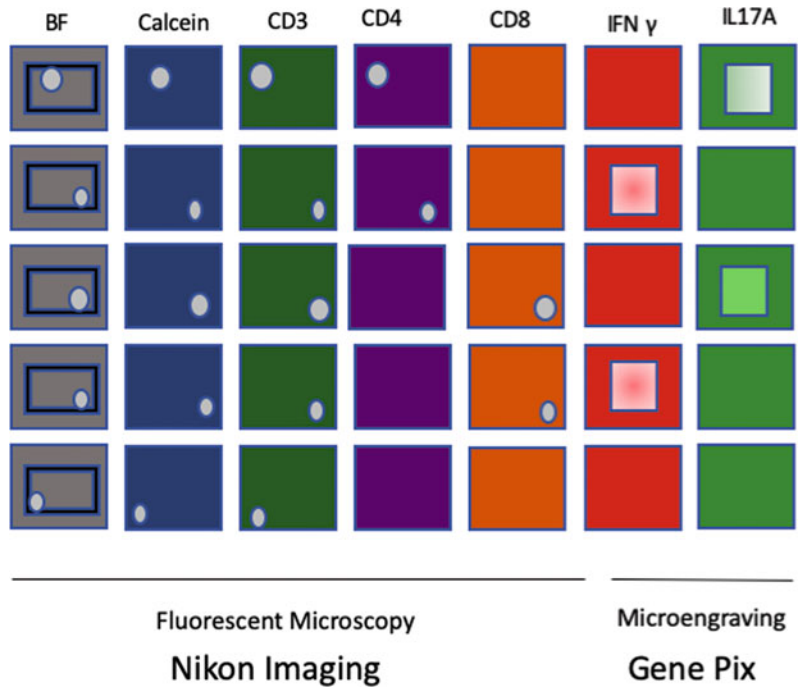


Fig. 3.6 An example of T helper cell profiling using single-cell microengraving technology. Representation of cells (bright field [BF]) as observed in the process of microengraving in nanowells labeled with calcein (indicating live cells). Micrographs of matching microarray by microengraving show detection signals for CD3, CD4, CD8, IFNγ, and IL-17A responsive cells as represented in the demonstration

single-cell transcriptome, and single-cell TCR expression. Our ability to comprehend and organize this set of new data will be limited if we are not able to organize the data in a biologically meaningful way. The second aspect that we must grapple with is its usefulness in antigen discovery. The current approach in examining and testing the antigen specificity of individual T cell clone is labor intensive using vast peptide-MHC libraries under culture conditions mimicking the in-vivo environment. In order to improve diagnosis and define therapy, specifically personalized medicine, it is imperative that we can identify and test the antigen specificity based on the single-cell TCRs. Emerging single-cell technologies bring promising discoveries but there are challenges that need to be addressed in order to fully comprehend the TCR repertoires and its application.

Acknowledgements CQN was supported financially in part by PHS grants DE023433, DE018958, DE028544 from the National Institutes of Health. RW is supported by the T90/R90 Comprehensive Training Program in Oral Biology (NIDCR, T90 DE21990). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Chothia C, Boswell DR, Lesk AM (1988) The outline structure of the T-cell alpha beta receptor. EMBO J 7:3745–3755
2. Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. Nature 334:395
3. Smith-Garvin JE, Koretzky GA, Jordan MS (2009) T cell activation. Annu Rev Immunol 27:591–619. https://doi.org/10.1146/annurev.immunol.021908.132706
4. Rudolph MG, Stanfield RL, Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol 24:419–466. https://doi.org/10.1146/annurev.immunol.23.021704.115658
5. Claverie J, Prochnickachalufour A, Bougueleret L (1989) Implications of a Fab-like structure for the T-cell receptor. Immunol Today 10:10–14. https://doi.org/10.1016/0167-5699(89)90058-3
6. Kaufmann SH (1996) gamma/delta and other unconventional T lymphocytes: what do they see and what

do they do? Proc Natl Acad Sci U S A 93:2272–2279. https://doi.org/10.1073/pnas.93.6.2272
7. Hayday AC (2000) [gamma][delta] cells: a right time and a right place for a conserved third way of protection. Annu Rev Immunol 18:975–1026. https://doi.org/10.1146/annurev.immunol.18.1.975
8. Holtmeier W, Kabelitz D (2005) gammadelta T cells link innate and adaptive immune responses. Chem Immunol Allerg 86:151–183. https://doi.org/10.1159/000086659
9. Arstila TP et al (1999) A direct estimate of the human alphabeta T cell receptor diversity. Science 286:958–961. https://doi.org/10.1126/science.286.5441.958
10. van Dongen JJ, Wolvers-Tettero IL (1991) Analysis of immunoglobulin and T cell receptor genes. Part I: Basic and technical aspects. Clin Chim Acta 198:1–91. https://doi.org/10.1016/0009-8981(91)90246-9
11. Pannetier C et al (1993) The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. Proc Natl Acad Sci U S A 90:4319–4323. https://doi.org/10.1073/pnas.90.9.4319
12. Cabaniols JP, Fazilleau N, Casrouge A, Kourilsky P, Kanellopoulos JM (2001) Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. J Exp Med 194:1385–1390. https://doi.org/10.1084/jem.194.9.1385
13. Sadofsky MJ (2001) The RAG proteins in V(D)J recombination: more than just a nuclease. Nucleic Acids Res 29:1399–1409. https://doi.org/10.1093/nar/29.7.1399
14. Borg NA et al (2005) The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. Nat Immunol 6:171–180. https://doi.org/10.1038/ni1155
15. Stern LJ, Wiley DC (1994) Antigenic peptide binding by class I and class II histocompatibility proteins. Structure 2:245–251
16. Sim BC, Zerva L, Greene MI, Gascoigne NR (1996) Control of MHC restriction by TCR Valpha CDR1 and CDR2. Science 273:963–966. https://doi.org/10.1126/science.273.5277.963
17. Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA (1992) Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. Science 257:919–927. https://doi.org/10.1126/science.1323877
18. Madden DR, Garboczi DN, Wiley DC (1993) The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. Cell 75:693–708. https://doi.org/10.1016/0092-8674(93)90490-h
19. Janeway CA, Bottomly K (1994) Signals and signs for lymphocyte responses. Cell 76:275–285. https://doi.org/10.1016/0092-8674(94)90335-2
20. June CH, Bluestone JA, Nadler LM, Thompson CB (1994) The B7 and CD28 receptor families. Immunol Today 15:321–331. https://doi.org/10.1016/0167-5699(94)90080-9
21. Skapenko A, Leipe J, Lipsky PE, Schulze-Koops H (2005) The role of the T cell in autoimmune inflammation. Arthritis Res Ther 7:S4. https://doi.org/10.1186/ar1703
22. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M (1995) Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol 155:1151–1164
23. Bouneaud C, Kourilsky P, Bousso P (2000) Impact of negative selection on the T cell repertoire reactive to a self-peptide: a large fraction of T cell clones escapes clonal deletion. Immunity 13:829–840. https://doi.org/10.1016/S1074-7613(00)00080-7
24. Liblau R et al (1991) T cell response to myelin basic protein epitopes in multiple sclerosis patients and healthy subjects. Eur J Immunol 21:1391–1395. https://doi.org/10.1002/eji.1830210610
25. Rosenblum MD, Remedios KA, Abbas AK (2015) Mechanisms of human autoimmunity. J Clin Invest 125:2228–2233. https://doi.org/10.1172/JCI78088
26. Mills KH (2011) TLR-dependent T cell activation in autoimmunity. Nat Rev Immunol 11:807–822. https://doi.org/10.1038/nri3095
27. Itoh N et al (1993) Mononuclear cell infiltration and its relation to the expression of major histocompatibility complex antigens and adhesion molecules in pancreas biopsy specimens from newly diagnosed insulin-dependent diabetes mellitus patients. J Clin Invest 92:2313–2322. https://doi.org/10.1172/JCI116835
28. Varela-Calvino R, Calviño-Sampedro C, Gómez-Touriño I, Cordero OJ (2017) Apportioning blame: autoreactive CD4. Arch Immunol Ther Exp 65:275–284. https://doi.org/10.1007/s00005-016-0452-4
29. Tzartos JS et al (2008) Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. Am J Pathol 172:146–155. https://doi.org/10.2353/ajpath.2008.070690
30. Kebir H et al (2007) Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. Nat Med 13:1173–1175. https://doi.org/10.1038/nm1651
31. Compston A, Coles A (2008) Multiple sclerosis. Lancet 372:1502–1517. https://doi.org/10.1016/S0140-6736(08)61620-7
32. Kryczek I et al (2008) Induction of IL-17+ T cell trafficking and development by IFN-gamma: mechanism and pathological relevance in psoriasis. J Immunol 181:4733–4741. https://doi.org/10.4049/jimmunol.181.7.4733

33. Cai Y, Fleming C, Yan J (2012) New insights of T cells in the pathogenesis of psoriasis. Cell Mol Immunol 9:302–309. https://doi.org/10.1038/cmi.2012.15
34. Karczewski J, Dobrowolska A, Rychlewska-Hańczewska A, Adamski Z (2016) New insights into the role of T cells in pathogenesis of psoriasis and psoriatic arthritis. Autoimmunity 49:435–450. https://doi.org/10.3109/08916934.2016.1166214
35. Rosenberg WM, Moss PA, Bell JI (1992) Variation in human T cell receptor V beta and J beta repertoire: analysis using anchor polymerase chain reaction. Eur J Immunol 22:541–549. https://doi.org/10.1002/eji.1830220237
36. Moss PA, Rosenberg WM, Zintzaras E, Bell JI (1993) Characterization of the human T cell receptor alpha-chain repertoire and demonstration of a genetic influence on V alpha usage. Eur J Immunol 23:1153–1159. https://doi.org/10.1002/eji.1830230526
37. Moss PA, Bell JI (1995) Sequence analysis of the human alpha beta T-cell receptor CDR3 region. Immunogenetics 42:10–18
38. Moss PA, Bell JI (1996) Comparative sequence analysis of the human T cell receptor TCRA and TCRB CDR3 regions. Hum Immunol 48:32–38
39. Boudinot P et al (2008) New perspectives for large-scale repertoire analysis of immune receptors. Mol Immunol 45:2437–2445. https://doi.org/10.1016/j.molimm.2007.12.018
40. Pannetier C, Delassus S, Darche S, Saucier C, Kourilsky P (1993) Quantitative titration of nucleic acids by enzymatic amplification reactions run to saturation. Nucleic Acids Res 21:577–583. https://doi.org/10.1093/nar/21.3.577
41. Choi YW et al (1989) Interaction of Staphylococcus aureus toxin "superantigens" with human T cells. Proc Natl Acad Sci U S A 86:8941–8945. https://doi.org/10.1073/pnas.86.22.8941
42. Genevée C et al (1992) An experimentally validated panel of subfamily-specific oligonucleotide primers (V alpha 1-w29/V beta 1-w24) for the study of human T cell receptor variable V gene segment usage by polymerase chain reaction. Eur J Immunol 22:1261–1269. https://doi.org/10.1002/eji.1830220522
43. Lehner PJ et al (1995) Human HLA-A0201-restricted cytotoxic T lymphocyte recognition of influenza A is dominated by T cells bearing the V beta 17 gene segment. J Exp Med 181:79–91. https://doi.org/10.1084/jem.181.1.79
44. Pantaleo G et al (1994) Major expansion of CD8+ T cells with a predominant V beta usage during the primary immune response to HIV. Nature 370:463–467. https://doi.org/10.1038/370463a0
45. Pannetier C, Even J, Kourilsky P (1995) T-cell repertoire diversity and clonal expansions in normal and clinical samples. Immunol Today 16:176–181
46. Babel N, Stervbo U, Reinke P, Volk HD (2019) The identity card of T cells-clinical utility of T-cell receptor repertoire analysis in transplantation. Transplantation 103:1544–1555. https://doi.org/10.1097/TP.0000000000002776
47. Cochet M et al (1992) Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. Eur J Immunol 22:2639–2647. https://doi.org/10.1002/eji.1830221025
48. Currier JR, Robinson MA (2001) Spectratype/immunoscope analysis of the expressed TCR repertoire. Curr Protoc Immunol Chapter 10:Unit 10.28. https://doi.org/10.1002/0471142735.im1028s38
49. Chaara W et al (2018) RepSeq data representativeness and robustness assessment by Shannon entropy. Front Immunol 9:1038. https://doi.org/10.3389/fimmu.2018.01038
50. Gallard A, Foucras G, Coureau C, Guéry JC (2002) Tracking T cell clonotypes in complex T lymphocyte populations by real-time quantitative PCR using fluorogenic complementarity-determining region-3-specific probes. J Immunol Methods 270:269–280. https://doi.org/10.1016/s0022-1759(02)00336-8
51. Six A et al (2013) The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. Front Immunol 4:413. https://doi.org/10.3389/fimmu.2013.00413
52. Casrouge A et al (2000) Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. J Immunol 164:5782–5787. https://doi.org/10.4049/jimmunol.164.11.5782
53. Seo KS, Park JY, Terman DS, Bohach GA (2010) A quantitative real time PCR method to analyze T cell receptor Vbeta subgroup expansion by staphylococcal superantigens. J Transl Med 8:2. https://doi.org/10.1186/1479-5876-8-2
54. Wettstein P, Strausbauch M, Therneau T, Borson N (2008) The application of real-time PCR to the analysis of T cell repertoires. Nucleic Acids Res 36:e140. https://doi.org/10.1093/nar/gkn634
55. Manfras BJ, Rudert WA, Trucco M, Boehm BO (1997) Analysis of the alpha/beta T-cell receptor repertoire by competitive and quantitative family-specific PCR with exogenous standards and high resolution fluorescence based CDR3 size imaging. J Immunol Methods 210:235–249. https://doi.org/10.1016/s0022-1759(97)00197-x
56. VanderBorght A et al (1999) Identification of over-represented T cell receptor genes in blood and tissue biopsies by PCR-ELISA. J Immunol Methods 223:47–61. https://doi.org/10.1016/s0022-1759(98)00201-4
57. Lim A et al (2002) Combination of MHC-peptide multimer-based T cell sorting with the Immunoscope permits sensitive ex vivo quantitation and follow-up of human CD8+ T cell immune responses. J Immunol Methods 261:177–194. https://doi.org/10.1016/s0022-1759(02)00004-2

58. Calis JJ, Rosenberg BR (2014) Characterizing immune repertoires by high throughput sequencing: strategies and applications. Trends Immunol 35:581–590. https://doi.org/10.1016/j.it.2014.09.004
59. Woodsworth DJ, Castellarin M, Holt RA (2013) Sequence analysis of T-cell repertoires in health and disease. Genome Med 5:98. https://doi.org/10.1186/gm502
60. Clemente MJ et al (2013) Deep sequencing of the T-cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. Blood 122:4077–4085. https://doi.org/10.1182/blood-2013-05-506386
61. Lynch JN, Donermeyer DL, Weber KS, Kranz DM, Allen PM (2013) Subtle changes in TCRα CDR1 profoundly increase the sensitivity of CD4 T cells. Mol Immunol 53:283–294. https://doi.org/10.1016/j.molimm.2012.08.020
62. Chlewicki LK, Holler PD, Monti BC, Clutter MR, Kranz DM (2005) High-affinity, peptide-specific T cell receptors can be generated by mutations in CDR1, CDR2 or CDR3. J Mol Biol 346:223–239. https://doi.org/10.1016/j.jmb.2004.11.057
63. Rosati E et al (2017) Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol 17:61. https://doi.org/10.1186/s12896-017-0379-9
64. Lin C (2016) Analysis of TCR repertoires for tracking and evaluating distinct T cell subsets. Stem Cell Investig 3:12. https://doi.org/10.21037/sci.2016.04.01
65. Correia-Neves M, Waltzinger C, Mathis D, Benoist C (2001) The shaping of the T cell repertoire. Immunity 14:21–32
66. Sant'Angelo DB et al (1998) A molecular map of T cell development. Immunity 9:179–186
67. Kirsch I, Vignali M, Robins H (2015) T-cell receptor profiling in cancer. Mol Oncol 9:2063–2070. https://doi.org/10.1016/j.molonc.2015.09.003
68. Ma Y, Shi N, Li M, Chen F, Niu H (2015) Applications of next-generation sequencing in systemic autoimmune diseases. Genom Proteom Bioinformatics 13:242–249. https://doi.org/10.1016/j.gpb.2015.09.004
69. Okino ST, Kong M, Sarras H, Wang Y (2016) Evaluation of bias associated with high-multiplex, target-specific pre-amplification. Biomol Detect Quantif 6:13–21. https://doi.org/10.1016/j.bdq.2015.12.001
70. De Simone M, Rossetti G, Pagani M (2018) Single cell T cell receptor sequencing: techniques and future challenges. Front Immunol 9:1638. https://doi.org/10.3389/fimmu.2018.01638
71. Wu SG et al (2019) High throughput sequencing of T-cell receptor repertoire using dry blood spots. J Transl Med 17:47. https://doi.org/10.1186/s12967-019-1796-4
72. Hou X et al (2016) High throughput sequencing of T cell antigen receptors reveals a conserved TCR repertoire. Medicine 95:e2839. https://doi.org/10.1097/MD.0000000000002839
73. Rytlewski J et al (2019) Model to improve specificity for identification of clinically-relevant expanded T cells in peripheral blood. PLoS One 14:e0213684. https://doi.org/10.1371/journal.pone.0213684
74. Kivioja T et al (2011) Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods 9:72–74. https://doi.org/10.1038/nmeth.1778
75. Bolotin DA et al (2012) Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. Eur J Immunol 42:3073–3083. https://doi.org/10.1002/eji.201242517
76. Head SR et al (2014) Library construction for next-generation sequencing: overviews and challenges. Biotechniques 56:61–64., 66, 68, passim. https://doi.org/10.2144/000114133
77. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58:586–597. https://doi.org/10.1016/j.molcel.2015.05.004
78. Muraro PA et al (2014) T cell repertoire following autologous stem cell transplantation for multiple sclerosis. J Clin Invest 124:1168–1172. https://doi.org/10.1172/JCI71691
79. Han A, Glanville J, Hansmann L, Davis MM (2014) Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat Biotechnol 32:684–692. https://doi.org/10.1038/nbt.2938
80. Bai X et al (2015) Characteristics of tumor infiltrating lymphocyte and circulating lymphocyte repertoires in pancreatic cancer by the sequencing of T cell receptors. Sci Rep 5:13664. https://doi.org/10.1038/srep13664
81. Gros A et al (2014) PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating human tumors. J Clin Invest 124:2246–2259. https://doi.org/10.1172/JCI73639
82. Chattopadhyay PK, Gierahn TM, Roederer M, Love JC (2014) Single-cell technologies for monitoring immune systems. Nat Immunol 15:128–135. https://doi.org/10.1038/ni.2796
83. Nagafuchi Y, Shoda H, Fujio K (2019) Immune profiling and precision medicine in systemic lupus erythematosus. Cell 8:140. https://doi.org/10.3390/cells8020140
84. Miconnet I (2012) Probing the T-cell receptor repertoire with deep sequencing. Curr Opin HIV AIDS 7:64–70. https://doi.org/10.1097/COH.0b013e32834ddcae
85. Newell EW, Davis MM (2014) Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. Nat Biotechnol 32:149–157. https://doi.org/10.1038/nbt.2783
86. Kiselev VY, Andrews TS, Hemberg M (2019) Challenges in unsupervised clustering of single-cell RNA-seq data. Nat Rev Genet 20:273–282. https://doi.org/10.1038/s41576-018-0088-9

87. Byrne A et al (2017) Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. Nat Commun 8:16027. https://doi.org/10.1038/ncomms16027
88. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16:133–145. https://doi.org/10.1038/nrg3833
89. Shembekar N, Chaipan C, Utharala R, Merten CA (2016) Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics. Lab Chip 16:1314–1331. https://doi.org/10.1039/c6lc00249h
90. Zykwinska A, Marquis M, Sinquin C, Cuenot S, Colliec-Jouault S (2016) Assembly of HE800 exopolysaccharide produced by a deep-sea hydrothermal bacterium into microgels for protein delivery applications. Carbohydr Polym 142:213–221. https://doi.org/10.1016/j.carbpol.2016.01.056
91. Mazutis L et al (2013) Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc 8:870–891. https://doi.org/10.1038/nprot.2013.046
92. Buettner F et al (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol 33:155–160. https://doi.org/10.1038/nbt.3102
93. Thomsen ER et al (2016) Fixed single-cell transcriptomic characterization of human radial glial diversity. Nat Methods 13:87–93. https://doi.org/10.1038/nmeth.3629
94. Gee MH et al (2018) Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. Cell 172:549–563.e516. https://doi.org/10.1016/j.cell.2017.11.043
95. Konry T, Dominguez-Villar M, Baecher-Allan C, Hafler DA, Yarmush ML (2011) Droplet-based microfluidic platforms for single T cell secretion analysis of IL-10 cytokine. Biosens Bioelectron 26:2707–2710. https://doi.org/10.1016/j.bios.2010.09.006
96. Diercks AH et al (2009) A microfluidic device for multiplexed protein detection in nano-liter volumes. Anal Biochem 386:30–35. https://doi.org/10.1016/j.ab.2008.12.012
97. Sarkar S, Motwani V, Sabhachandani P, Cohen N, Konry T (2015) T cell dynamic activation and functional analysis in nanoliter droplet microarray. J Clin Cell Immunol 6:334. https://doi.org/10.4172/2155-9899.1000334
98. Stubbington MJT et al (2016) T cell fate and clonality inference from single-cell transcriptomes. Nat Methods 13:329–332. https://doi.org/10.1038/nmeth.3800
99. Staudt N, Müller-Sienerth N, Wright GJ (2014) Development of an antigen microarray for high throughput monoclonal antibody selection. Biochem Biophys Res Commun 445:785–790. https://doi.org/10.1016/j.bbrc.2013.12.033
100. Chokkalingam V et al (2013) Probing cellular heterogeneity in cytokine-secreting immune cells using droplet-based microfluidics. Lab Chip 13:4740–4744. https://doi.org/10.1039/c3lc50945a
101. McDaniel JR et al (2018) Identification of tumor-reactive B cells and systemic IgG in breast cancer based on clonal frequency in the sentinel lymph node. Cancer Immunol Immunother 67:729–738. https://doi.org/10.1007/s00262-018-2123-2
102. Segaliny AI et al (2018) Functional TCR T cell screening using single-cell droplet microfluidics. Lab Chip 18:3733–3749. https://doi.org/10.1039/c8lc00818c
103. Klein AM et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161:1187–1201. https://doi.org/10.1016/j.cell.2015.04.044
104. DeKosky BJ et al (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotechnol 31:166–169. https://doi.org/10.1038/nbt.2492
105. Tang F et al (2009) mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. https://doi.org/10.1038/nmeth.1315
106. Linnemann C et al (2013) High-throughput identification of antigen-specific TCRs by TCR gene capture. Nat Med 19:1534–1541. https://doi.org/10.1038/nm.3359
107. Chen B et al (2016) High-throughput analysis and protein engineering using microcapillary arrays. Nat Chem Biol 12:76–81. https://doi.org/10.1038/nchembio.1978
108. Macosko EZ et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
109. Han Q et al (2012) Polyfunctional responses by human T cells result from sequential release of cytokines. Proc Natl Acad Sci U S A 109:1607–1612. https://doi.org/10.1073/pnas.1117194109
110. Yamanaka YJ et al (2012) Single-cell analysis of the dynamics and functional outcomes of interactions between human natural killer cells and target cells. Integr Biol 4:1175–1184. https://doi.org/10.1039/c2ib20167d
111. Fitzgerald V et al (2015) Exploiting highly ordered subnanoliter volume microcapillaries as microtools for the analysis of antibody producing cells. Anal Chem 87:997–1003. https://doi.org/10.1021/ac503547j
112. Junkin M et al (2016) High-content quantification of single-cell immune dynamics. Cell Rep 15:411–422. https://doi.org/10.1016/j.celrep.2016.03.033
113. Brouzes E et al (2009) Droplet microfluidic technology for single-cell high-throughput screening. Proc

Natl Acad Sci U S A 106:14195–14200. https://doi.org/10.1073/pnas.0903542106
114. Pavesi A et al (2017) A 3D microfluidic model for preclinical evaluation of TCR-engineered T cells against solid tumors. JCI Insight 2:89762. https://doi.org/10.1172/jci.insight.89762
115. Xue Q et al (2017) Single-cell multiplexed cytokine profiling of CD19 CAR-T cells reveals a diverse landscape of polyfunctional antigen-specific response. J Immunother Cancer 5:85. https://doi.org/10.1186/s40425-017-0293-7
116. El Debs B, Utharala R, Balyasnikova IV, Griffiths AD, Merten CA (2012) Functional single-cell hybridoma screening using droplet-based microfluidics. Proc Natl Acad Sci U S A 109:11570–11575. https://doi.org/10.1073/pnas.1204514109
117. Sista RS et al (2008) Heterogeneous immunoassays using magnetic beads on a digital microfluidic platform. Lab Chip 8:2188–2196. https://doi.org/10.1039/b807855f
118. Love JC, Ronan JL, Grotenbreg GM, van der Veen AG, Ploegh HL (2006) A microengraving method for rapid selection of single cells producing antigen-specific antibodies. Nat Biotechnol 24:703–707. https://doi.org/10.1038/nbt1210
119. Adler AS et al (2017) Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics. MAbs 9:1282–1296. https://doi.org/10.1080/19420862.2017.1371383
120. Hatch AC et al (2011) 1-Million droplet array with wide-field fluorescence imaging for digital PCR. Lab Chip 11:3838–3845. https://doi.org/10.1039/c1lc20561g
121. Williams R et al (2006) Amplification of complex gene libraries by emulsion PCR. Nat Methods 3:545–550. https://doi.org/10.1038/nmeth896
122. Demaree B, Weisgerber D, Lan F, Abate AR (2018) An ultrahigh-throughput microfluidic platform for single-cell genome sequencing. J Vis Exp. https://doi.org/10.3791/57598
123. Eyer K et al (2017) Single-cell deep phenotyping of IgG-secreting cells for high-resolution immune monitoring. Nat Biotechnol 35:977–982. https://doi.org/10.1038/nbt.3964
124. Menon V (2018) Clustering single cells: a review of approaches on high-and low-depth single-cell RNA-seq data. Brief Funct Genom 17:240–245. https://doi.org/10.1093/bfgp/elx044
125. Poulin JF, Tasic B, Hjerling-Leffler J, Trimarchi JM, Awatramani R (2016) Disentangling neural cell diversity using single-cell transcriptomics. Nat Neurosci 19:1131–1141. https://doi.org/10.1038/nn.4366
126. Singh M et al (2019) High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun 10:3120. https://doi.org/10.1038/s41467-019-11049-4
127. Kim SM et al (2012) Analysis of the paired TCR α- and β-chains of single human T cells. PLoS One 7:e37338. https://doi.org/10.1371/journal.pone.0037338
128. Howie B et al (2015) High-throughput pairing of T cell receptor α and β sequences. Sci Transl Med 7:301ra131. https://doi.org/10.1126/scitranslmed.aac5624
129. Afik S et al (2017) Targeted reconstruction of T cell receptor sequence from single cell RNA-seq links CDR3 length to T cell differentiation state. Nucleic Acids Res 45:e148. https://doi.org/10.1093/nar/gkx615
130. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-seq: single-cell RNA-seq by multiplexed linear amplification. Cell Rep 2:666–673. https://doi.org/10.1016/j.celrep.2012.08.003
131. Sasagawa Y et al (2017) Erratum to: Quartz-seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol 18:9. https://doi.org/10.1186/s13059-017-1154-x
132. Eltahla AA et al (2016) Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol Cell Biol 94:604–611. https://doi.org/10.1038/icb.2016.16
133. Love KR, Bagh S, Choi J, Love JC (2013) Microtools for single-cell analysis in biopharmaceutical development and manufacturing. Trends Biotechnol 31:280–286. https://doi.org/10.1016/j.tibtech.2013.03.001
134. Torres AJ, Hill AS, Love JC (2014) Nanowell-based immunoassays for measuring single-cell secretion: characterization of transport and surface binding. Anal Chem 86:11562–11569. https://doi.org/10.1021/ac4030297
135. Esfandiary L et al (2016) Single-cell antibody nanowells: a novel technology in detecting anti-SSA/Ro60- and anti-SSB/La autoantibody-producing cells in peripheral blood of rheumatic disease patients. Arthritis Res Ther 18:107. https://doi.org/10.1186/s13075-016-1010-5
136. Yonaha F et al (1992) Restricted junctional usage of T cell receptor V beta 2 and V beta 13 genes, which are overrepresented on infiltrating T cells in the lips of patients with Sjogren's syndrome. Arthritis Rheum 35:1362–1367. https://doi.org/10.1002/art.1780351118
137. Sumida T et al (1992) T cell receptor repertoire of infiltrating T cells in lips of Sjogren's syndrome patients. J Clin Invest 89:681–685. https://doi.org/10.1172/JCI115635
138. Freimark B, Pickering L, Concannon P, Fox R (1989) Nucleotide sequence of a uniquely expressed human T cell receptor beta chain variable region gene (V beta) in Sjogren's syndrome. Nucleic Acids Res 17:455. https://doi.org/10.1093/nar/17.1.455
139. Matsumoto I et al (1996) Common T cell receptor clonotype in lacrimal glands and labial salivary

glands from patients with Sjogren's syndrome. J Clin Invest 97:1969–1977. https://doi.org/10.1172/JCI118629
140. Smith MD et al (1994) Selective expression of V beta families by T cells in the blood and salivary gland infiltrate of patients with primary Sjogren's syndrome. J Rheumatol 21:1832–1837
141. Sumida T, Matsumoto I, Maeda T, Nishioka K (1997) T-cell receptor in Sjogren's syndrome. Br J Rheumatol 36:622–629. https://doi.org/10.1093/rheumatology/36.6.622
142. Dash P et al (2011) Paired analysis of TCRalpha and TCRbeta chains at the single-cell level in mice. J Clin Invest 121:288–295. https://doi.org/10.1172/JCI44752
143. Kim SM et al (2012) Analysis of the paired TCR alpha- and beta-chains of single human T cells. PLoS One 7:e37338. https://doi.org/10.1371/journal.pone.0037338
144. Han A et al (2013) Dietary gluten triggers concomitant activation of CD4+ and CD8+ alphabeta T cells and gammadelta T cells in celiac disease. Proc Natl Acad Sci U S A 110:13073–13078. https://doi.org/10.1073/pnas.1311861110
145. Joachims ML et al (2016) Single-cell analysis of glandular T cell receptors in Sjogren's syndrome. JCI Insight 1:e85609. https://doi.org/10.1172/jci.insight.85609
146. Voigt A et al (2018) Unique glandular ex-vivo Th1 and Th17 receptor motifs in Sjogren's syndrome patients using single-cell analysis. Clin Immunol 192:58–67. https://doi.org/10.1016/j.clim.2018.04.009
147. Wanchoo A et al (2017) Single-cell analysis reveals sexually dimorphic repertoires of Interferon-gamma and IL-17A producing T cells in salivary glands of Sjogren's syndrome mice. Sci Rep 7:12512. https://doi.org/10.1038/s41598-017-12627-6
148. Karabiyik A, Peck AB, Nguyen CQ (2013) The important role of T cells and receptor expression in Sjogren's syndrome. Scand J Immunol 78:157–166. https://doi.org/10.1111/sji.12079
149. Mahony J et al (2007) Development of a respiratory virus panel test for detection of twenty human respiratory viruses by use of multiplex PCR and a fluid microbead-based assay. J Clin Microbiol 45:2965–2970. https://doi.org/10.1128/JCM.02436-06
150. Perfetto SP, Chattopadhyay PK, Roederer M (2004) Seventeen-colour flow cytometry: unravelling the immune system. Nat Rev Immunol 4:648–655. https://doi.org/10.1038/nri1416
151. Proserpio V, Mahata B (2016) Single-cell technologies to study the immune system. Immunology 147:133–140. https://doi.org/10.1111/imm.12553
152. Svensson V, Vento-Tormo R, Teichmann SA (2018) Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 13:599–604. https://doi.org/10.1038/nprot.2017.149

Shivai Gupta received a master's degree from Georgetown University, Washington DC in 2016. She is currently a PhD student under the supervision of Dr. Cuong Nguyen at University of Florida's College of Medicine. Her research and experience include having worked in infectious disease at the USFDA on flaviviruses such as Zika and dengue virus. She has also worked in the formulation vaccine development for respiratory syncytial virus at Novavax Inc. Her experience also includes clinical trial operations for the NIH's Division of AIDS. Ms. Gupta's current project focuses on therapeutic approaches for Sjögren's syndrome using in-silico high-throughput small molecule drug screening.

Richard Witas obtained his bachelor's degree from the University of South Florida in 2014, where he studied the role of exosomes in shuttling the neurotoxic peptide Aβ in Alzheimer's disease. He is currently a Ph.D. student under the supervision of Prof. Cuong Nguyen. His research is centered on characterizing how impaired functioning of cells of the innate immune system may promote the onset of Sjögren's syndrome. Specifically, he seeks to understand how aberrant TAM receptor tyrosine kinase signaling within phagocytes contributes to an autoimmune response in Sjögren's syndrome.

Alexandria Voigt obtained her Bachelor's degree in Biochemistry from the Georgia Institute of Technology. She studied meiotic non-disjunction events in fruit flies on live, single oocytes, then targeting of embryonic stem cells to produce a mouse model of myotonic dystrophy, and studied various other mouse models of neuromuscular genetic disorders. Currently, her research focuses on the isolation and identification of the T- and B-cell receptor repertories via single-cell technologies and elucidating the etiology of Sjögren's syndrome.

Touyana Semenova obtained her MD from Irkutsk State Medical University in 2007 and Ph.D. degree in 2016 from the Graduate School of Chemistry and Life Sciences, Grenoble Alpes University, France. She has completed the first joint Ph.D. thesis in Medical Sciences between France and Russia. She has an experience in clinical virology, development of clinical research studies and laboratory assays. Dr. Semenova is currently a postdoctoral research associate in Dr. Cuong Nguyen's group. She has coauthored over ten publications. Her current research focuses on the development of neutralizing antibodies against flaviviruses using single-cell technology.

Cuong Q. Nguyen graduated from the University of Nebraska-Lincoln in 1999 where he majored in Biology and Biochemistry. He obtained his PhD from the University of Florida in 2006. His doctoral thesis focused on the role of complement in B cell function and activation in Sjögren's syndrome. He conducted part of his postdoctoral research under the guidance of Dr. Christopher Love at MIT where he studied single-cell microengraving. Dr. Nguyen's primary research focus has been on

understanding the etiology of Sjögren's syndrome using both human and animal models. His laboratory has developed a number of animal models that make it possible to examine many facets of the disease etiology. Dr. Nguyen's research team was one of the first groups to identify the presence of IL-17/Th17 cells in Sjögren's syndrome and its clinical correlation to the disease. His continuing research on Th17 cell biology with its signature cytokines emphasizes the importance and interaction of these immune cells in the autoimmune process. Lastly, using high-throughput sequencing in combination with single-cell microengraving technology, Dr. Nguyen's team investigates the clonal diversity of the T cell receptors of pathogenic T cells. His seminal studies have found that there is an elevated prevalence of pathogenic effector T cells in the salivary glands with a sexually dimorphic selection bias of T cell receptor repertoires. Dr. Nguyen's recent publication determines that activated Th1 and Th17 cells of primary Sjögren's syndrome patients showed restricted clonal diversities with some unique antigen binding motifs.

DNA Methylation in Pulmonary Fibrosis

4

Shuang Zhou, Xiangdong Wang, Hongzhi Gao, and Yiming Zeng

Abstract

DNA methylations, including global methylation pattern and specific gene methylation, are associated with pathogenesis and progress of pulmonary fibrosis. This chapter illustrates alteration of DNA methylation in pulmonary fibrosis as a predictive or prognostic factor. Treatment with the DNA methylation inhibitors will be an emerging anti-fibrosis therapy, although we are still in the pre-clinical stage of using epigenetic markers as potential targets for biomarkers and therapeutic interventions.

Keywords

Methylation · DNA · Lung fibrosis · Pulmonary fibrosis · IPF

S. Zhou · H. Gao (✉)
Clinical Center for Molecular Diagnosis and Therapy, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, China

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China
e-mail: Xiangdong.wang@clintransmed.org

Y. Zeng
2nd Affiliated Hospital, Fujian Medical University, Quanzhou, Fujian, China

4.1 Introduction

Idiopathic pulmonary fibrosis (IPF) is a serious form of pulmonary fibrosis, with which patients have the median survival time of about 2–3 years [1]. IPF is also a type of chronic lung disease characterized by a progressive scarring of the lung parenchyma and irreversible decline in lung function with hypoxemia and dyspnea. The prevalence and mortality of pulmonary fibrosis are on the rise with age, especially among people over 50 years old [2]. The incidence of IPF in men is higher than that in women and is more common in smokers [3]. Even after smoking cessation, the status of IPF cannot be improved. The pathogenesis of IPF is not completely clear and the clinical manifestation of IPF is highly variable. However, there are still some recognized potential risk factors such as environmental exposure, microbial agents, or gastroesophageal reflux. Recent studies have shown that gene expression and epigenetic regulation, especially the DNA methylation regulation, play an important role in the development of IPF [4–6].

DNA methylation is an inherited epigenetic process, involving the covalent transfer of the c-5 position of the DNA cytosine loop by the catalysis of DNA methyltransferases (DNMTs) [7]. The methylation alters gene function but does not change the sequence. The majority of DNA methylation occurs on the fifth carbon atom of cytosines that precede a guanine nucleotide or CpG sites [8]. DMA methylation is a dynamic

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_4

and inheritable process. Methylation of CpG island promoters prevents the binding of transcription factors and results in gene silencing and repression. On the contrary, hypomethylation and demethylation are associated with upregulation of gene expression [9]. DNMTs and methyl-binding proteins (MBPs) are major enzymes to catalyze DNA methylation [10], essential for transcriptional regulation and normal development and related to genomic imprinting, repression of transposable elements, X-chromosome inactivation, carcinogenesis, and aging [7, 11].

Epigenetic changes are associated with numerous diseases including cancers and pulmonary fibrosis, where large hypomethylated blocks of genomes and promoter hypermethylation of classic suppressor genes were found [8]. Studies on DNA methylation analysis confirmed that DNA methylation is common and important in pulmonary fibrosis. And numerous specific genes are involving in pathogenesis, such as Thy-1 (CD90), prostaglandin receptor 2 (PTGER2), cyclo-oxygenase-2 (COX-2), $p14^{ARF}$, or chemokine IP-10 [12–16]. This chapter will focus on the global genome methylation pattern and targeted DNA methylation status in the pathogenesis of lung fibrosis, and then discuss the potential therapies of methylation inhibitors [17, 18].

4.2 Genome-Wide DNA Methylation in IPF

Methodologies for methylation measurement include next generation high throughput sequencing, whole genome bisulfite sequencing (WGBS), microarray, methylated DNA immunoprecipitation sequencing (Me DIP-Seq), bisulfite genomic sequence (BGS), and methylation-specific PCR (MSP). WGBS, Me DIP-Seq, microarray, and BGS are widely used in genome-wide DNA methylation analysis. For example, the human CpG islands microarray and WGBS were used to detect the alteration of the whole DNA extracted from the lung tissues of patients with or without IPF [15]. The extensive DNA methylation changes were found within CpG islands in IPF lung samples, different from methylation profiles of healthy, although partial methylated areas have many similarities [15]. The DNA methylation and RNA expression changed in lung tissue from IPF using human methylation chip and RNA hybridization chip. Altered DNA methylation is consistent with the mRNA expression of many genes, indicating the importance of DNA methylation in the pathogenesis of IPF [8]. Unfortunately, it is hard to clarify the alternations of DNA methylation within the individual cell type and difference between cell types, since most studies are based on the entire lung tissue.

The genome-wide differences in DNA methylation were detected in fibroblasts isolated from lung tissue of IPF patients, as compared with patients with lung nodules [19]. The methylation differences are mainly concentrated in genes associated with cell proliferation, extracellular matrix generation, potassium channel, and organ organogenesis and corresponded with alteration of gene expression at mRNA and protein levels [19].

4.3 IPF Specificity of Thy-1 DNA Methylation

Several specific genes were considered as IPF-specific and their DNA hypermethylation is consistent with the downregulated expression, such as Thy-1, COX-2, PTGER2, $p14^{ARF}$, and chemokine IP-10 [13–16, 20, 21]. The reduction in the expression of those genes can directly induce the initiation of fibro-genesis, activation of fibroblast proliferation, and resistance to apoptosis [1]. Of those, Thy-1 cell surface antigen (Thy-1) is also known as CD90, a 25–37 kDa glycoprotein, localizing to lipid rafts and on the external leaflet of the lipid bilayer [22]. The activation of Thy-1 promotes T cell activation and affects multiple non-immunologic biological processes, such as cellular adhesion, migration, cell death, wound healing, neurite outgrowth, tumor repression, and fibrosis. Thy-1 as a highly conserved molecule has two membrane-bound and soluble forms and the biological role of

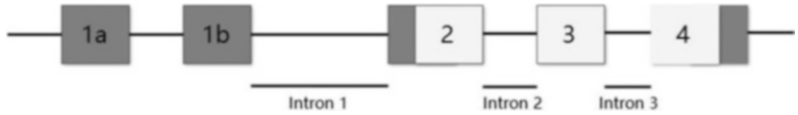


Fig. 4.1 Thy-1 gene structure. Exons 1a and 1b encode two distinct alternative spliced mRNA; exon 3 for the mature protein, and the 50-end of exon 4 for the trans-membrane sequence. Portions of the gene encoding for the mature Thy-1 protein are marked as light gray orthogons. Dark gray orthogons complete the exons

Thy-1 dependent upon cell type and tissue specificity [23]. Thy-1 is often used as a marker for cell types and has a crucial effect on cell biology, of which the dysregulation is related to fibrotic diseases and malignancy [23]. *Thy-1* located in chromosome 9 in mice and chromosome 11q22.3 in human are both initially expressed in the form of 161 a.a pro-form and have different post-transcriptional modifications [24]. Two different proteins encoded from the alleles differ only in position 89, of which one is arginine and the other is glycine. Thy-1 in human has only one allele for thymine, and the first 19 a.a pro-form positions the signal to targets the endoplasmic reticulum (ER) [25]. Thy-1 has two isoforms in mice: Thy-1.2 in Bal/c mice and Thy-1.1 in AKR mice have a glutamine and an arginine at the position 89, respectively. Genetic characteristics of *Thy-1* genes are similar among human, mouse, and rat [26]. Human *Thy-1* contains four exons, of which exon 1 (Fig. 4.1a, b) produce two mRNA splicing variants after transcription and exon 2 contains the translation starting site, exon 3 encodes the amino acids 7–106, and exon 4 is mainly responsible for the C-terminal end and poly-A tail [27] (Fig. 4.1).

Thy-1 participates in a number of signaling cascades and acts as a universal signal modulator in proliferation, survival, cell adhesion, and cytokine/growth factor responses [23]. Thy-1 undergoes signal transduction in non-immunologic cells by integrins, growth factors, cytokines, and protein tyrosine kinases. The roles of those signaling cascades mainly focus on cell proliferation, apoptosis, cellular adhesion, and migration. Thy-1 interacts with itself, adaptors, scaffolds, or signaling molecules, such as reggies-1/2, Src family of C-terminal Src kinase (Csk)-binding protein (CBP) and protein tyrosine kinases (SFK), in the cell membrane of several cell types to convey signals to the cell interior. Thy-1 is an important component of protein complexes, to initiate cell signaling from rafts (Fig. 4.2). In addition, Thy-1 interacts with other receptors at the plasma membrane such as $\alpha_V\beta_5$ integrin in fibroblasts [28]. Thy-1(−) fibroblasts move faster and migrate more efficiently in wound healing than Thy-1(+) ones [28]. A mechanism to regulate fibroblast migration is involved in SFK and Rho GTPase activation [27]. It is proposed that Thy-1 expression regulates Src and FAK kinase activation, as well as phosphorylation of p190RhoGAP by increasing RhoA-GTP levels, to stress fiber and focal adhesion formation [29]. Decreased migration of Thy-1 (+) fibroblast subpopulations may occur as the consequence of a complex Thy-1-triggered signaling process, in addition to passive Thy-1-to-matrix adhesion [27]. It implies Thy-1-dependent roles in fibroblast-matrix adhesion and migration.

The loss of Thy-1 expression in lung fibroblasts correlates with many aspects of the fibrogenic phenotype including proliferation [25]. The proliferated myofibroblasts in the fibroblast foci were found Thy-1 negative in IPF, rather than in the normal fibroblasts [30]. Thy-1 can not only regulate the expression of myogenic gene, promote myofibroblastic differentiation, but also determine the survival of lung fibroblasts. Yan Y. Sanders et al. [20] demonstrated that Thy (−) fibroblasts proliferated in myofibroblastic foci, inhibiting the myofibroblast differentiation of fibroblasts, which was restored by DNA methyltransferase inhibitors. The epigenetic downregulation of Thy-1 occurred in cell transformation and clinical malignant tumor [20]. Rat lung fibroblasts without Thy-1 on the surface, low expression of myogenic genes and low protein levels of sarcomeric myosin, α-SMA, and MyoD, had high responses to

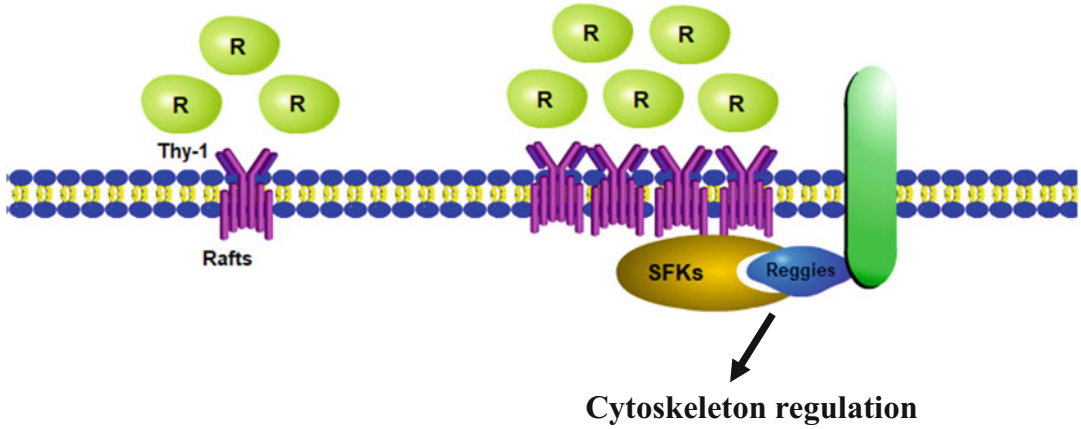


Fig. 4.2 Signaling induced by Thy-1. Thy-1 binds to its ligand (R) and undergoes molecular clustering at the plasma membrane. Thy-1 interacts with itself, with adaptors, scaffolds, or signaling molecules, such as reggies-1/2, Src family of C-terminal Src kinase (Csk)-binding protein (CBP) and protein tyrosine kinases (SFK), in the cell membrane of several cell types to convey signals to the cell interior

pro-myofibroblastic stimuli including TGF-β [30].

Loss of Thy-1 expression appears to be associated with the differentiation of myofibroblasts both in mouse bleomycin model and IPF patients [31]. The relation between Thy-1 and myofibroblasts phenotype seems to be tissue-specific and dependent. Loss of Thy-1 expression also resulted in the hypermethylation of the Thy-1 promoter in IPF Samples and was restored through demethylation, similar between human and rat lung fibroblasts [20].

4.4 IPF Specificity of COX-2 DNA Methylation

Cyclooxygenases (COXs) are a 67–72 kDa integral membrane protein, are located on the nuclear membrane and the endoplasmic reticulum (ER), and contain three isoforms [32]. COX-1 is expressed constitutively like "housekeeping" enzyme associated with homeostasis, COX-2 is the inducible form and is upregulated in both inflammation and cancer, and COX-3 is expressed in spinal cord and brain although its functions remain unclear [33]. Cyclooxygenase-2 (COX-2) is referred to prostaglandin endoperoxide synthase (PTGS)I as a key enzyme that catalyzes the conversion of arachidonic acid (AA) to prostaglandins (PGs) [34]. COX-2 plays a crucial role in some pathophysiological processes, including angiogenesis, inflammation, tumorigenesis, and tumor drug resistance, and becomes a new target for cancer treatment [35]. In solid tumors such as colorectal cancer, prostate cancer, breast cancer, and most recently hematological malignancies, COX-2 mainly functions as a regulator of cell proliferation and apoptosis [33]. The activation and overexpression of COX-2 were found in tumor cells related to tumor progression and aggressiveness [36]. COX-2 expression could be induced by anticancer chemoradiotherapy, resulting in drug resistance [36]. The inhibition of COX-2 was proposed as an attractive new strategy for cancer treatment in patients [37]. Non-steroidal anti-inflammatory drugs (NSAIDs), broad spectrum COX-2-inhibitors, or COX-2-specific inhibitors were found to have side-effects, such as myocardial infarction [36]. The development of new anti-COX-2 drugs with less side-effects seems particularly urgent [34, 38].

COX-1 gene is located on chromosome 9 (9q32-9q33.3), nearly 40 kilobase (kb) pairs, containing 11 exons and its mRNA is 2.8 kb. COX-2 is located on chromosome 1 (1q25.2-25.3), containing ten exons approximately

8.3 kb and transcript about 4.5 kb [39]. In the flanking region of *COX-2*, there are 50 bps of the regulation area of gene transcription, containing a TATA box and a few putative transcription-factor binding sites of NF-IL-6, NF-κB, and a TGF-β response element, which demonstrates a complex combination of the factors associated with COX-2 gene regulation [40]. Single nucleotide polymorphism (SNP) in the gene promoter affects transcription of COX-2 gene. The most frequently functional polymorphisms of COX-2 gene, _765G>C (rs20417) and _1195G>A (rs689466), are correlated with inflammatory disorders, such as chronic periodontitis [41], inflammatory bowel diseases, and subclinical atherosclerosis [41]. This is probably because those gene polymorphisms may alter the function of COX-2 by regulation of COX-2 expression and affect the synthesis of prostaglandins in the pathogenesis of inflammatory diseases [42].

Prostaglandin E2 (PGE_2), the major catalyzed product of COX-2, plays a key role in the tumorigenesis of colorectal cancer [43]. The COX-2/PGE_2-JAK2/STAT3 signaling pathway may be the drug target for berberine to mediate the effect on metastasis and invasiveness of cancer. The berberine reduced COX-2/PGE_2 levels, inhibited JAK2/STAT3 activation, decreased expression of downstream target genes MMP-2/-9, and caused less metastasis and invasiveness in cancer [44] (Fig. 4.3). PGE_2 is associated with occurrence of malignant tumors and plays a beneficial role in lung fibrotic diseases. This is partially due to the function of PGE_2 to limit the proliferation of lung fibroblasts and to inhibit myofibroblast differentiation, migration, and collagen secretion. Figure 4.4 diagrams the homeostatic and anti-fibrotic behavior of PGE2 signaling pathway in fibroblasts and lung epithelial cells (AECs) [45].

The expression of COX-2 was downregulated in IPF and upregulated in COPD as well as in IPF and sclerosis [46, 47]. COX-2 downregulation and reduced PGE2 production are related to myofibroblasts in the development and progression of IPF [48]. The downregulation of COX-2 could reduce PGE2 and induce the continuous proliferation of fibroblasts, which is considered as a new viewpoint in the pathogenesis of IPF [49]. Lung fibroblasts derived from IPF patients were unable to induce PGE2 synthesis, even if stimulated by proinflammatory cytokines and LPS, probably due to the abnormal expression of COX-2 [45, 50]. In patients with IPF, the PGE2 level of bronchoalveolar lavage fluid was significantly lower than that of normal individuals, which is because PGE2 could reduce the proliferation of fibroblast and collagen aggregation by inhibiting COX-2-dominated synthesis and promotion of degradation, beneficial for inhibiting pulmonary fibrosis [51].

COX-2 was downregulated in lung tissue from patients with IPF [15, 52]. By upregulation of DNMT3a expression, PGE2 increases the gene-specific DNA methylation of lung fibroblasts, such as MGMT gene and IGFBP2 gene [53].

The transcriptional regulatory factor c8orf4 for COX-2 was demethylated via 5-AZAdc, a DNA methylation inhibitor to reverse decreased level of COX-2 mRNA in a dose-dependent pattern [15, 53]. C8orf4 regulates the expression of COX-2 in lung fibroblasts by binding of the proximal promoter by the hypermethylation of the transcription regulator as an indirect epigenetic mechanism to regulate COX-2 expression and COX-2 derived PGE2 synthesis in pulmonary fibrosis [15].

4.5 $p14^{ARF}$ and Function

The $p14^{ARF}$ protein as a tumor suppressor protein is an alternate reading frame protein (ARF) encoded by CDKN2A gene. ARF is a 14 kDa, 132 a.a protein named $p14^{ARF}$ in human, and a 19 kDa, 169 a.a protein named $p19^{ARF}$ in mice [54]. $P14^{ARF}$ is a cell cycle regulation protein to block the cell cycle in the G1 and G2 phases and inhibit the growth of abnormal cells by activating p53 indirectly [55]. $p14^{ARF}$ protein binds to and interferes with the Mdm2 protein, a p53 negative-regulator, and then stabilizes and activates p53 pathway [54, 56]. The role of $p14^{ARF}$ in carcinogenesis was evidenced by the finding that ARF-null mice have a high tendency to induce

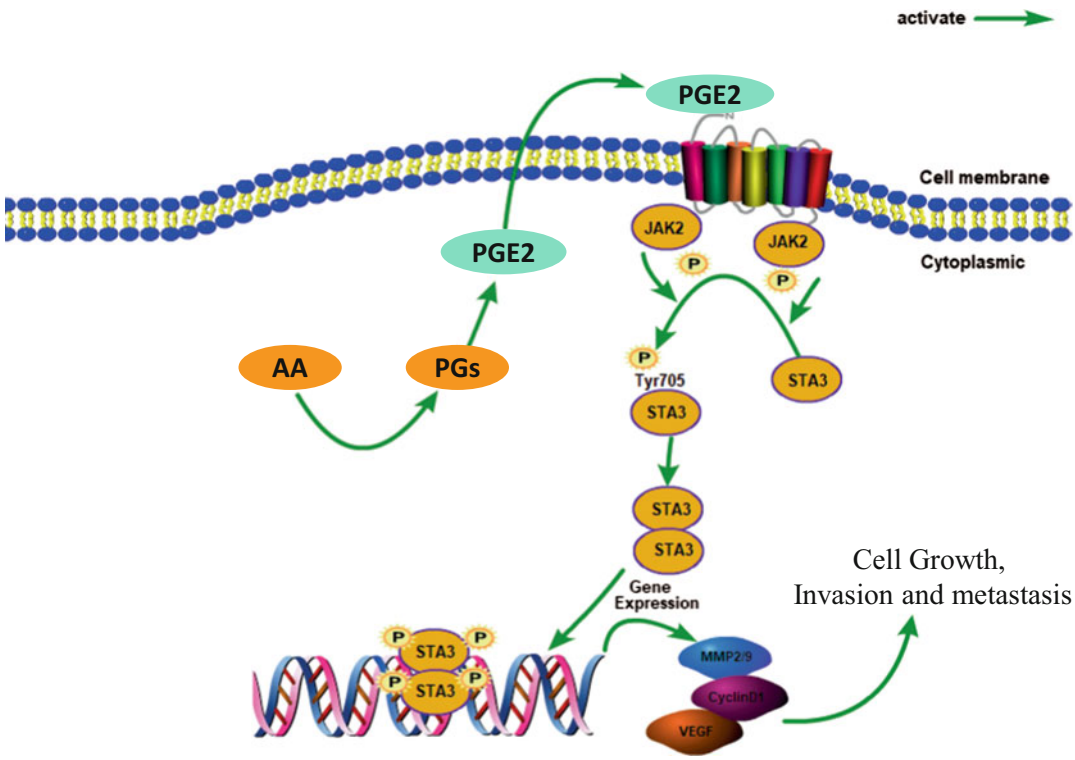


Fig. 4.3 COX-2/PGE2-JAK2/STAT3 signaling pathway. PGE2, the main catalyzed product of COX-2 from arachidonic acid, could bind to the EP receptor on the cell membrane, thereby activating the JAK2, followed by the phosphorylating of STAT3 in the Tyr705 site. Berberine inhibits invasion and metastasis of colorectal cancer cells via COX-2/PGE2 mediated JAK2/STAT3 signaling pathway

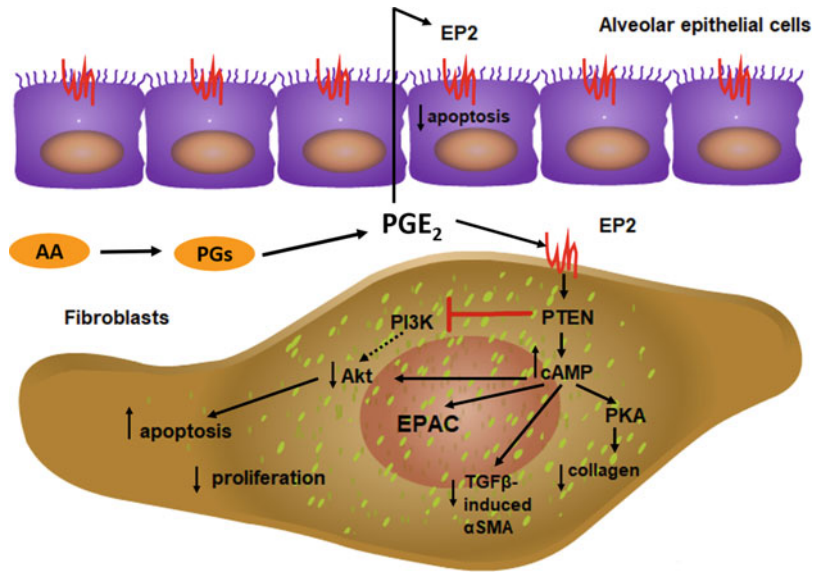


Fig. 4.4 PGE2 signaling pathway in lung fibrosis. Diagrams the homeostatic and anti-fibrotic behavior of PGE2 signaling pathway in fibroblasts and lung epithelial cells (AECs)

tumors, e.g., carcinomas, gliomas, lymphomas, and sarcomas, leading to death early in life [57].

The INK4a–ARF locus (CDKN2A in humans) on chromosome 9p21 encodes two structure-similar tumor suppressor proteins with different functions, $p14^{ARF}$ ($p19^{ARF}$ in the mouse) and p16INK4a to indirectly control the activities of p53 and the retinoblastoma protein (RB) transcription factor, respectively [58]. $p14^{ARF}$ and INK4a mRNA consist of 3 exons of which exons 2 and 3 are the same with two different exon 1 transcripts (α and β) [59, 60]. Although $p14^{ARF}$ has an unrelated structure, it can also cause cell cycle arrest in G1 and G2 phase [61]. $P14^{ARF}$ gene as a tumor suppressor gene plays an important role in the progression and pathogenesis of tumor, since it is usually mutated or deleted [62, 63].

The dysfunction of the p14ARF-Mdm2-p53 pathway, also known as p53 pathway, is one of the most important signals of cancer pathogenesis. The $p14^{ARF}$ in the p53 pathway binds with Mdm2 in the nucleolus, resulting in the inability of Mdm2 to degrade p53 [64, 65] (Fig. 4.5). The activity of Mdm2 can be inhibited by $p14^{ARF}$, to indirectly block the degradation of p53. When p53 is activated, the consequences of the ARF-p53 binding depend on the cell cycle state [66]. $P14^{ARF}$ controls the expression of p53, and then activated p53 secondarily regulates the expression of $p14^{ARF}$ by negative feedback [67]. Overexpression of $p14^{ARF}$ in the nucleus contributes to the loss of shuttling ability of Mdm2 and induces p53 mutations [68]. This pathway is inactivated by p14ARF deletion, p53 mutation, or amplification of Mdm2, which is complex and interactive but common and important.

The p53/$p14^{ARF}$ signaling pathway is often downregulated in patients with colorectal cancer, and $p14^{ARF}$ is highly methylated in the early stages of colorectal cancer [69]. The methylation of $p14^{ARF}$ may have predictive value for early colorectal cancer patients, but not as a prognostic factor. The target drug for $p14^{ARF}$ demethylation may be a new direction for the development of new colorectal cancer drugs [69]. The $p14^{ARF}$ gene can be inactivated in many cancers, due to deletion, promoter hypermethylation, or mutations [69]. In the evolution of oligodendrogliomas, the hypermethylation-resulted aberrant p14ARF expression and the deletions of p14ARF/p16INK4a are associated with the progression to anaplastic oligodendroglioma [70, 71]. Studies on the methylation status of the $p14^{ARF}$ promoter suggested that $p14^{ARF}$ can be a useful biomarker for the pathological TNM stage, prognosis, and clinical outcome of cancer patients [72]. Homozygous deletion of the $p14^{ARF}$ gene loci was detected in multiple carcinomas and was associated with tumorigenesis. DNA methylation

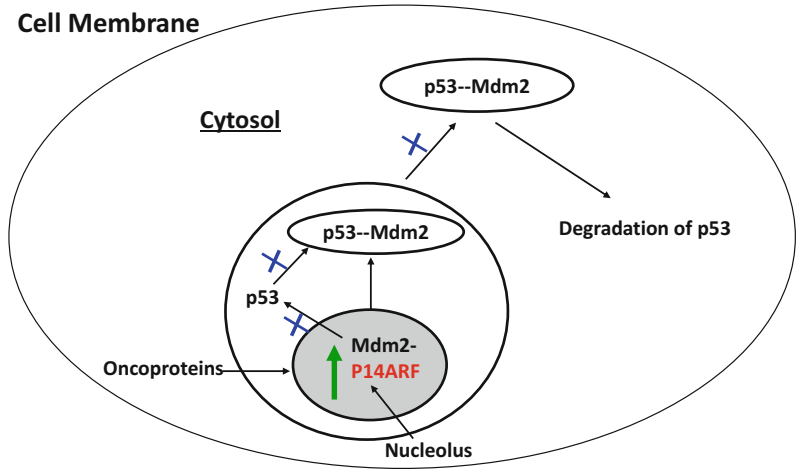


Fig. 4.5 p14ARF-Mdm2-p53 pathway in breast cancer. Mdm2 translocates from the nucleolus to the nucleoplasm and binds to p53. The Mdm2-p53 complex then migrates to the cytoplasm, resulting in the degradation of p53

can regulate $p14^{ARF}$ mRNA levels, and the methylation status of $p14^{ARF}$ is related to the occurrence of primary liver cancer and TNM staging [73]. The promoter methylation status of $p14^{ARF}$ in fibroblasts isolated from IPF and normal lung demonstrated that hypermethylated $p14^{ARF}$ occurred in half of the IPF fibroblasts and was correlated with the decreased expression of the gene and protein as well as increased resistance to apoptosis [16].

Hypermethylation and downregulated expression of PTGER2 also play an important role in the development of IPF. Levels of DNA hypermethylation were higher in fibroblasts isolated from mice and human lungs with pulmonary fibrosis, leading to a decrease in EP2 expression level and PGE2 resistance [14]. Therapies with DNA methylation inhibitors (e.g., 5-Aza-2-′-deoxycytidine and zebularine) reversed the reduced mRNA and protein expression of EP2, and restored PGE2 activities in fibrotic fibroblasts. Those results indicate that DNA hypermethylation play the decisive role in the downregulation of PTGER2 expression and subsequent PGE2 resistance. The enhancement of Akt signal transduction may be a new mechanism of the promotion of DNA hypermethylation in the formation of lung fibrosis [14].

4.6 Conclusion and Prospective

DNA methylation is one of mechanisms by which the epigenetic regulation plays a crucial role in lung fibrosis, cancer, and chronic diseases. Global methylation pattern and specific gene methylation status as an important regulatory factor contribute to the development of pulmonary fibrosis. DNA methylation of associated genes is associated with the occurrence and progression of pulmonary fibrosis and change the phenotype and destiny of fibroblasts through the regulation of cell activation, differentiation, and balance of fibrotic and anti-fibrotic gene expressions.

Methylation patterns and severities of the promoter regions of Thy-1, COX-2, $p14^{ARF}$, and PTGER2 genes should be considered as disease-specific biomarkers to predict the occurrence and development of IPF. The intracellular mechanisms and heterogeneity of DNA methylation in the regulation of signal pathway activities should be investigated by single-cell DNA and RNA sequencing [74–76]. The promoter methylation of the target genes can contribute to the pathogenesis and development of pulmonary fibrosis through multiple signal pathways, which should be furthermore identified and validated with advanced biotechnologies [77–81].

References

1. Sundarakrishnan A, Chen Y, Black LD, Aldridge BB, Kaplan DL (2018) Engineered cell and tissue models of pulmonary fibrosis. Adv Drug Deliv Rev 129:78
2. King TE Jr, Pardo A, Selman M (2011) Idiopathic pulmonary fibrosis. Lancet 378(9807):1949–1961
3. Rangarajan S, Locy ML, Luckhardt TR, Thannickal VJ (2016) Targeted therapy for idiopathic pulmonary fibrosis: where to now? Drugs 76(3):291–300
4. Macneal K, Schwartz DA (2012) The genetic and environmental causes of pulmonary fibrosis. Proc Am Thorac Soc 9(3):120–125
5. Ziobro R, Henry B, Edwards MJ, Lentsch AB, Gulbins E (2013) Ceramide mediates lung fibrosis in cystic fibrosis. Biochem Biophys Res Commun 434(4):705–709
6. Wolters PJ, Collard HR, Jones KD (2014) Pathogenesis of idiopathic pulmonary fibrosis. Ann Rev Pathol Mech Dis 9:157–179
7. Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. Neuropsychopharmacology 38 (1):23–38
8. Sanders YY, Ambalavanan N, Halloran B, Zhang X, Liu H, Crossman DK et al (2012) Altered DNA methylation profile in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 186(6):525–535
9. Marchal C, Miotto B (2015) Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns. J Cell Physiol 230 (4):743–751
10. Du Q, Luu PL, Stirzaker C, Clark SJ (2015) Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7(6):1051–1073
11. Gendrel AV, Heard E (2014) Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu Rev Cell Dev Biol 30:561–580
12. Rabinovich EI, Kapetanaki MG, Steinfeld I, Gibson KF, Pandit KV, Yu G et al (2012) Global methylation patterns in idiopathic pulmonary fibrosis. PLoS One 7 (4):e33770
13. Robinson CM, Neary R, Levendale A, Watson CJ, Baugh JA (2012) Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the

development of a pro-fibrotic phenotype. Respir Res 13(1):74
14. Huang SK, Fisher AS, Scruggs AM, White ES, Hogaboam CM, Richardson BC et al (2010) Hypermethylation of PTGER2 confers prostaglandin E2 resistance in fibrotic fibroblasts from humans and mice. Am J Pathol 177(5):2245–2255
15. Evans IC, Barnes JL, Garner IM, Pearce DR, Maher TM, Shiwen X et al (2016) Epigenetic regulation of cyclooxygenase-2 by methylation of c8orf4 in pulmonary fibrosis. Clin Sci 130(8):575–586
16. Cisneros J, Hagood J, Checa M, Ortiz-Quintero B, Negreros M, Herrera I et al (2012) Hypermethylation-mediated silencing of p14ARF in fibroblasts from idiopathic pulmonary fibrosis. Am J Phys Lung Cell Mol Phys 303(4):L295–L303
17. Zhang X, Hu M, Lyu X, Li C, Thannickal VJ, Sanders YY (2017) DNA methylation regulated gene expression in organ fibrosis. Biochim Biophys Acta Mol basis Dis 1863(9):2389–2397
18. Spagnolo P, Sverzellati N, Rossi G, Cavazza A, Tzouvelekis A, Crestani B et al (2015) Idiopathic pulmonary fibrosis: an update. Ann Med 47(1):15–27
19. Huang SK, Scruggs AM, McEachin RC, White ES, Peters-Golden M (2014) Lung fibroblasts from patients with idiopathic pulmonary fibrosis exhibit genome-wide differences in DNA methylation compared to fibroblasts from nonfibrotic lung. PLoS One 9 (9):e107055
20. Sanders YY, Pardo A, Selman M, Nuovo GJ, Tollefsbol TO, Siegal GP et al (2008) Thy-1 promoter hypermethylation: a novel epigenetic pathogenic mechanism in pulmonary fibrosis. Am J Respir Cell Mol Biol 39(5):610–618
21. Roman J, Mutsaers SE (2018) Epigenetic control of CXCL10: regulating the counterregulator in idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol 58:419
22. Rege TA, Hagood JS (2006) Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. FASEB J 20(8):1045–1054
23. Hagood JS (2019) Thy-1 as an integrator of diverse extracellular signals. Front Cell Dev Biol 7:26
24. Abd-Elmotelb M (2012) Thy-1 (CD90) Expression in dental pulp stem cells: Guy's, King's and St. Thomas's School of Dentistry
25. Sauzay C, Voutetakis K, Chatziioannou A, Chevet E, Avril T (2019) CD90/Thy-1, a cancer-associated cell surface signaling molecule. Front Cell Dev Biol 7:66
26. Bradley JE, Ramirez G, Hagood JS (2009) Roles and regulation of Thy-1, a context-dependent modulator of cell phenotype. Biofactors 35(3):258–265
27. Herrera-Molina R, Valdivia A, Kong M, Alvarez A, Cardenas A, Quest AF et al (2013) Thy-1-interacting molecules and cellular signaling in cis and trans. Int Rev Cell Mol Biol 305:163–216
28. Herrera-Molina R, Valdivia A, Kong M, Alvarez A, Cárdenas A, Quest AF et al (2013) Thy-1-interacting molecules and cellular signaling in cis and trans. Int Rev Cell Mol Biol 305:163–216
29. Rege TA, Pallero MA, Gomez C, Grenett HE, Murphy-Ullrich JE, Hagood JS (2006) Thy-1, via its GPI anchor, modulates Src family kinase and focal adhesion kinase phosphorylation and subcellular localization, and fibroblast migration, in response to thrombospondin-1/hep I. Exp Cell Res 312 (19):3752–3767
30. Sanders YY, Kumbla P, Hagood JS (2007) Enhanced myofibroblastic differentiation and survival in Thy-1(-) lung fibroblasts. Am J Respir Cell Mol Biol 36 (2):226–235
31. Nicola T, Hagood JS, James ML, Macewen MW, Williams TA, Hewitt MM et al (2009) Loss of Thy-1 inhibits alveolar development in the newborn mouse lung. Am J Phys Lung Cell Mol Phys 296(5):L738–L750
32. Yu T, Lao X, Zheng H (2016) Influencing COX-2 activity by COX related pathways in inflammation and cancer. Mini-Rev Med Chem 16(15):1230–1243
33. Sobolewski C, Cerella C, Dicato M, Ghibelli L, Diederich M (2010) The role of cyclooxygenase-2 in cell proliferation and cell death in human malignancies. Int J Cell Biol 2010:215158
34. Liao X, Wang W, Fan C, Yang N, Zhao J, Zhang Y et al (2017) Prokaryotic expression, purification and characterization of human cyclooxygenase-2. Int J Mol Med 40(1):75–82
35. Ghosh N, Chaki R, Mandal V, Mandal SC (2010) COX-2 as a target for cancer chemotherapy. Pharmacol Rep 62(2):233–244
36. Khan Z, Khan N, Tiwari RP, Sah NK, Prasad GB, Bisen PS (2011) Biology of Cox-2: an application in cancer therapeutics. Curr Drug Targets 12 (7):1082–1093
37. Gulyas M, Mattsson JSM, Lindgren A, Ek L, Lamberg Lundström K, Behndig A et al (2018) COX-2 expression and effects of celecoxib in addition to standard chemotherapy in advanced non-small cell lung cancer. Acta Oncol 57(2):244–250
38. Zarghi A, Arfaei S (2011) Selective COX-2 inhibitors: a review of their structure-activity relationships. Iran J Pharmaceut Res 10(4):655–683
39. Tay A, Squire JA, Goldberg H, Skorecki K (1994) Assignment of the human prostaglandin-endoperoxide synthase 2 (PTGS2) gene to 1q25 by fluorescence in situ hybridization. Genomics 23(3):718–719
40. Park GY, Christman JW (2006) Involvement of cyclooxygenase-2 and prostaglandins in the molecular pathogenesis of inflammatory lung diseases. Am J Phys Lung Cell Mol Phys 290(5):L797–L805
41. Piranda DN, Festa-Vasconcellos JS, Amaral LM, Bergmann A, Vianna-Jorge R (2010) Polymorphisms in regulatory regions of cyclooxygenase-2 gene and

breast cancer risk in Brazilians: a case-control study. BMC Cancer 10:613
42. Andersen V, Nimmo E, Krarup HB, Drummond H, Christensen J, Ho GT et al (2011) Cyclooxygenase-2 (COX-2) polymorphisms and risk of inflammatory bowel disease in a Scottish and Danish case-control study. Inflamm Bowel Dis 17(4):937–946
43. Mizuno R, Kawada K (2019) Prostaglandin E2/EP signaling in the tumor microenvironment of colorectal cancer. Int J Mol Sci 20(24):E6254
44. Liu X, Ji Q, Ye N, Sui H, Zhou L, Zhu H et al (2015) Berberine inhibits invasion and metastasis of colorectal cancer cells via COX-2/PGE2 mediated JAK2/STAT3 signaling pathway. PLoS One 10(5):e0123478
45. Bozyk PD, Moore BB (2011) Prostaglandin E2 and the pathogenesis of pulmonary fibrosis. Am J Respir Cell Mol Biol 45(3):445–452
46. Coward WR, Feghali-Bostwick CA, Jenkins G, Knox AJ, Pang L (2014) A central role for G9a and EZH2 in the epigenetic silencing of cyclooxygenase-2 in idiopathic pulmonary fibrosis. FASEB J 28(7):3183–3196
47. Cheng J, Dackor RT, Bradbury JA, Li H, DeGraff LM, Hong LK et al (2016) Contribution of alveolar type II cell-derived cyclooxygenase-2 to basal airway function, lung inflammation, and lung fibrosis. FASEB J 30 (1):160–173
48. Gabasa M, Royo D, Molina-Molina M, Roca-Ferrer J, Pujols L, Picado C et al (2013) Lung myofibroblasts are characterized by down-regulated cyclooxygenase-2 and its main metabolite, prostaglandin E2. PLoS One 8(6):e65445
49. Maher TM, Evans IC, Bottoms SE, Mercer PF, Thorley AJ, Nicholson AG et al (2010) Diminished prostaglandin E2 contributes to the apoptosis paradox in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 182(1):73–82
50. Wilborn J, Crofford LJ, Burdick MD, Kunkel SL, Strieter RM, Peters-Golden M (1995) Cultured lung fibroblasts isolated from patients with idiopathic pulmonary fibrosis have a diminished capacity to synthesize prostaglandin E2 and to express cyclooxygenase-2. J Clin Invest 95(4):1861–1868
51. McAnulty RJ, Hernandez-Rodriguez NA, Mutsaers SE, Coker RK, Laurent GJ (1997) Indomethacin suppresses the anti-proliferative effects of transforming growth factor-beta isoforms on fibroblast cell cultures. Biochem J 321(Pt 3):639–643
52. Xaubet A, Fu W, Li M (2010) A haplotype of cyclooxygenase-2 gene is associated with idiopathic pulmonary fibrosis. Sarcoid Vasc Diff Lung Dis 27 (2):121–130
53. Huang SK, Scruggs AM, Donaghy J, McEachin RC, Fisher AS, Richardson BC et al (2012) Prostaglandin E2 increases fibroblast gene-specific and global DNA methylation via increased DNA methyltransferase expression. FASEB J 26(9):3703–3714
54. Sherr CJ (2006) Divorcing ARF and p53: an unsettled case. Nat Rev Cancer 6(9):663–673
55. Silva J, Silva JM, Dominguez G, Garcia JM, Cantos B, Rodriguez R et al (2003) Concomitant expression of p16INK4a and p14ARF in primary breast cancer and analysis of inactivation mechanisms. J Pathol 199 (3):289–297
56. Cortot AB, Younes M, Martel-Planche G, Guibert B, Isaac S, Souquet PJ et al (2014) Mutation of TP53 and alteration of p14(arf) expression in EGFR- and KRAS-mutated lung adenocarcinomas. Clin Lung Cancer 15 (2):124–130
57. Stott FJ, Bates S, James MC, McConnell BB, Starborg M, Brookes S et al (1998) The alternative product from the human CDKN2A locus, p14ARF, participates in a regulatory feedback loop with p53 and MDM2. EMBO J 17(17):5001–5014
58. Zhang Y, Hyle J, Wright S, Shao Y, Zhao X, Zhang H et al (2019) A cis-element within the ARF locus mediates repression of p16INK4A expression via long-range chromatin interactions. Proc Natl Acad Sci
59. Inoue K, Fry EA (2018) Tumor suppression by the EGR1, DMP1, ARF, p53, and PTEN Network. Cancer Investig 36(9-10):520–536
60. López F, Sampedro T, Llorente JL, Hermsen M, Álvarez-Marcos C (2017) Alterations of p14 ARF, p15 INK4b, and p16 INK4a genes in primary laryngeal squamous cell carcinoma. Pathol Oncol Res 23 (1):63–71
61. Bayramov B, Gunes S, Buyukalpelli R, Aydın O, Henkel R (2018) Promoter methylation analysis of CDH1 and p14ARF genes in patients with urothelial bladder cancer. OncoTarget Ther 11:4189
62. Kotake Y, Naemura M, Murasaki C, Inoue Y, Okamoto H (2015) Transcriptional regulation of the p16 tumor suppressor gene. Anticancer Res 35 (8):4397–4401
63. Flaherty K, Jones D, Yussuf S, Davis S, Huselid E, Wang W et al (2015) Conditional mouse and zebrafish models of INK4-mediated tumor suppression reveal ARF-independent regulation of cellular senescence. Cancer Res 75:Abstract nr 1264
64. Werner LR, Huang S, Francis DM, Armstrong EA, Ma F, Li C et al (2015) Small molecule inhibition of MDM2–p53 interaction augments radiation response in human tumors. Mol Cancer Ther 14(9):1994–2003
65. Fazal L, Ahn M, Bevan L, Buck I, Castro J, Chessari G et al (2018) Development of a potent class of small molecule inhibitors of the MDM2-p53 protein-protein interaction. Cancer Res 78:Abstract nr 1652
66. Yaddanapudi SCS (2016) The human ARF tumor suppressor regulates drosha nucleolar localization and rRNA processing activity: Washington University in St. Louis
67. Joerger AC, Fersht AR (2016) The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches. Annu Rev Biochem 85:375–404
68. Cui L, Zhou F, Chen C, Wang CC (2019) Overexpression of CCDC69 activates p14 ARF/MDM2/p53 pathway and confers cisplatin sensitivity. J Ovar Res 12(1):4

69. Chaar I, Amara S, Elamine OE, Khiari M, Ounissi D, Khalfallah T et al (2014) Biological significance of promoter hypermethylation of p14/ARF gene: relationships to p53 mutational status in Tunisian population with colorectal carcinoma. Tumour Biol 35 (2):1439–1449
70. Delmonico L, dos Santos MA, Franco MF, Esteves EB, Scherrer L, de Moura Gallo CV et al (2015) CDKN2A (p14ARF/p16INK4a) and ATM promoter methylation in patients with impalpable breast lesions. Hum Pathol 46(10):1540–1547
71. Tripathi A, Sharma R, Kejriwal N, Ambasta RK, Kumar P (2016) Epigenetic post transcriptional mutation in neuro-oncology. In: Epigenetic advancements in cancer. Springer, New York, NY, pp 177–205
72. Mansouri A, Karamchandani J, Das S (2017) Molecular genetics of secondary glioblastoma. In: Glioblastoma. Codon Publications, Brisbane, QLD
73. Zhang H, Nie W, Huang F (2015) The correlation relationship between P14ARF gene DNA methylation and primary liver cancer. Med Sci Monit 21:3077
74. Song D, Yang D, Powell CA, Wang X (2019) Cell-cell communication: old mystery and new opportunity. Cell Biol Toxicol 35(2):89–93
75. Wang W, Gao D, Wang X (2018) Can single-cell RNA sequencing crack the mystery of cells? Cell Biol Toxicol 34(1):1–6
76. Zeng Y, Chen X, Gao H, Wang X (2018) An artificial intelligent single cell is part of the cell dream world. Cell Biol Toxicol 34(4):247–249
77. Mirtavoos-Mahyari H, Ghafouri-Fard S, Khosravi A, Motevaseli E, Esfahani-Monfared Z, Seifi S et al (2019) Circulating free DNA concentration as a marker of disease recurrence and metastatic potential in lung cancer. Clin Transl Med 8(1):14
78. El Khoury D, Fayjaloun S, Nassar M, Sahakian J, Aad PY (2019) Updates on the effect of mycotoxins on male reproductive efficiency in mammals. Toxins (Basel) 11(9):E515
79. Sanchez A, Kuras M, Murillo JR, Pla I, Pawlowski K, Szasz AM et al (2019) Novel functional proteins coded by the human genome discovered in metastases of melanoma patients. Cell Biol Toxicol
80. Wang W, Hou J, Zheng N, Wang X, Zhang J (2019) Keeping our eyes on CRISPR: the "Atlas" of gene editing. Cell Biol Toxicol 35(4):285–288
81. Marcell Szasz A, Malm J, Rezeli M, Sugihara Y, Betancourt LH, Rivas D et al (2019) Challenging the heterogeneity of disease presentation in malignant melanoma-impact on patient treatment. Cell Biol Toxicol 35(1):1–14

Shuang Zhou received the Master degree of Genetics, Majoring in Molecular Genetics and Genetic Engineering in Sichuan University. Currently, she serves as a technician and researcher at Clinic Center for Molecular Diagnosis and Therapy, the Second Affiliated Hospital of Fujian Medical University. She is mainly responsible for molecular diagnosis and target-gene detection projects by using QPCR and NGS technology. Meanwhile, through the LC-MS platform, she engages in the research of lipid metabolomics and respiratory diseases such as pulmonary fibrosis. Recently, she undertook 3 projects funded by Fujian provincial as well as municipal Scientific Foundation of China. She is the author of 7 scientific publications with the impact factor about 15.

Xiangdong Wang is a distinguished professor of medicine, director of Shanghai Institute of Clinical Bioinformatics, executive director of Clinical Science Institute of Fudan University Zhongshan Hospital, director of Fudan University Center of Clinical Bioinformatics, deputy director of Shanghai Respiratory Research Institute, and visiting professor of King's College London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 300 scientific publications with the impact factor about 900, citation number about 6920, h-index 48, i10-index 221, and cited journal impact factor about 8000.

Hongzhi Gao is an Associated Professor/Chief physician of Neurosurgery, the Vice Director of Clinic Center for Molecular Diagnosis and Therapy, the Second Affiliated Hospital of Fujian Medical University, and the Enterprise Supervisor of Tianjin University. He has a long-standing interest in the molecular diagnosis, the stem cell therapy, and functional neurosurgery for the nervous systemic disorders, including inherited and infectious disease. Dr. Gao is also an active key researcher in many international cooperative projects of citrin deficiency disease, with many publications.

Yiming Zeng is a Chairman of Academic Committee of the Second Affiliated Hospital of Fujian Medical University, Expert of State Council Expert for Special Allowance, and Director of Sleep Medicine Key Laboratory of Fujian Province. He achieved the Outstanding Contribution of Middle-aged Expert of National Health and Family Planning Commission of the People's Republic of China (NHFPC). His main research is focused on clinical, interventional pulmonology, sleep-breathing disorders, and noninvasive mechanical ventilation. He is the author of more than 150 scientific publications.

Methylation of Inflammatory Cells in Lung Diseases

5

Yifei Liu, Hongzhi Gao, Xiangdong Wang, and Yiming Zeng

Abstract

This chapter overviews roles of DNA methylation in inflammatory cell biology with the focuses on lymphocytes and macrophages/monocytes in lung diseases, although the molecular mechanisms by which target genes are methylated and regulated in lung diseases remain unclear. Most of epigenetic studies on DNA methylation of target genes in lung diseases mainly demonstrated the correlation of DNA methylation of target genes with the levels of other corresponding factors, with the specificity of clinical phenomes, and with the severity of lung diseases. There is an urgent need to identify and validate the specificity and regulatory mechanisms of inflammatory cell epigenetics in depth. The epigenetic heterogeneity among different subsets of T cells and among promoters or non-promoters of target genes should be furthermore clarified in acute or chronic lung diseases and cancers. The hyper/hypo-methylation and modifications of chromosol and extrachromosomal DNA may result in alternations in proteins within inflammatory cells, which can be identified as disease-specific biomarkers and therapeutic targets.

Keywords

Inflammatory cells · Lung disease · Epigenetics · DNA methylation

5.1 Introduction

The disorders of the immune system are regulated by heredity and environment, such as tumors, systemic lupus erythematosus, and asthma. Epigenetics provide new insights for understanding of tumor and autoimmune diseases and lung diseases. Epigenetic alterations contribute to the development of inflammation and immune disorders, and to the instability of phenotypic inheritance and cell differentiation [1]. Epigenetics mainly includes DNA methylation to regulate transcription and expression of target genes, genomic imprinting, and chromatin remodeling (e.g., post-transcriptional regulation, such as non-coding RNA, microRNA, antisense oligonucleotides, and riboswitch RNAs), and post-translational modifications of proteins (e.g.,

Yifei Liu and Hongzhi Gao contributed equally to this work.

Y. Liu · H. Gao
Clinical Center for Molecular Diagnosis and Therapy, Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, China

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China
e-mail: Xiangdong.wang@clintransmed.org

Y. Zeng (✉)
2nd Affiliated Hospital, Fujian Medical University, Quanzhou, Fujian, China

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_5

methylation and acetylation of histones, other modifications of histones, and covalent modifications of non-histones). DNA methylation refers to the process to transfer from the methyl group of *S*-adenosine methionine to the cytosine of CpG island into 5-methylpyrimidine under DNA methyltransferase [2].

The methylation of cytosine after DNA replication changes the conformation of DNA to avoid the binding of DNA to the DNA binding protein, leading to the long-term non-expression activity and silence of those non-coding regions. Genes can be transcribed by a non-methylated promoter, even in adjacent non-transcribed areas where genes are highly methylated. The genetically modified epigenetic modification is transmitted to the daughter cells in the process of somatic cell proliferation stabilized by DNA methyltransferase (Dnmtl), rather than at different stages of embryonic development. Altered levels of DNA methylation are most pronounced in early embryonic development, essential for the development of embryos and the selection of alleles. Changes in DNA methylation can affect the differentiation and reactivity of T cells by regulating cytokines and other related genes, responsible for the pathogenesis of immune diseases [3–5].

Inflammatory cells, e.g., macrophages, mast cells, endothelial cells, lymphocytes, granulocytes, neutrophils, basophils, eosinophils, and other cells, play a vital role in immune response. Among those, T lymphocyte methylation was reported to be related with lung diseases. T cells are quite complex and heterogeneous at different stages of development or different functions of subgroups. T cells are divided into helper T cells (Th) with functions to assist humoral immunity and cellular immunity, suppressor T cells (Ts) to inhibit cellular immunity and humoral immunity, effector T cells (Te) with the function of releasing lymphatic factors, and cytotoxic T cells (Tc) with the function of killing target cells. According to the features and surface marks, T cells can be roughly divided into cytotoxic T cells, T helper cells, and memory T cells [6–9]. Altered methylation of immune cells such as macrophages was found in lung diseases [10].

5.2 The Methylation of Regulatory T cells

Regulatory T cells (Treg) account for 5–10% of peripheral blood CD4+ T cells and are divided into natural Treg (nTreg) and acquired Treg (aTreg) [11]. nTreg comes from thymus gland mainly through the cellular contact mechanism to exert inhibitive function [12], while aTreg is peripheral mature T cells induced by persistent exposure of antigens and cytokines such as IL-10 and TGF-β. Of those signals and regulatory factors, the fork head/winged helix transcription factor (FOXP3) plays the important role in the regulation of nTreg function through the methylation degree of multiple CpG islands related to the expression of FOXP3 gene [13, 14].

The demethylation of exon in the upstream of FOXP3 non-coding region plays a role in maintaining the stability of FOXP3, evidenced by the fact that the damage of Treg cells was related to the increased methylation level of FOXP3 [15, 16]. The exposure to air pollutants induced the occurrence of hyper-methylation at FOXP3 in Treg, leading to the development of asthma [17]. In addition, studies have shown that prolonged exposure to high concentrations of CO, NO_2, and PM2.5 can lead to changes in FOXP3 methylation levels [18]. In another study, decreased methylation of the FOXP3 promoter was associated with increased lung function [19]. Studies have found that high concentration of polycyclic aromatic hydrocarbons exposure can increase the methylation of FOXP3 gene, reduce the expression of FOXP3 and turn Treg into Th2 cells [20], as shown in Fig. 5.1. Folic acid, a methyl donor, could increase the susceptibility to diseases caused by DNA hypermethylation and airway susceptibility [21]. Treg methylation plays an important role in the recovery of lung injury, since DNA methyltransferase inhibitors act on the FOXP3 gene in Treg and

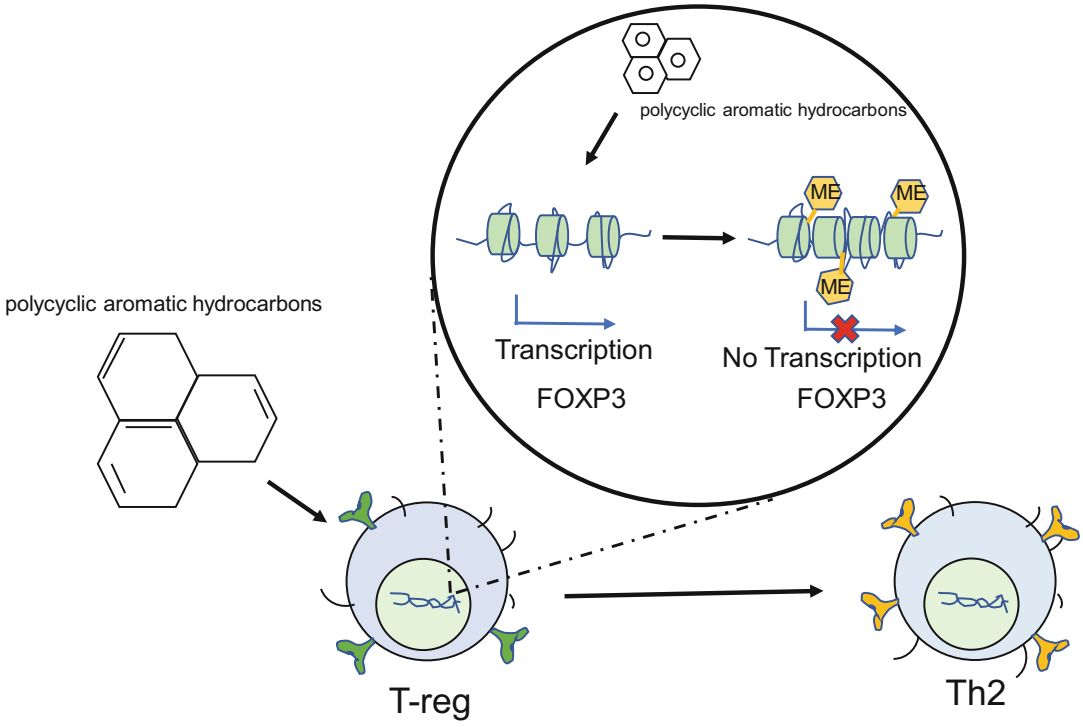


Fig. 5.1 High concentration of polycyclic aromatic hydrocarbons exposure can increase the Increased methylation of FOXP3 gene could reduce the expression of FOXP3 and turn Treg into Th2 cells

increase the number of Treg to accelerate the repair of lung injury [16].

5.3 Th1 DNA Methylation

The differentiation process of Th is controlled by the multiple factors and complex regulation networks of transcription factors. Th1 mainly generates IFN-γ, IL-2, IL-12, and TNF-beta/alpha as the Th1 cytokines, contributing to cellular immunity [22–24]. Th2 mainly produces Th2 cytokines such as IL-4, IL-5, IL-6, and IL-10 [25] to participate the humoral immune response and stimulate the proliferation of B cells and production of antibodies. Those Th cell subsets are characterized by a series of specific gene expressions, including cytokines, cell surface receptors, and regulatory factors.

IFN-γ plays an important role in the regulation of Th1 differentiation, where IFN-γ was produced by natural killer cells (NK). IL-27 and IL-12 from antigen presenting cells (APC) activate STAT1 and STAT4, respectively [26]. The neutralization of IFN-γ prolonged the differentiation of Th1 through signal pathways of STAT1, STAT4, and T-bet [27]. STAT1 and transcription factors are activated by T cell antigen receptor (TCR) together with T-bet, and promote the secretion of IFN-γ and the Th1 phenotype of positive feedback loop. STAT4 is activated by IL-12 signal and is involved in the regulation of the Th1 genetic program. STAT1 and STAT4 also contribute to the regulation of Tbx21 expression [28].

The Th1/Th2 balance is a relatively stable condition under normal circumstances and changes with epigenetic disorders [29, 30]. IFNG as a negative regulator of airway allergic immune response corresponds with IFN-γ as a typical Th1 cytokine, to contribute to the occurrence and development of asthma. The demethylation in the promoter region of the allergen reverse regulating gene and Th1 cytokine IFNG gene was related to the differential

direction of CD4+ T cells [31]. Frequent exposure to trichloroethylene increases the methylation level of IFNG promoter in CD4+ T cells [32]. In 21 pairs of twin patients with asthma, the methylation of FOXP3 gene at 13 CpG loci and IFN-γ gene promoter region at 6 CpG loci increased in Treg and effector T cells [33], as shown in Fig. 5.2. The methylation level of IFNG gene promoter region increased in experimental asthma models induced by antigens, particles, and toxins, e.g., egg albumin induced hypermethylation of IFNG gene and hypo-expression of IFN-γ [34]. However, molecular mechanisms by which IFNG gene promoter region is methylated and the process is regulated in pulmonary diseases remain unclear.

5.4 Th2 DNA Methylation

IL-4 gene promoter regions contain methylation states widely in the CD4+ T cells, while can be demethylated in the allergen-specific T cells, especially when the T cells were stimulated by the allergen. This process contributes to the development of asthma by reversing the hypermethylation of the IFNG gene promoter region, to promoting the binding of Th2-related transcription factors GATA-3 and STAT6 to the corresponding sites, and then inducing a large number of IL-4 production and the differentiation from T cells into Th2 cells [35, 36], as shown in Fig. 5.3. Th2 cells were differentiated and IL4 gene was demethylated in the process of asthma development. For example, maternal exposure to nitrogen dioxide could increase immunoglobulins (Ig) E, airway hyperresponsiveness, respiratory inflammation, and asthma, which was then reduced during postpartum development [37].

IL-13 is produced by TH2 cells. Decreased methylation level in the promoter region of IL-13 increased IL-13 expression and aggravated the lung inflammation in experimental asthma [38]. STAT6 is closely related to Th2 and is activated by IL-4 and IL-13 [39]. DNA methyltransferase inhibitors were found to increase the expression of STAT6 in T cells and influence the differentiation direction of T cells [39]. The expression of IL-6 gene increased in Th2, when the methylation in the promoter region of DNA was reduced [40]. The DNMT inhibitor could also increase the expression of IFNG and decrease methylation in the promoter region of IFNG in CD4+ T cells [41], as explained in Fig. 5.4. Prenatal ETS exposure can lead to a significant increase of house dust mite allergies and inflammatory responses associated with methylation in offspring. The methylation may affect developmental plasticity and programming of T cells and increases the over-production of Th2 cytokines (IL-4, IL-5, and IL-13) during airway remodeling [42].

5.5 Macrophage/Monocyte DNA Methylation

Macrophages and monocytes are phagocytes responsible for the non-specific defense and specific defense and for activation of lymphocytes or other immune cells in lung diseases [43–46]. The altered DNA methylation in alveolar macrophages was associated with smoking [47]. The DNA methylation status of CpG motifs in alveolar macrophages was different between smokers and non-smokers, measured by methylation array and pathway analysis [10]. On basis of the whole genome methylation state of alveolar macrophage DNA sample, methylation analysis and gene expression data, the FLT1 gene was selected as the target gene associated with smoke. Smokers had higher levels of FLT1 gene expression in three major splice variants, which were mainly caused by epigenetic changes in the promoter. This may provide new insights into the pathogenesis of pulmonary diseases. Monocytes derived from hematopoietic stem cells in bone marrow can become macrophages in lung tissues after the settle. The decreased methylation of ZPBP2 in human peripheral blood mononuclear cell will increase the expression of the genes allocated in the 17q12-q21 region and lead to the development of asthma [48–52].

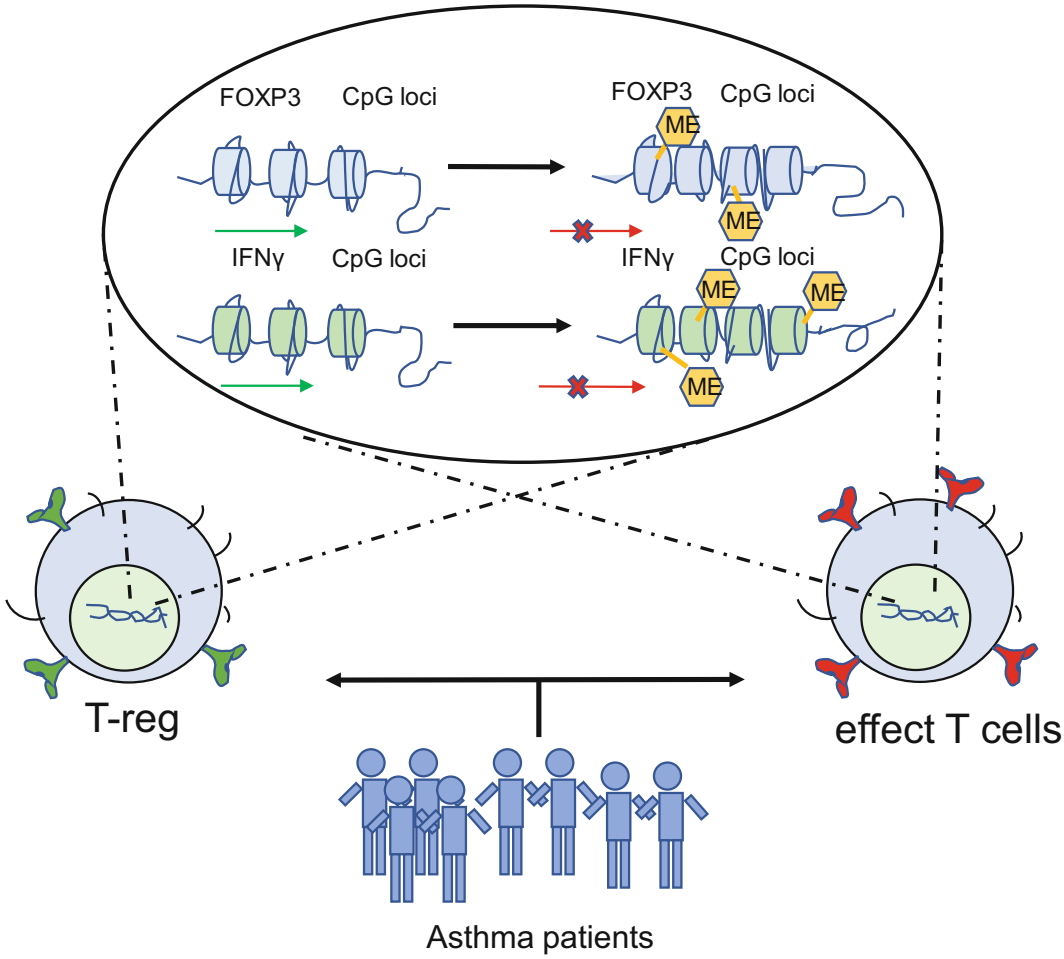


Fig. 5.2 In 21 pairs of twin patients with asthma, the methylation of FOXP3 gene at 13 CpG loci and IFN-γ gene promoter region at 6 CpG loci increased in Treg and effector T cells

5.6 Summary and Perspective

This chapter overviews roles of DNA methylation in inflammatory cell biology with the focuses on lymphocytes and macrophages/monocytes in lung diseases, although the molecular mechanisms by which target genes are methylated and regulated in lung diseases remain unclear. Most of epigenetic studies on DNA methylation of target genes in lung diseases mainly demonstrated the correlation of DNA methylation of target genes with the levels of other corresponding factors, with the specificity of clinical phenomes, and with the severity of lung diseases. There is an urgent need to identify and validate the specificity and regulatory mechanisms of inflammatory cell epigenetics in depth, like targets in other diseases [53–56]. The epigenetic heterogeneity among different subsets of T cells and among promoters or non-promoters of target genes should be furthermore clarified in acute or chronic lung diseases and cancers. The hyper/hypo-methylation and modifications of chromosol and extrachromosomal DNA may result in alternations in proteins within inflammatory cells, which can be identified as disease-specific biomarkers and therapeutic targets [56–62].

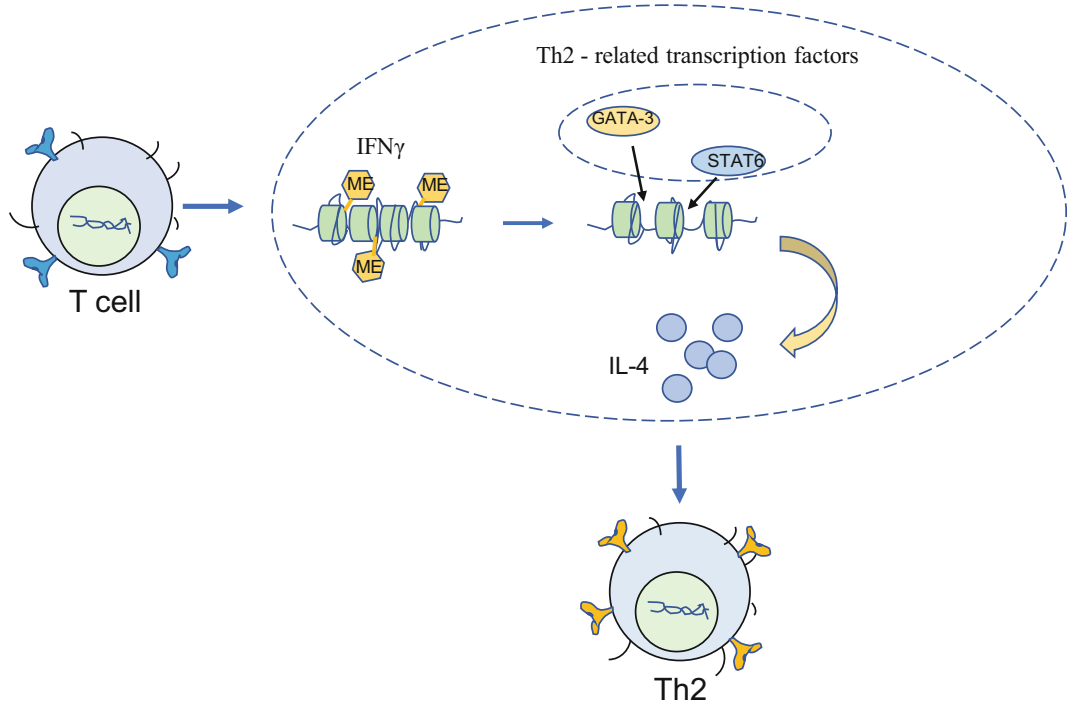


Fig. 5.3 The IL-4 promoter region can be demethylated in allergen-specific T cells. This process contributes to the development of asthma by reversing the hypermethylation of the IFNG gene promoter region, to promoting the binding of Th2-related transcription factors GATA-3 and STAT6 to the corresponding sites, and then inducing a large number of IL-4 production and the differentiation from T cells into Th2 cells

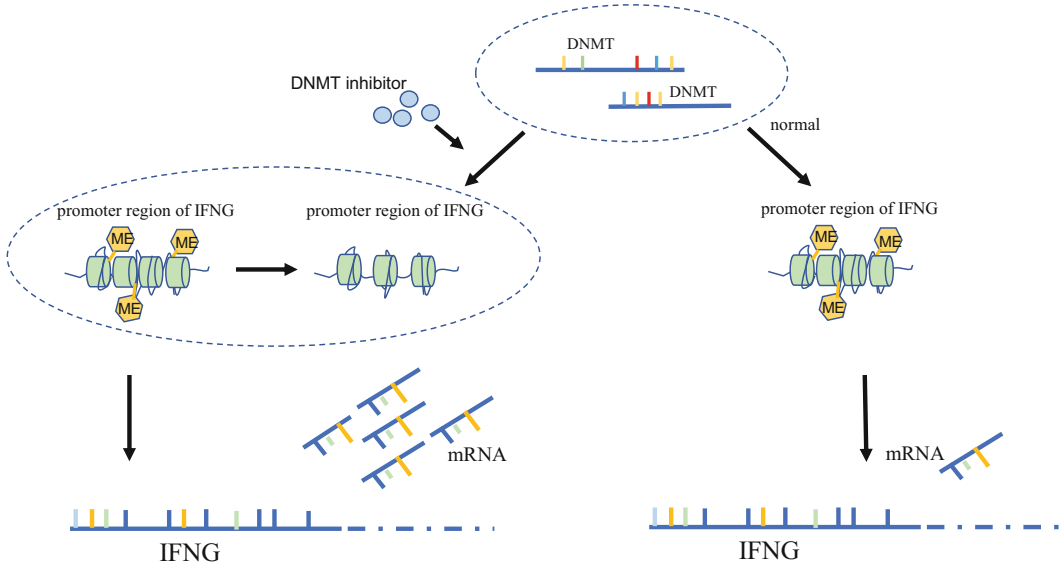


Fig. 5.4 DNMT inhibitor could increase the expression of IFNG and decrease methylation in the promoter region of IFNG in CD4+ T cells

References

1. Suarez-Alvarez B, Baragano Raneros A, Ortega F, Lopez-Larrea C (2013) Epigenetic modulation of the immune function: a potential target for tolerance. Epigenetics 8(7):694–702
2. Koppelman GH, Nawijn MC (2011) Recent advances in the epigenetics and genomics of asthma. Curr Opin Allergy Clin Immunol 11(5):414–419
3. Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M (1990) Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. Arthritis Rheum 33(11):1665–1673
4. Huck S, Deveaud E, Namane A, Zouali M (1999) Abnormal DNA methylation and deoxycytosine-deoxyguanine content in nucleosomes from lymphocytes undergoing apoptosis. FASEB J 13 (11):1415–1422
5. Yung R, Powers D, Johnson K, Amento E, Carr D, Laing T et al (1996) Mechanisms of drug-induced lupus. II. T cells overexpressing lymphocyte function-associated antigen 1 become autoreactive and cause a lupuslike disease in syngeneic mice. J Clin Invest 97(12):2866–2871
6. Barnes PJ (2008) Immunology of asthma and chronic obstructive pulmonary disease. Nat Rev Immunol 8 (3):183–192
7. Romagnani S (1991) Type 1 T helper and type 2 T helper cells: functions, regulation and role in protection and disease. Int J Clin Lab Res 21(2):152–158
8. Clark RA (2015) Resident memory T cells in human health and disease. Sci Transl Med 7(269):269rv1
9. Andersen MH, Schrama D, Thor Straten P, Becker JC (2006) Cytotoxic T cells. J Invest Dermatol 126 (1):32–41
10. Philibert RA, Sears RA, Powers LS, Nash E, Bair T, Gerke AK et al (2012) Coordinated DNA methylation and gene expression changes in smoker alveolar macrophages: specific effects on VEGF receptor 1 expression. J Leukoc Biol 92(3):621–631
11. Morikawa H, Sakaguchi S (2014) Genetic and epigenetic basis of Treg cell development and function: from a FoxP3-centered view to an epigenome-defined view of natural Treg cells. Immunol Rev 259 (1):192–205
12. Leonard JD, Gilmore DC, Dileepan T, Nawrocka WI, Chao JL, Schoenbach MH et al (2017) Identification of natural regulatory T cell epitopes reveals convergence on a dominant autoantigen. Immunity 47(1):107–17. e8
13. Lee SK, Kim JY, Jang BW, Hur SE, Na BJ, Lee M et al (2011) Foxp3(high) and Foxp3(low) Treg cells differentially correlate with T helper 1 and natural killer cells in peripheral blood. Hum Immunol 72 (8):621–626
14. Taylor MD, van der Werf N, Harris A, Graham AL, Bain O, Allen JE et al (2009) Early recruitment of natural CD4+ Foxp3+ Treg cells by infective larvae determines the outcome of filarial infection. Eur J Immunol 39(1):192–206
15. Wang T, Garcia JG, Zhang W (2012) Epigenetic regulation in particulate matter-mediated cardiopulmonary toxicities: a systems biology perspective. Curr Pharmacogen Personal Med 10(4):314–321
16. Singer BD, Mock JR, Aggarwal NR, Garibaldi BT, Sidhaye VK, Florez MA et al (2015) Regulatory T cell DNA methyltransferase inhibition accelerates resolution of lung inflammation. Am J Respir Cell Mol Biol 52(5):641–652
17. Nadeau K, McDonald-Hyman C, Noth EM, Pratt B, Hammond SK, Balmes J et al (2010) Ambient air pollution impairs regulatory T-cell function in asthma. J Allergy Clin Immunol 126(4):845–52.e10
18. Prunicki M, Stell L, Dinakarpandian D, de Planell-Saguer M, Lucas RW, Hammond SK et al (2018) Exposure to NO2, CO, and PM2.5 is linked to regional DNA methylation differences in asthma. Clin Epigenetics 10:2
19. Lovinsky-Desir S, Jung KH, Jezioro JR, Torrone DZ, de Planell-Saguer M, Yan B et al (2017) Physical activity, black carbon exposure, and DNA methylation in the FOXP3 promoter. Clin Epigenetics 9:65
20. Liu J, Zhang L, Winterroth LC, Garcia M, Weiman S, Wong JW et al (2013) Epigenetically mediated pathogenic effects of phenanthrene on regulatory T cells. J Toxicol 2013:967029
21. Prescott SL, Clifton V (2009) Asthma and pregnancy: emerging evidence of epigenetic interactions in utero. Curr Opin Allergy Clin Immunol 9(5):417–426
22. Asosingh K, Swaidani S, Aronica M, Erzurum SC (2007) Th1- and Th2-dependent endothelial progenitor cell recruitment and angiogenic switch in asthma. J Immunol 178(10):6482–6494
23. Hu WK, Lu XX, Yang S, Xu GP, Lan F, Chen SX et al (2009) Expression of the Th1-specific cell-surface protein Tim-3 increases in a murine model of atopic asthma. J Asthma 46(9):872–877
24. Chakraborty S, Panda AK, Bose S, Roy D, Kajal K, Guha D et al (2017) Transcriptional regulation of FOXP3 requires integrated activation of both promoter and CNS regions in tumor-induced CD8(+) Treg cells. Sci Rep 7(1):1628
25. Selvaraja M, Abdullah M, Arip M, Chin VK, Shah A, Amin NS (2019) Elevated interleukin-25 and its association to Th2 cytokines in systemic lupus erythematosus with lupus nephritis. PLoS One 14 (11):e0224707
26. Zhu J, Yamane H, Paul WE (2010) Differentiation of effector CD4 T cell populations (∗). Annu Rev Immunol 28:445–489
27. Zhu J, Paul WE (2010) Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. Immunol Rev 238(1):247–262
28. Schulz EG, Mariani L, Radbruch A, Hofer T (2009) Sequential polarization and imprinting of type 1 T helper lymphocytes by interferon-gamma and interleukin-12. Immunity 30(5):673–683

29. Begin P, Nadeau KC (2014) Epigenetic regulation of asthma and allergic disease. Allergy, Asthma Clin Immunol 10(1):27
30. Sanders VM (2006) Epigenetic regulation of Th1 and Th2 cell development. Brain Behav Immun 20 (4):317–324
31. White GP, Hollams EM, Yerkovich ST, Bosco A, Holt BJ, Bassami MR et al (2006) CpG methylation patterns in the IFNgamma promoter in naive T cells: variations during Th1 and Th2 differentiation and between atopics and non-atopics. Pediatr Allergy Immunol 17(8):557–564
32. Gilbert KM, Blossom SJ, Erickson SW, Broadfoot B, West K, Bai S et al (2016) Chronic exposure to trichloroethylene increases DNA methylation of the Ifng promoter in CD4(+) T cells. Toxicol Lett 260:1–7
33. Runyon RS, Cachola LM, Rajeshuni N, Hunter T, Garcia M, Ahn R et al (2012) Asthma discordance in twins is linked to epigenetic modifications of T cells. PLoS One 7(11):e48796
34. Krauss-Etschmann S, Meyer KF, Dehmel S, Hylkema MN (2015) Inter- and transgenerational epigenetic inheritance: evidence in asthma and COPD? Clin Epigenetics 7:53
35. Kwon NH, Kim JS, Lee JY, Oh MJ, Choi DC (2008) DNA methylation and the expression of IL-4 and IFN-gamma promoter genes in patients with bronchial asthma. J Clin Immunol 28(2):139–146
36. Agarwal S, Rao A (1998) Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity 9(6):765–775
37. Yue H, Yan W, Ji X, Gao R, Ma J, Rao Z et al (2017) Maternal exposure of BALB/c mice to indoor NO2 and allergic asthma syndrome in offspring at adulthood with evaluation of DNA methylation associated Th2 polarization. Environ Health Perspect 125 (9):097011
38. Yu Q, Zhou B, Zhang Y, Nguyen ET, Du J, Glosson NL et al (2012) DNA methyltransferase 3a limits the expression of interleukin-13 in T helper 2 cells and allergic airway inflammation. Proc Natl Acad Sci U S A 109(2):541–546
39. Kim EG, Shin HJ, Lee CG, Park HY, Kim YK, Park HW et al (2010) DNA methylation and not allelic variation regulates STAT6 expression in human T cells. Clin Exp Med 10(3):143–152
40. Nile CJ, Read RC, Akil M, Duff GW, Wilson AG (2008) Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis Rheum 58 (9):2686–2693
41. Brand S, Kesper DA, Teich R, Kilic-Niebergall E, Pinkenburg O, Bothur E et al (2012) DNA methylation of TH1/TH2 cytokine genes affects sensitization and progress of experimental asthma. J Allergy Clin Immunol 129(6):1602–10.e6
42. Christensen S, Jaffar Z, Cole E, Porter V, Ferrini M, Postma B et al (2017) Prenatal environmental tobacco smoke exposure increases allergic asthma risk with methylation changes in mice. Environ Mol Mutagen 58(6):423–433
43. Brain JD (1980) Pulmonary macrophages: when do they prevent and when do they cause COPD? Chest 77(2 Suppl):264–265
44. Tager M, Piecyk A, Kohnlein T, Thiel U, Ansorge S, Welte T (2000) Evidence of a defective thiol status of alveolar macrophages from COPD patients and smokers. Chronic obstructive pulmonary disease. Free Radic Biol Med 29(11):1160–1165
45. Tetley TD (2002) Macrophages and the pathogenesis of COPD. Chest 121(5 Suppl):156S–159S
46. Ishii T, Wallace AM, Zhang X, Gosselink J, Abboud RT, English JC et al (2006) Stability of housekeeping genes in alveolar macrophages from COPD patients. Eur Respir J 27(2):300–306
47. Droemann D, Goldmann T, Tiedje T, Zabel P, Dalhoff K, Schaaf B (2005) Toll-like receptor 2 expression is decreased on alveolar macrophages in cigarette smokers and COPD patients. Respir Res 6:68
48. Al Tuwaijri A, Gagne-Ouellet V, Madore AM, Laprise C, Naumova AK (2016) Local genotype influences DNA methylation at two asthma-associated regions, 5q31 and 17q21, in a founder effect population. J Med Genet 53(4):232–241
49. Naumova AK, Leppert M, Barker DF, Morgan K, Sapienza C (1998) Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. Am J Hum Genet 62 (6):1493–1499
50. Kanagaratham C, Chiwara V, Ho B, Moussette S, Youssef M, Venuto D et al (2018) Loss of the zona pellucida-binding protein 2 (Zpbp2) gene in mice impacts airway hypersensitivity and lung lipid metabolism in a sex-dependent fashion. Mamm Genome 29 (3-4):281–298
51. Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Lariviere M, Moussette S et al (2009) Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet 85 (3):377–393
52. Naumova AK, Olien L, Bird LM, Slamka C, Fonseca M, Verner AE et al (1995) Transmission-ratio distortion of X chromosomes among male offspring of females with skewed X-inactivation. Dev Genet 17(3):198–205
53. Toh TB, Lim JJ, Chow EK (2019) Epigenetics of hepatocellular carcinoma. Clin Transl Med 8(1):13
54. Chirshev E, Oberg KC, Ioffe YJ, Unternaehrer JJ (2019) Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. Clin Transl Med 8(1):24
55. Luo Y (2019) Refining CRISPR-based genome and epigenome editing off-targets. Cell Biol Toxicol 35 (4):281–283
56. Gil J, Betancourt LH, Pla I, Sanchez A, Appelqvist R, Miliotis T et al (2019) Clinical protein science in

translational medicine targeting malignant melanoma. Cell Biol Toxicol 35(4):293–332
57. Wu D, Cheng Y, Wang X (2019) CSGT group. Definition of clinical gene tests. Cell Biol Toxicol 35 (2):83–87
58. Kong S, Zhang Y (2019) Deciphering Hi-C: from 3D genome to function. Cell Biol Toxicol 35(1):15–32
59. Ghufran MS, Soni P, Kanade SR (2019) Aflatoxin-induced upregulation of protein arginine methyltransferase 5 is mediated by protein kinase C and extracellular signal-regulated kinase. Cell Biol Toxicol 35(1):67–80
60. Ansari D, Toren W, Zhou Q, Hu D, Andersson R (2019) Proteomic and genomic profiling of pancreatic cancer. Cell Biol Toxicol 35(4):333–343
61. Mirtavoos-Mahyari H, Ghafouri-Fard S, Khosravi A, Motevaseli E, Esfahani-Monfared Z, Seifi S et al (2019) Circulating free DNA concentration as a marker of disease recurrence and metastatic potential in lung cancer. Clin Transl Med 8(1):14
62. Gong J, Chehrazi-Raffle A, Placencio-Hickok V, Guan M, Hendifar A, Salgia R (2019) The gut microbiome and response to immune checkpoint inhibitors: preclinical and clinical strategies. Clin Transl Med 8(1):9

Yifei Liu, Master of Science, Researcher of Clinical Center for Molecular Diagnosis and Therapy of the Second Affiliated Hospital of Fujian medical University. His main research is focuses on bioinformatics and data mining, with a particular focus on tumor-related superenhancers.

Hongzhi Gao is an Associated Professor/Chief physician of Neurosurgery, the Vice Director of Clinical Center for Molecular Diagnosis and Therapy, The Second Affiliated Hospital of Fujian Medical University, and the Enterprise Supervisor of Tianjin University. He has a long-standing interest in the molecular diagnosis, the stem cell therapy, and functional neurosurgery for the nervous systemic disorders, including inherited and infectious disease. Dr. Gao is also an active key researcher in many international cooperative projects of citrin deficiency disease, with many publications.

Xiangdong Wang is a distinguished professor of medicine, director of Shanghai Institute of Clinical Bioinformatics, executive director of Clinical Science Institute of Fudan University Zhongshan Hospital, director of Fudan University Center of Clinical Bioinformatics, deputy director of Shanghai Respiratory Research Institute, and visiting professor of King's College London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 300 scientific publications with the impact factor about 900, citation number about 6920, h-index 48, i10-index 221, and cited journal impact factor about 8000.

Yiming Zeng is a Chairman of Academic Committee of the Second Affiliated Hospital of Fujian Medical University, Expert of State Council Expert for Special Allowance, and Director of Sleep Medicine Key Laboratory of Fujian Province. He achieved the Outstanding Contribution of Middle-aged Expert of National Health and Family Planning Commission of the People's Republic of China (NHFPC). His main research is focused on clinical, interventional pulmonology, sleep-breathing disorders, and noninvasive mechanical ventilation. He is the author of more than 150 scientific publications.

Research Advances on DNA Methylation in Idiopathic Pulmonary Fibrosis

6

Qian-Kun Luo, Hui Zhang, and Li Li

Abstract

Idiopathic pulmonary fibrosis (IPF) is a chronic complex lung disease with no specific treatment and poor prognosis, characterized by the pulmonary progressive fibrosis and dysfunctions that lead to respiratory failure. Several factors may impact the progress of IPF, including age, cigarette smoking, and dusts, of which genetic and epigenetic factors mainly contribute to lung tissue fibrosis. DNA methylation is one of epigenetic processes that occur in many diseases and regulate chromosomal and extrachromosomal DNA functions in response to environmental exposures. The methylation plays pivotal roles in regulation of gene expression to facilitate the formation of fibroblastic foci and lung fibrosis. This chapter will describe alterations and effects of the DNA methylation on gene expression, the potential application of DNA methylation as a biomarker, and significance as therapeutic targets. Those understanding will provide us new insight into the treatment and prognosis of IPF.

Keywords

DNA methylation · Genetic variants · Idiopathic pulmonary fibrosis · Epigenetic

Q.-K. Luo
Zhengzhou University People's Hospital, Henan Provincial People's Hospital, Zhengzhou, Henan, China

H. Zhang
Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

L. Li (✉)
Department of Scientific Research and Discipline Construction, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China
e-mail: lili@henu.edu.cn

6.1 Introduction

Idiopathic pulmonary fibrosis (IPF) is an interstitial lung disease with progressive proliferation of fibroblasts and accumulation of fibrotic tissue that is associated with age, environmental and microbial exposures, and genetic and epigenetic changes [1]. Many factors contribute to the formation of IPF, including repeated microinjury of an aging, dysfunctional alveolar epithelium, epithelial-fibroblast communication, extracellular matrix over-production, myofibroblast proliferation, and lung tissue remodeling [1–7]. Morbidity and mortality increase obviously as the

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_6

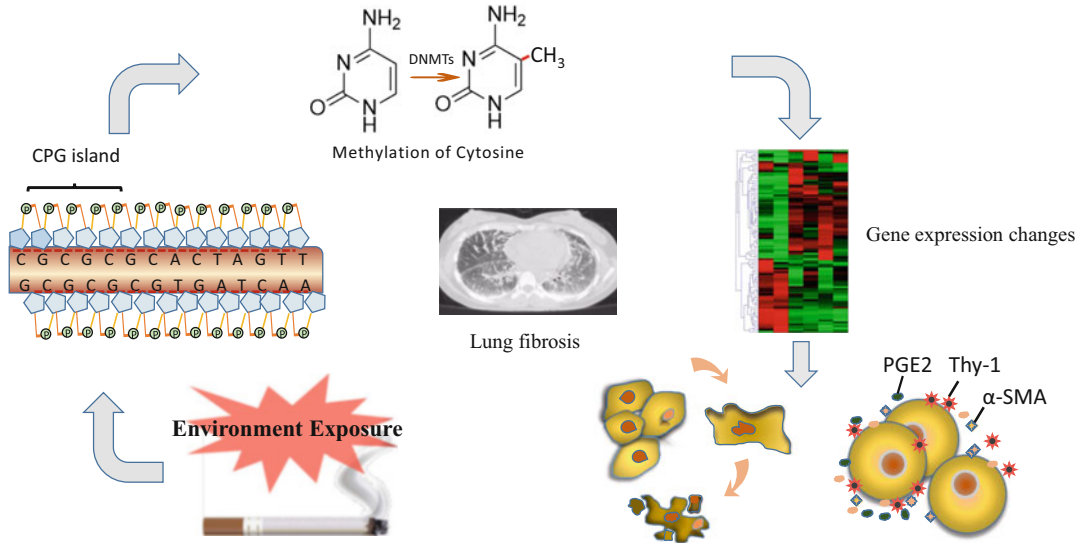


Fig. 6.1 Methylation of the fifth carbon of cytosine in CpG island has effects on gene expressions which affect cell proliferation, apoptosis, differentiation, and synthesis of many factors

population ages [3]. Studies have verified correlations between exposure to cigarette smoking, wood, metal dust, stone, silica, and microbial agents in patients with IPF. In addition to the environmental factors, genetic variation has been investigated as a potent risk factor for developing lung fibrosis. Some patients with variants may have a better survival, while others with genetic mutations may be susceptible to familial interstitial pneumonia. For example, the MUC5B promoter variant was considered an important genetic risk for IPF, although the mechanisms by which genetic mutations are associated with IPF remain unclear [4–7].

DNA methylation is altered in many diseases and often involved with environmental exposures. Studies demonstrated that cigarette smoking and aging have significant impacts on epigenetic alterations, revealing a relationship between IPF and DNA methylation [8]. Methylation of DNA modifies the fifth carbon of cytosine residues (CpG dinucleotides) with a methyl group transferred from *S*-adenyl methionine by DNA methyltransferases (Fig. 6.1). Methylation of CpG islands in gene bodies and promoters has significant effects on gene expression and development of diseases. Clinical studies demonstrated the underlying functions for DNA methylation as a biomarker for prognosis of cancers and others [9]. Studies on genome-wide methylation indicated that there were 2130 differentially methylated regions (DMRs) between IPF lung tissues and control samples [10]. Yan Y. Sanders et al. also identified that 870 genes were differentially methylated in IPF lung tissue compared to normal lung tissue [11]. Methylation in those DMRs may regulate expression of many target genes and miRNAs and the regulatory sites in genes involved in IPF [12, 13]. Previous studies defined molecular processes affected by DNA methylation in IPF and mechanisms of how DNA methylation prompts the development of IPF.

6.2 Genome-Wide DNA Methylation in IPF

To explore the DNA methylation changes in IPF, genome-wide DNA methylation has been carried out in several studies recently. Rabinovich et al. studied genomic DNA methylation profiles in IPF lung tissue, lung adenocarcinomas, and control lung tissue and found 625 distinct DMRs in IPF compared to normal lung tissue [14]. Only a small fraction of differentially methylated CpG islands was in promoters, where the DNA methylation

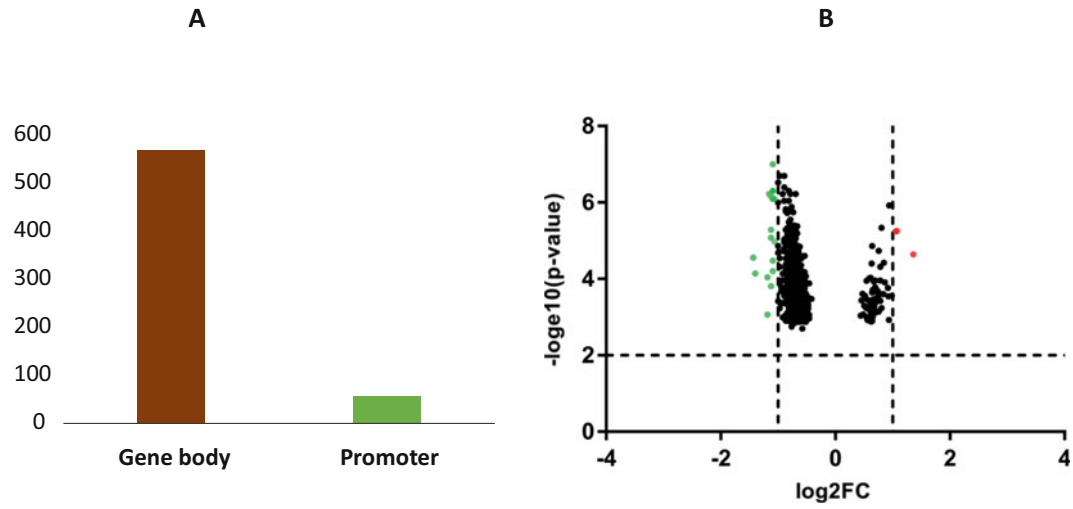


Fig. 6.2 (**a**) Differentially methylated CpG islands are located in gene bodies such as intron, exon, and intergenic. Only 8.9% DMRs are located in promoter. (**b**) Volcano plot of 625 hypomethylated CpG islands in IPF. The *X axis* differential methylation is represented by mean fold change. *P*-value (*Y axis*) is adjusted by −log10. Green dots are the significantly hypomethylated genes, red dots are the hypermethylated genes. (This figure was modified from REF. [14], Plos One.)

marks were related to biological processes, including gene expression, development of cancer, and cell proliferation and apoptosis (Fig. 6.2). However, effects of methylation in non-promoter regions remain unknown. The study also validated that the expressions of Serine/Threonine Kinase 17b and histone cluster 1 H2ah were upregulated in IPF, while their promoters were hypomethylated [14].

Another study analyzed genome-wide DNA methylation in 12 IPF patients compared with seven normal control patients by DNA methylation microarray and found that 870 genes were differentially methylated in IPF [11]. RNA expression array analysis was used to verify DNA methylation-regulated gene expression. There were eight genes significantly regulated by DNA methylation that were associated with IPF [11]. However, many differentially methylated genes were not correlated with gene expression levels because of the small sample. This particular study detected CpG islands and the CpG island shores, which were shown to be associated with gene expression, specified histology, and stem cells in cancer. Huang et al. cultured fibroblasts obtained from IPF lung

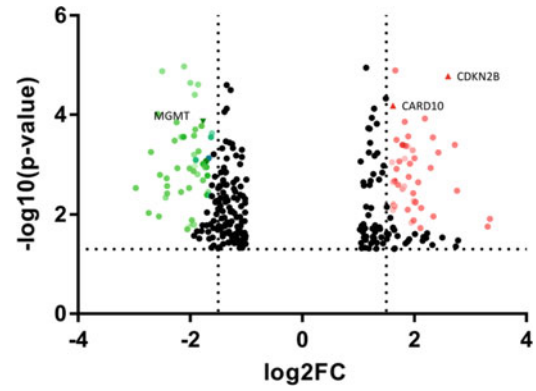


Fig. 6.3 Differentially methylated CpG loci between IPF cells and control cell lines. Three genes (CDKN2B, CARD10, MGMT) were identified to be associated with the expression of genes contributing to the proliferation of IPF fibroblasts. (This figure was modified from REF. [15], Plos One.)

biopsies and detected the number of CpG sites differentially methylated in IPF as compared to control cells [30]. The differentially methylated genes play important roles in proliferation of fibroblasts and production of extracellular matrix (ECM). Of those, three genes were identified with altered expression according to DNA methylation (Fig. 6.3), although the mechanism of DNA

methylation regulating gene expression was unclear. A study assessed DNA methylation in approximately 100 IPF lung tissue samples and detected CpG island shores and 2130 DMRs with significance, among which 43% are hypermethylated in IPF lung tissues after controlling environmental effects and other biases. Most of the DMRs are located in gene bodies and CpG island shores. Several novel genes including CASZ1, SOX7, and TRIM71 were furthermore validated for the association of those genes with IPF. CASZ1 was found to be a strong candidate gene for association with IPF [10].

6.3 DNA Methylation of Gene Promoters in IPF

Many targeted genes were identified to be methylated in gene promoter and the expression of gene or miRNAs were modified in IPF recently, as listed in Table 6.1. The mucin 5b (MUC5B) gene, which contains variants, is suggested as an important risk factor for the development of IPF. The variant (rs35705950) in the promoter region of MUC5B was found to be related to familial IPF [16]. The mechanism by which variant rs35705950 contributes to pulmonary fibrosis has been a hot area for study in recent years [17]. The study on DNA methylation uncovered the methylated target region of the MUC5B variant by mass spectrometry-based methods. Several hypermethylated or hypomethylated DMRs were identified to be associated with IPF. The target FOXA2 binding site (Chr11: 1241254-1241270) in a conserved region (Chr11: 1241262-1241269) had the strongest effect on MUC5B expression. The methylation surrounding the FOXA2 motif may alter the binding of FOXA2 and regulate the coaction of the MUC5B promoter variant and transcription factors, leading to the overexpression of MUC5B and MUC5B-dominated bio-function

Table 6.1 Current investigations of DNA methylation in idiopathic pulmonary fibrosis. Gene expression is upregulated (↑) or downregulated (↓) in association with DNA methylation

Targeted genes	Mechanism	Regulation status	Samples	Contribution
CASZ1 [10]	Hypomethylation	↑	IPF lung tissue, human alveolar cell line	Fibroproliferation
STK17B, STK3, and HIST1H2AH [14]	Hypomethylation	↑	IPF lung tissue	Cell apoptosis and histone modification
MGMT/CDKN2B, CARD10 [15]	Hypomethylation/ Hypermethylation	↑/↓	Human lung fibroblast	Cell proliferation
ZNF467, CLDN5/ TP53INP1, DDAH1 [11]	Hypermethylation/ Hypomethylation	↓/↑	IPF lung tissue	Pathogenesis of IPF
MUC5B [12]	Methylation around FOXA2 binding region	↑	Human lung fibroblast	Familial IPF
COX2 [13]	Hypermethylation of c8orf4 promoter	↓	Human lung fibroblast	Antifibrotic
TGF-β/BMP [21]	Demethylation of BMPER	↑/↓	Human lung fibroblast and mice cell line	Lung fibrosis
MiRNA-17~92	DNA methylation	↓	IPF lung tissue and human lung fibroblast	Myofibroblast differentiation
Thy-1 [23–25]	Hypermethylation	↓	Human lung fibroblast	Myofibroblast differentiation
P14 (ARF) [26]	Hypermethylation	↓	Human lung fibroblast	Proapoptotic
α-SMA [27, 28]	Hypermethylation	↓	Mice lung tissue and mice alveolar epithelial cell	Myofibroblast differentiation
PGE_2 [29, 30]	Hypermethylation of PTGER2	PGE_2 resistance	Mice lung tissue and human lung fibroblast	Lung fibrosis

[12]. The study indicates that the FOXA2 binding may be one of critical factors in the composite regulatory network to control MUC5B expression. Other transcription factors, such as HOXA9, STAT3, and ZBTB7A, may regulate MUC5B expression through the interaction with FOXA2.

Keerthisingam C.B. et al. reported that expression changes of cyclooxygenase-2 (COX2) and prostaglandin (PG) E2 were associated with lung fibrotic responses [18]. The DNA methylation could promote COX2 silencing in cancer and may decrease COX2 and PGE_2 expression in fibrotic lung fibroblasts. However, COX2 expression was not regulated directly by methylation of its promoter, evidenced by the fact that the G9a and enhancer of zeste homolog 2 (EZH2) may silence COX2 through interactions with histone hypermethylation [19]. Another report demonstrated that the hypermethylation of the c8orf4 factor might alter COX2 expression and promote the development of lung fibrosis [13], although the specific regulatory mechanism is still under exploration. Bone morphogenetic proteins (BMPs) and transforming growth factor-β (TGF-β) were verified as key factors in inducing lung fibrosis [20]. The TGF-β/BMP pathway could regulate proliferation and transdifferentiation of fibroblasts in response to lung injury. The BMP endothelial cell precursor-derived regulator (BMPER) is a strong factor to affect the activity of TGF-β/BMP pathway in IPF. BMPER is found to have effects on the invasion and migration of lung fibroblasts. The DNA methylation-induced demethylation can downregulate BMPER promoter activity and expression [21], where the target sequence and BMPER promoter methylation are not detected.

miRNAs can regulate gene expression by interfering with transcription or by degrading target mRNAs, since miRNA expression altered in IPF. The miRNA-17~92 cluster is downregulated in fibrotic lung tissue to probably target fibrotic genes, including TGF-β, collagen, metalloproteinases, and DNMT-1. The DNA methylation of the miRNA-17~92 cluster promoter can silence its expression and be restored with 5′-aza-2′-deoxycytidine (5-aza) through the reduction of the synthesis of DNA methyltransferase (DNMT)-1 and genes that promote myofibroblast differentiation [22]. Thymocyte differentiation antigen-1 (Thy-1) regulates cell–cell and cell–matrix interactions, by which myofibroblast production may be associated with lung fibrosis development. Sanders YY et al. have found that hypermethylation and downregulation of Thy-1 occur in IPF and that suppression of Thy-1 expression regulated by promoter methylation can be restored by the DNMT-1 inhibitor 5-aza [23, 24]. Another study investigated the effect of hypoxia on lung fibroblasts and found that the Thy-1 promoter is highly methylated in hypoxic fibroblasts, while Thy-1 expression is downregulated [25]. Moreover, the proapoptotic gene P14 was also identified as hypermethylated in its promoter CpG islands in IPF lung tissue [26].

Myofibroblasts are important in ECM accumulation and fibrotic cytokine production in lung fibrosis. α-smooth muscle actin (α-SMA) is a pivotal factor for the production of myofibroblasts from fibroblasts. The expression of α-SMA was associated with methylation of CpG islands in its promoter. siRNA and 5-aza can inhibit DNMT activity and induce the synthesis of α-SMA, while overexpression of DNMT downregulates α-SMA production [27]. α-SMA expression is also regulated by methyl CpG binding protein 2 (MeCP2) and mild lung fibrosis occurred in MeCP2 mutant mice [28]. It strongly suggests that DNA methylation regulated by DNMT plays an important role in the expression of α-SMA and fibrogenesis. Prostaglandin E_2 (PGE_2) was downregulated in lung lavage fluid of IPF patients with poor prognosis at lower levels of PGE_2. PGE_2 has been verified to have antifibrotic effects in previous studies. Huang SK et al. reported that PGE_2 resistance in IPF is correlated with decreased expression of E prostanoid 2 receptor (EP2), and hypermethylation of the prostaglandin E receptor 2 gene (PTGER2) promoter is a key factor inducing PGE_2 resistance [29, 30].

6.4 Perspectives

The cause of IPF was considered to be chronic inflammation that established lung parenchymal fibrosis, while anti-inflammatory and immunosuppressive therapies fail to significantly improve the prognosis of patients. Repetitive and sustained lung injuries associated with environment exposures, genetic mutation, and aging are regarded as the major mechanisms bringing about lung fibrosis [2, 31]. However, the mechanisms underlying myofibroblast differentiation from fibroblasts and accumulation and proliferation of these cells remain unknown. Many studies on fibroblast differentiation and gene expression emerged in recent years, including studies on genome-wide association and epigenetic alterations [17, 32]. In addition, telocytes were identified to possibly affect the cell–cell interactions and apoptosis through integrins-PI3K pathways in the lung [33, 34].

The epigenetic modulation has the significant impact in pathogenesis of many diseases through the comprehensive processes and mechanisms detected by genomics, proteomics, as well as clinical transomics [35–39]. Discovery of CpG island shores and histone variants brought a new perspective in gene regulation and epigenetic mechanisms [40, 41]. DNA methylation can be a dynamic biomarker for biological responses to environment exposures, dietary habits, cigarette smoking, drinking, and age. The alteration of DNA methylation in IPF influenced by multiple risk factors is more significant. With the development of DNA methylation assays, it is questioned whether DNA methylation can meet the criteria of disease-specific biomarkers for clinical application to dynamically monitor disease severity, duration, phase, and response to therapy like others [42, 43]. Genome-wide DNA methylation analysis allows researchers to easily measure DNA methylation in large samples and decipher the mechanisms by which DNA methylation regulates gene expression, methylated CpG islands in promoter regions of genes. These discoveries may bring us novel therapeutic strategies or diagnostic biomarkers to improve the prognosis of IPF patients. However, current studies on simplex lung tissues may lack accuracy because of the complex pathogenesis of lung fibrosis and the dynamic nature of epigenetic features. Environmental effects were taken into consideration in their genome-wide DNA methylation study on alveolar lavage and blood in different stages of disease [10, 44].

DNA methylation was majorly investigated in bulk cells obtained from entire lung tissue, including alveolar epithelial and endothelial cells, which may contain many specific epigenetic signals in different cell types. Techniques have been applied to deconstruct the tissue into cell-specific components [45]. For example, cigarette exposure is confirmed to play a critical role in epigenetic alteration in IPF, while it is a great challenge to correlate cigarette exposure and epigenetic alterations in IPF on account of the late disease stage when tissues are taken from lungs. In addition, the cigarette factor is associated with the development of many diseases, such as pulmonary fibrosis, chronic obstructive pulmonary disease, and lung cancer [46, 47]. It is hard to distinguish which DNA methylation changes affected by cigarette smoke contribute specifically to the development of IPF.

Another challenge is to combine DNA methylation with other epigenetic mechanisms, such as noncoding RNAs, histone modifications, and chromatin remodeling. A key reason for poor overlap between DNA methylation and genetic alteration may be the complex interactions of these epigenetic mechanisms. The combination of expression quantitative trait loci (eQTLs) and genome-wide association (GWA) studies provides us a new strategy to identify methylation marks that regulate gene expression and genetic variants based on DNA methylation (methyl-QTL) [10, 48, 49]. Furthermore, 5-aza has now been applied as a therapy for lung cancers associated with histone deacetylase [50].

In summary, dysfunction of DNA methylation is one of critical dominants in the development of IPF, although there are still a large number of regulatory processes to be explored. The deep understanding of DNA methylation will benefit the discovery and development of a new

therapeutic strategy for IPF treatment and novel prognostic tools.

Acknowledgement *Conflicts of Interest Statement*: Qian-kun Luo, Hui Zhang, and Li Li have no conflicts of interest.

References

1. Sgalla G, Lovene B, Calvelo M et al (2018) Idiopathic pulmonary fibrosis: pathogenesis and management. Respir Res 19(1):32. https://doi.org/10.1186/s12931-018-0730-2
2. Richeldi L, Collard HR, Jones MG (2017) Idiopathic pulmonary fibrosis. Lancet 89(10082):1941–1952. https://doi.org/10.1186/s12931-018-0730-2
3. Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG et al (2015) Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. Nature 517(7536):621–625. https://doi.org/10.1038/nature14112
4. Fang T, Wang M, Xiao H, Wei X (2019) Mitochondrial dysfunction and chronic lung disease. Cell Biol Toxicol 35(6):493–502. https://doi.org/10.1007/s10565-019-09473-9
5. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE et al (2017) Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. Lancet Respir Med 5(11):869–880. https://doi.org/10.1016/S2213-2600(17)30387-9
6. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG et al (2013) Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA 309 (21):2232–2239. https://doi.org/10.1001/jama.2013.5827
7. Fukuda S, Enkhbaatar P, Nelson C, Cox RA, Wolfson MR, Shaffer TH et al (2018) Lack of durable protection against cotton smoke-induced acute lung injury in sheep by nebulized single chain urokinase plasminogen activator or tissue plasminogen activator. Clin Transl Med 7(1):17. https://doi.org/10.1186/s40169-018-0196-3
8. Liu F, Killian JK, Yang M, Walker RL, Hong JA et al (2010) Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. Oncogene 29(25):3650–3664. https://doi.org/10.1038/onc
9. Bock C, Halbritter F, Carmona FJ, Tierling S et al (2016) Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol 34(7):726–737. https://doi.org/10.1038/nbt.3605
10. Yang IV, Pedersen BS, Rabinovich E, Hennessy CE et al (2014) Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 190(11):1263–1272. https://doi.org/10.1164/rccm.201408-1452oc
11. Sanders YY, Ambalavanan N, Halloran B et al (2012) Altered DNA methylation profile in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 186 (6):525–535. https://doi.org/10.1164/rccm.201201-0077oc
12. Kawamura Y, Takouda J, Yoshimoto K, Nakashima K (2018) New aspects of glioblastoma multiforme revealed by similarities between neural and glioblastoma stem cells. Cell Biol Toxicol 34(6):425–440. https://doi.org/10.1007/s10565-017-9420-y
13. Evans IC, Barnes JL, Garner IM et al (2016) Epigenetic regulation of cyclooxygenase-2 by methylation of c8orf4 in pulmonary fibrosis. Clin Sci 130 (8):575–586. https://doi.org/10.1042/CS20150697
14. Rabinovich EI, Kapetanaki MG, Steinfeld I et al (2012) Global methylation patterns in idiopathic pulmonary fibrosis. PLoS One 7(4):e33770. https://doi.org/10.1371/journal.pone.0033770
15. Huang SK, Scruggs AM, McEachin RC et al (2014) Lung fibroblasts from patients with idiopathic pulmonary fibrosis exhibit genome-wide differences in DNA methylation compared to fibroblasts from nonfibrotic lung. PLoS One 9(9):e107055. https://doi.org/10.1371/journal.pone.0107055
16. Putman RK, Rosas IO, Hunninghake GM (2014) Genetics and early detection in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 189(7):770–778. https://doi.org/10.1164/rccm.201312-2219PP
17. Helling BA, Yang IV (2015) Epigenetics in lung fibrosis: from pathobiology to treatment perspective. Curr Opin Pulm Med 21(5):454–462. https://doi.org/10.1097/MCP.0000000000000191
18. Keerthisingam CB, Jenkins RG, Harrison NK et al (2001) Cyclooxygenase-2 deficiency results in a loss of the anti-proliferative response to transforming growth factor-beta in human fibrotic lung fibroblasts and promotes bleomycin-induced pulmonary fibrosis in mice. Am J Pathol 158(4):1411–1422. https://doi.org/10.1016/s0002-9440(10)64092-8
19. Coward WR, Feghali-Bostwick CA, Jenkins G et al (2014) A central role for G9a and EZH2 in the epigenetic silencing of cyclooxygenase-2 in idiopathic pulmonary fibrosis. FASEB J 28(7):3183–3196. https://doi.org/10.1096/fj.13-241760
20. Shi L, Dong N, Fang X, Wang X (2016) Regulatory mechanisms of TGF-β1-induced fibrogenesis of human alveolar epithelial cells. J Cell Mol Med 20 (11):2183–2193. https://doi.org/10.1111/jcmm.12918
21. Huan C, Yang T, Liang J et al (2015) Methylation-mediated BMPER expression in fibroblast activation in vitro and lung fibrosis in mice in vivo. Sci Rep 5:14910. https://doi.org/10.1038/srep14910
22. Dakhlallah D, Batte K, Wang Y, Carmen Z et al (2013) Epigenetic regulation of miR-17~92 contributes to the pathogenesis of pulmonary fibrosis. Am J Respir Crit Care Med 187(4):397–405. https://doi.org/10.1164/rccm.201205-0888oc

23. Sanders YY, Kumbla P, Hagood JS (2007) Enhanced myofibroblastic differentiation and survivalin Thy-1(-) lung fibroblasts. Am J Respir Cell Mol Biol 36 (2):226–235. https://doi.org/10.1165/rcmb.2006-0178OC
24. Sanders YY, Pardo A, Selman M et al (2008) Thy-1 promoter hypermethylation: a novel epigenetic pathogenic mechanism in pulmonary fibrosis. Am J Respir Cell Mol Biol 39(5):610–618. https://doi.org/10.1165/rcmb.2007-0322OC
25. Robinson CM, Neary R, Levendale A, Watson CJ, Baugh JA (2012) Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the development of a pro-fibrotic phenotype. Respir Res 13:74. https://doi.org/10.1186/1465-9921-13-74
26. Cisneros J, Hagood J, Checa M et al (2012) Hypermethylation-mediated silencing of p14(ARF) in fibroblasts from idiopathic pulmonary fibrosis. Am J Phys Lung Cell Mol Phys 303(4):L295–L303. https://doi.org/10.1152/ajplung.00332.2011
27. Hu B, Gharaee-Kermani M, Wu Z, Phan SH (2010) Epigenetic regulation of myofibroblast differentiation by DNA methylation. Am J Pathol 177(1):21–28. https://doi.org/10.2353/ajpath.2010.090999
28. Hu B, Gharaee-Kermani M, Wu Z, Phan SH (2011) Essential role of MeCP2 in the regulation of myofibroblast differentiation during pulmonary fibrosis. Am J Pathol 178(4):1500–1508. https://doi.org/10.1016/j.ajpath.2011.01.002
29. Huang SK, Wettlaufer SH, Hogaboam CM, Flaherty KR et al (2008) Variable prostaglandin E2 resistance in fibroblasts from patients with usual interstitial pneumonia. Am J Respir Crit Care Med 177(1):66–74. https://doi.org/10.1164/rccm.200706-963OC
30. Huang SK, Fisher AS, Scruggs AM, White ES et al (2010) Hypermethylation of PTGER2 confers prostaglandin E2 resistance in fibrotic fibroblasts from humans and mice. Am J Pathol 177(5):2245–2255. https://doi.org/10.2353/ajpath.2010.100446
31. Kc R, Shukla SD, Gautam SS, Hansbro PM, O'Toole RF (2018) The role of environmental exposure to non-cigarette smoke in lung disease. Clin Transl Med 7(1):39. https://doi.org/10.1186/s40169-018-0217-2
32. Fingerlin TE, Murphy E, Zhang W et al (2013) Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 45(6):613–620. https://doi.org/10.1038/ng.2609
33. Song D, Tang L, Wang L, Huang J, Zeng T, Fang H, Wang X (2020) Roles of TGFβ1 in the expression of phosphoinositide 3-kinase isoform genes and sensitivity and response of lung telocytes to PI3K inhibitors. Cell Biol Toxicol 36:51. https://doi.org/10.1007/s10565-019-09487-3
34. Song D, Yang D, Powell CA, Wang X (2019) Cell-cell communication: old mystery and new opportunity. Cell Biol Toxicol 35(2):89–93. https://doi.org/10.1007/s10565-019-09470-y
35. Kadel D, Zhang Y, Sun HR, Zhao Y, Dong QZ, Qin LX (2019) Current perspectives of cancer-associated fibroblast in therapeutic resistance: potential mechanism and future strategy. Cell Biol Toxicol 35 (5):407–421. https://doi.org/10.1007/s10565-019-09461-z
36. Glumac PM, LeBeau AM (2018) The role of CD133 in cancer: a concise review. Clin Transl Med 7(1):18. https://doi.org/10.1186/s40169-018-0198-1
37. Wang X (2018) Clinical trans-omics: an integration of clinical phenomes with molecular multiomics. Cell Biol Toxicol 34(3):163–166. https://doi.org/10.1007/s10565-018-9431-3
38. Qiao T, Wang X (2019) A new light of proteomics in cell biology and toxicology. Cell Biol Toxicol 35 (4):289–291. https://doi.org/10.1007/s10565-019-09492-6
39. Toh TB, Lim JJ, Chow EK (2019) Epigenetics of hepatocellular carcinoma. Clin Transl Med 8(1):13. https://doi.org/10.1186/s40169-019-0230-0
40. Chirshev E, Oberg KC, Ioffe YJ, Unternaehrer JJ (2019) Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. Clin Transl Med 8(1):24. https://doi.org/10.1186/s40169-019-0240-y
41. Mirtavoos-Mahyari H, Ghafouri-Fard S, Khosravi A, Motevaseli E, Esfahani-Monfared Z, Seifi S, Salimi B, Oskooei VK, Ghadami M, Modarressi MH (2019) Circulating free DNA concentration as a marker of disease recurrence and metastatic potential in lung cancer. Clin Transl Med 8(1):14. https://doi.org/10.1186/s40169-019-0229-6
42. Liu X, Wu J (2018) History, applications, and challenges of immune repertoire research. Cell Biol Toxicol 34(6):441–457. https://doi.org/10.1007/s10565-018-9426-0
43. Zhang Y (2018) Cell toxicity mechanism and biomarker. Clin Transl Med 7(1):34. https://doi.org/10.1186/s40169-018-0212-7
44. Tzouvelekis A, Kaminski N (2015) Epigenetics in idiopathic pulmonary fibrosis. Biochem Cell Biol 93 (2):159–170. https://doi.org/10.1139/bcb-2014-0126
45. Yang IV, Schwartz DA (2015) Epigenetics of idiopathic pulmonary fibrosis. Transl Res 165(1):48–60. https://doi.org/10.1016/j.trsl.2014.03.011
46. Shi L, Zhu B, Xu M, Wang X (2018) Selection of AECOPD-specific immunomodulatory biomarkers by integrating genomics and proteomics with clinical informatics. Cell Biol Toxicol 34(2):109–123. https://doi.org/10.1007/s10565-017-9405-x
47. Amodio N, D'Aquila P, Passarino G, Tassone P, Bellizzi D (2017) Epigenetic modifications in multiple myeloma: recent advances on the role of DNA and histone methylation. Expert Opin Ther Targets 21 (1):91–101. https://doi.org/10.1080/14728222.2016.1266339
48. Colombo G, Garavaglia ML, Astori E, Giustarini D, Rossi R, Milzani A, Dalle-Donne I (2019) Protein carbonylation in human bronchial epithelial cells exposed to cigarette smoke extract. Cell Biol Toxicol

35(4):345–360. https://doi.org/10.1007/s10565-019-09460-0
49. Montgomery SB, Dermitzakis ET (2011) From expression QTLs to personalized transcriptomics. Nat Rev Genet 12(4):277–282. https://doi.org/10.1038/nrg2969
50. Juergens RA, Wrangle J, Vendetti FP, Murphy SC et al (2011) Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov 1(7):598–607. https://doi.org/10.1158/2159-8290.CD-11-0214

Qian-Kun Luo is a Medicine Doctor at Zhengzhou University People's Hospital. His major is hepatobiliary pancreatic surgery. His main research is focused on genomics, proteomics, and targeted gene therapy. He is the author of four scientific publications.

Hui Zhang is an editor-in-chief assistant at Henan Provincial People's Hospital and a postgraduate in Zhengzhou University. She has read a large number of scientific articles on topics such as single-cell sequencing, gene editing, molecular drug targeting, stem cells, immune therapy, and heterogeneity during her editing work. In addition, years of experience in clinical surgery and ethics committee gave her a deep understanding of medical practice ethics and scientific research ethics.

Li Li is a Director of Department of Scientific Research and Discipline Construction, Henan Provincial people's Hospital. She is Member of clinical research group of Chinese Medical Association's Scientific Research Management Branch, Standing Committee member of Chinese Medical Association's Henan Research and Management Branch, and a Vice-chairman of Henan Discipline Management Branch of Chinese Hospital Management Society. She has engaged in the management of medical scientific research for 30 years and her main research is focused on health management scientific research big data, laboratory biosafety, and medical ethics. She published more than 20 scientific papers.

DNA Methylation in Chronic Obstructive Pulmonary Disease

7

Xiaoyang Chen, Furong Yan, Xiaoping Lin, Liyong Shi, Xiangdong Wang, and Yiming Zeng

Abstract

Chronic obstructive pulmonary disease (COPD) is a lung disease affected by both genetic and environmental factors. Therefore, the role of epigenetics in the pathogenesis of COPD has attracted much attention. As one of the three epigenetic mechanisms, DNA methylation has been extensively studied in COPD. The present review aims at overviewing the effect of DNA methylation on etiology, pathogenesis, pathophysiological changes, and complications of COPD. The clarification of aberrant methylation of target genes, which play important roles in the initiation and progression of COPD, will provide new disease-specific biomarker and targets for early diagnosis and therapy.

Keywords

Chronic obstructive pulmonary disease · DNA methylation · Epigenetic

7.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a life-threatening lung disease, characterized by chronic airway inflammation and airflow limitation and induced by the exposure to noxious particle especially by cigarette smoking and pollution. The development of COPD is associated with genetic factors as well as environmental exposures. The epigenetic mechanisms, by which the gene and environment interact, are subdivided into DNA methylation, post-translational histone methylation or acetylation and non-coding RNAs (Fig. 7.1). DNA methylation, addition of a methyl group to the five position of the cytosine, usually occurs in the CpG islands located at the core sequence and transcription start point of the structural gene promoter. DNA methylation sites can be inherited with the replication of DNA and catalyzed by DNA methyltransferases (DNMTs) (Fig. 7.2). Hypermethylation of CpG islands in gene promoters usually leads to downregulation of gene expression, whereas hypomethylation leads to upregulation [1] (Fig. 7.3). Therefore, the role of DNA methylation in COPD has become a hot spot area. This review is focus on how

X. Chen (✉) · X. Lin · L. Shi · Y. Zeng
Department of Pulmonary and Critical Care Medicine, Respiratory Medicine Center of Fujian Province, Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, China

F. Yan
Center for Molecular Diagnosis and Therapy, Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian Province, China

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China
e-mail: Xiangdong.wang@clintransmed.org

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_7

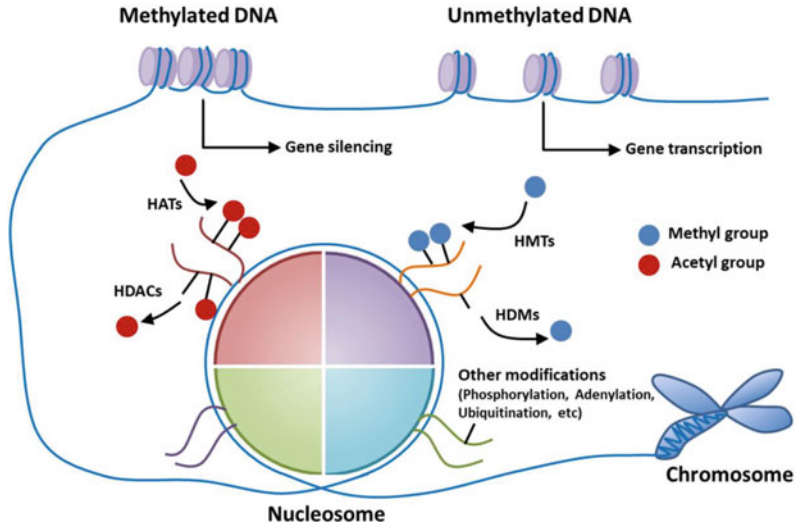


Fig. 7.1 The main ways of epigenetic modification. Most epigenetic modifications act at DNA methylation, histone methylation or acetylation. DNA methylation can be regulated by action of DNA methyltransferases (DNMT). Histone modifications are mainly manifested in acetylation and methylation. Histone methyltransferases (HMTs) and histone demethylase (HDMs) mediated histone methylation. Histone acetyl transferases (HATs) and histone deacetylases (HDACs) mediated histone acetylation

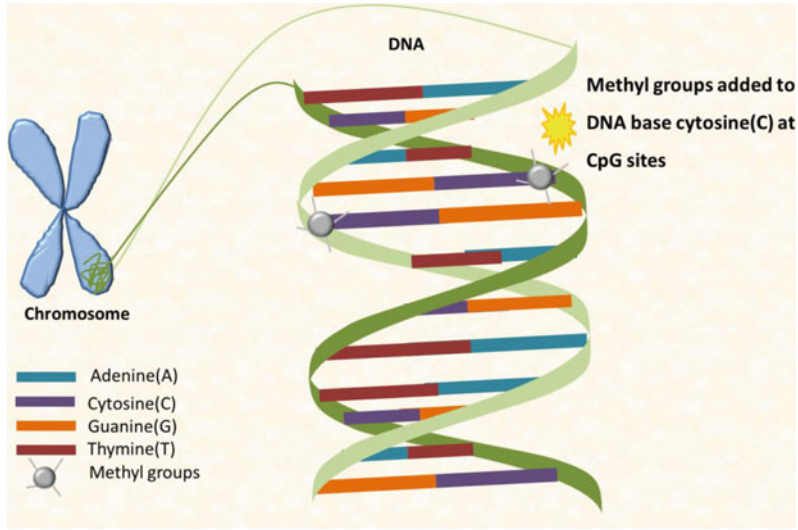


Fig. 7.2 The schematic representation of DNA methylation. The cytosine at CpG sites converts to 5′ methyl-cytosine by adding methyl groups. DNA methylation typically happened at cytosine which is followed by a guanine

environmental factors affect the development of COPD through gene methylation. The present review aims at overviewing the effect of DNA methylation on etiology, pathogenesis, pathophysiological changes, and complications of COPD.

7.2 Cigarette Smoking and DNA Methylation in COPD

Cigarette smoking is the most import environment risk factor of COPD, and other particles such as biomass fuel and urban air pollution particulate matter (PM2.5) are also in association with COPD prevalence [2]. Bojesen et al. studied

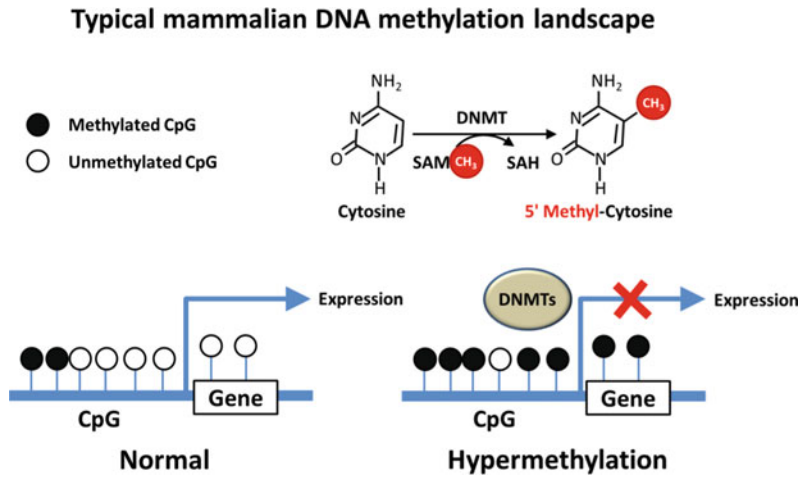


Fig. 7.3 A typical mammalian DNA methylation landscape. The process made a cytosine changing to a 5′methyl-cytosine by a family of DNMTs. When the promoter region is hypermethylated, the gene cannot be expressed. Notes: SAM: *S*-adenosylmethionine; SAH: *S*-adenosylhomocysteine; DNMTs: DNA methyltransferases

9234 individuals from the Copenhagen City Heart Study, and measured aryl-hydrocarbon receptor repressor (AHRR) (cg05575921) methylation using bisulfite treated leucocyte DNA [3]. Exacerbations of COPD in participants were followed for more than 20 years. Lower AHRR methylation was found to be associated with higher daily cigarette consumption, higher cumulative tobacco consumption, and longer smoking duration and shorter time since cessation. Hypomethylation of AHRR (cg05575921) was also associated with higher incidence of COPD exacerbation. This particular study suggested that AHRR (cg05575921) hypomethylation was associated not only with cigarette smoking but also with smoking-related morbidity and mortality.

Cigarette smoking could also cause DNA methylation other than AHRR. Cheng et al. analyzed DNA methylation on glutamate-cysteine ligase catalytic subunit (GCLC) promoter in clinical lung biopsy specimens from current-smoker, ex-smoker, and never-smoker patients with or without COPD [4]. GCLC regulates synthesis of glutathione (GSH), which is a critical antioxidant in the airway [5]. They found that the DNA methylation level of the GCLC promoter of cigarette smoking groups were significantly higher than those of non-smoking groups. Furthermore, the mRNA levels of GCLC in the lungs were correlated with the level of GCLC promoter hypermethylation. This finding was evidenced by in-vitro study.

A number of genome-wide DNA methylation data of cigarette smoking on COPD were generated in recent years. Emily et al. detected the methylation of buccal brushes DNA sample from COPD patients with current and former smoking history [6]. Seven CpG sites were found to be correlated with current cigarette smoking, including cg09853702, cg16323911 (long intergenic non-protein coding RNA 673, LINC00673), cg02162897 (Cytochrome P450 1B1, CYP1B1), cg03126561 (parvin alpha, PARVA), cg16199747 (syntaxin binding protein 5 antisense RNA 1, STXBP5-AS1), cg16187635 and cg21371809 (FERM domain containing 4A, FRMD4A). The methylation level of CYP1B1 was inversely correlated with lung function and the radiographic severity of emphysema in females. CYP1B1 plays a role in xenobiotic metabolism, probably contributing to COPD development. In addition, CYP1B1 is also associated with the estrogen metabolism and may play a role in difference of gender susceptibility towards smoking-related diseases [7, 8]. After the measurement of genome-wide DNA methylation in lung tissue, nitric oxide synthase 1 adaptor protein (NOS1AP), tumor necrosis factor, alpha-induced protein 2 (TNFAIP2), BH3 interacting domain death agonist (BID), gamma-aminobutyric acid type A receptor beta1 subunit (GABRB1), ataxin 7 (ATXN7), and THO

complex 7 (THOC7) were found to have differential methylations in lung tissues acquired from pneumonectomy, which were further validated by pyrosequencing. Pyrosequencing validation confirmed that hyper-methylations of NOS1AP, BID, and GABRB1 were associated with low expression of the corresponding genes in smokers and COPD patients [9].

Moreover, genome-wide DNA methylation research demonstrated that cigarette smoking per se could affect DNA methylation. In current smokers, DNA methylation of six differentially methylated positions (DMPs) were correlated with urine cotinine levels, including G protein subunit gamma 12 (GNG12), G protein-coupled receptor 15 (GPR15), melatonin receptor 1A (MTNR1A), AHRR, regulator of microtubule dynamics 3 (FAM82A2), and F2R like thrombin or trypsin receptor 3 (F2RL3). DNA methylation of six DMPs were correlated with pack-years, including 5′-nucleotidase, cytosolic IA (NTSC1A), zinc finger and BTB domain containing 9 (2BTB9), Janus kinase and microtubule interacting protein 3 (JAKMIP3), hemopexin (HPX), cyclin D1 (CCND1), and RING zinc finger-containing protein (RNF160) [10]. In ex-smokers, eight DMPs were correlated with duration of smoking cessation, including interferon gamma inducible protein 16 (IFI16), cytoplasmic linker associated protein 1 (CLASP1), Rho guanine nucleotide exchange factor 3 (ARHGEF3), protein O-glucosyltransferase 1 (KTELC1), sperm flagellar 2 (SPEF2), acyl-CoA thioesterase 13 (ACOT13), B-box and SPRY domain containing (BSPRY), and FAM82A2. The smoking-related differential methylation in blood was correlated with gene expression levels in lung tissue. This particular study demonstrated the effect of cigarette on DNA methylation, and the dose-dependent effect. Since smoking cessation was correlated with DNA methylation, there might be a time-dependent relationship of this effect [10]. Cigarette smoking could affect the expression of DNMTs in animal model and in COPD patients. While DNA methylation is mediated by DNMTs, cigarette smoking could affect the expression of target genes via differential methylation induced by abnormal expression of DNMTs [11]. Nevertheless, the prenatal smoke exposure was found to be associated with differential DNA methylation in newborns and children [12]. In an animal model, differential methylations of insulin like growth factor 1 receptor (IGF1r) occurred in a sex-dependent manner after prenatal exposure to smoke [13]. IGF1 and IGF1r play a role in lung development [14]. Gathering those evidence, it can be confirmed that cigarette smoking could affect DNA methylation. On the one hand, studies of genome-wide DNA methylation suggested that some suspicious genes after cigarette smoking were related with aberrant methylation, influencing the development of COPD. Some gene aberrant methylations were associated with the effect of cigarette smoking on COPD, correlated with expression levels of these genes. These results indicate that DNA methylation plays an important role in the mechanism of cigarette smoking on COPD.

7.3 Environmental Risk Factors and DNA Methylation in COPD

Many potential risk factors of COPD could induce DNA methylation. Sood et al. performed a cross-sectional study and methylation-specific polymerase chain reaction (PCR) assay in induced sputum sample and found that the wood smoke exposure was associated with reduction of lung function, correlated with differential methylations of p16 gene and zinc finger transcription factors binding protein 4 (GATA4) [15]. The air pollution, especially PM2.5 as a risk factor, could induce hypermethylation of p16 gene promoter and decrease DNMTs activity in primary bronchial epithelial cells derived from COPD patients, which could explain the global DNA hypomethylation [16]. The P16 gene could affect the genomic instability and the cell cycle progression [17]. Chen et al. investigated PM2.5 on COPD patients without smoking history and found that DNA hypomethylation of nitric oxide synthase 2A (NOS2A) from buccal sample was associated with elevated fractional exhaled nitric

Table 7.1 Relationships between environment factors and gene methylation of COPD

Environment factors	Genes	Gene function	Gene methylation	Possible impact on COPD	References
Cigarette smoking	AHRR	A tumor suppressor	Hypomethylation	Higher COPD exacerbation incidence; smoking-related morbidity and mortality.	[3]
Cigarette smoking	GCLC	Regulates synthesis of GSH	Hypermethylation	–	[4]
Cigarette smoking	CYP1B1	Regulates xenobiotic metabolism and estrogen metabolism	Hypomethylation	Reduction of lung function and severity of radiographic emphysema in female	[6–8]
Cigarette smoking	IGF1r	Regulates lung development	Hypomethylation	Reduction of lung function	[13]
Wood smoke exposure	GATA4		Hypermethylation	Reduction of lung function	[15]
Wood smoke exposure and particulate matter (PM2.5)	p16	Affects genomic instability and the cell cycle progression	Hypermethylation	Reduction of lung function	[15, 16]
Particulate matter (PM2.5)	NOS2A	Regulates NO synthesis	Hypomethylation	Elevated FeNO	[18]

oxide (FeNO) in COPD patients [18]. Thus, cigarette smoking and environmental risk factors, such as PM2.5 and biofuels, may also contribute to the pathogenesis of COPD through aberrant DNA methylation. The relationships between environment factors and gene methylation of COPD are detailed in Table 7.1.

7.4 Genome-Wide Methylation in COPD

Epigenetic mechanisms, specifically DNA methylation, are proposed to play an important role in asthma and cancer [19]. However, the profiling of DNA methylation in patients with COPD remains unclear. Isaac et al. performed genome-wide DNA methylation analysis in lung tissues from COPD patients and suggested genes, such as NOS1AP, TNFAIP2, BID, GABRB1, ATXN7, and THOC7, needed to be further validated by pyrosequencing. Differential DNA methylations of NOS1AP, BID, and GABRB1 were confirmed by pyrosequencing [9]. Jarrett et al. performed analyses of methylation quantitative trait loci (mQTL), association of single nucleotide polymorphisms (SNPs) with percent DNA methylation, and the colocalization of these results with previous COPD genome-wide association study (GWAS) findings using Bayesian methods and found aberrant methylation in potassium two pore domain channel subfamily K member 3 (KCNK3), eukaryotic elongation factor, selenocysteine-tRNA specific (EEFSEC), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta (PIK3CD), doublecortin domain containing 2C (DCDC2C), transcription elongation regulator 1 like (TCERG1L), FERM domain containing 4B (FRMD4B), and interleukin 27 (IL27) [20]. Those findings supported the role of DNA methylation in COPD pathogenesis.

A large-scale gene-specific investigation of DNA methylation demonstrated that COPD was associated with DNA methylation of 349 CpG sites in two family-based cohorts [21]. The gene ontology analysis suggested that the associated genes were involved in immune and inflammatory system pathways, in stress and external

stimuli, as well as in wound healing and coagulation cascades. The top five associated genes for COPD were chromosome 3 open reading frame 18 (C3orf18), serpin family A member 1 (SERPINA1), chloride voltage-gated channel 6 (CLCN6), core-binding factor, runt domain, alpha subunit 2; translocated to, 3 (CBFA2T3) and fucosyltransferase 7 (FUT7), while for forced expiratory volume in the first second (FEV1) were FXYD domain-containing ion transport regulator 1 (FXYD1), FUT7, SERPINA1, transient receptor potential cation channel subfamily M member 2 (TRPM2), and FERMT3. The SERPINA1 hypomethylation was significantly associated with COPD and phenotypes of lung dysfunction. The SERPINA1 gene encodes the alpha-1 antitrypsin protein, which is associated with COPD [21]. However, data from genome-wide methylation studies on COPD hardly interpreted due to the obvious variations, complexity of disease severity and stages, and uncontrolled samplings.

Jarrett et al. performed genome-wide DNA methylation analysis in lung tissue samples from 114 COPD patients and 46 control subjects and found that 11 out of 535 differential methylation sites had the difference between disease and controls [22]. The expression levels of those genes were correlated with DNA methylation, including cholinergic receptor muscarinic 1 (CHRM1), deltex 1, deltex E3 ubiquitin ligase 1 (DTX1), protein kinase C and casein kinase substrate in neurons 1 (PACSIN1), forkhead box K1 (FOXK1), sperm hammerhead 3 and multiple ankyrin repeat domains 2 (SHANK2), and forkhead box P2 (FOXP2). The associated genes identified in previous COPD genome-wide studies were integrated with differential methylation, where the top methylated sites from the intersection with previous GWAS were in CHRM1, glycosyltransferase 1 domain containing 1 (GLT1D1), and chromosome 10 open reading frame 11 (C10orf11). Of those, CHRM1 is involved in mediation of bronchoconstriction, suggesting that aberrant methylation of CHRM1 may play an important role in COPD [22].

7.5 Individual Gene Aberrant Methylation in COPD

Genome-wide methylation studies can help to identify and discover the potential genes associated with COPD. To validate concrete genes of which the aberrant DNA methylation plays a crucial role in pathogenesis of COPD, the methylation of single gene should be detected. Genome-wide association studies identified genes associated with COPD susceptibility on chromosomes 19q13 [23]. Nedeljkovic et al. detected the DNA methylation of rs7937 (RAB4B, EGLN2), a top associated gene in 19q13.2 region, and the gene expression level in both blood and lung tissue [24]. Rs7937 was significantly and consistently associated with differential DNA methylation in blood at four CpG sites in cis, independent of smoking. Aberrant methylation at cg11298343-EGLN2 was also found in COPD patients, and was correlated with expression level of EGLN2 [24]. The protein expression of EGLN2, prolyl hydroxylase domain-containing protein 1, regulates expression of transcription factor hypoxia induced factor (HIF), of which downstream genes were associated with inflammatory and immune process in COPD. Four gene variants, e.g., cholinergic receptor nicotinic alpha 3 subunit (CHRNA3), hydroxylysine kinase (HYKK), rs13180: iron responsive element binding protein 2 (IREB2), and rs8042238:IREB2 in chromosome 15q25.1 locus, were associated with COPD by epigenome-wide association analysis [24]. The aberrant methylation of IREB2, CHRNA3, and PSMA4 genes was associated with COPD, independently upon smoking. CHRNA3 encodes the nicotinic acetylcholine receptors, associated with risk of cigarette smoking and lung cancer [25].

Hong Li et al. assayed methylation of endothelial PAS domain protein 1 (EPAS1) by bisulfite sequencing PCR, and EPAS1 mRNA expression by real-time PCR in bronchoalveolar lavage fluid samples of COPD patients and healthy subjects and found that the aberrant methylation of EPAS1 promoter was correlated with the transcriptional level of EPAS1 mRNA in COPD [26]. Moreover,

toll like receptor 4 (TLR4) could promote hypermethylation of EPAS1 promoter and inhibited EPAS1 mRNA transcription in endothelial cells, suggesting that upregulation of TLR4 expression might decrease EPAS1 expression through the regulation of DNA methylation. EPAS1 gene encodes hypoxia inducible factor 2 subunit alpha (HIF2a), a transcription factor in the induction of genes regulated by hypoxia, which is an important pathophysiological characteristic of COPD. The relationship between EPAS1 methylation and COPD should be furthermore clarified. To find out the chief regulators of epigenetics in COPD, Yoo et al. developed a systematic approach to integrate genome-wide DNA methylation and gene expression in lung tissues with phenotype data of COPD patients and control subjects [27].

Key regulators of downstream target genes are more than all cis genes. Diffusing capacity of the lung for carbon monoxide (DLCO), body mass index, airflow obstruction, dyspnea and exercise capacity index (BODE), forced expiratory volume percentage predicted 1 (FEV1), FEV1/forced vital capacity (FVC) ratio, and emphysema percentage were used to represent the phenotypes of COPD severity [27]. Of 126 genes of key regulators in patients with COPD, ACSF3, SELO, and EPAS1 were found to be correlated with those phenotypes of COPD severity. EPAS1 was the only gene whose downstream genes, including vascular endothelial growth factor (VEGF), were significantly overlapping with all disease phenotype gene expression signature sets. The expression of EPAS1 protein was downregulated in lung tissues from COPD patients. The expressions of downstream genes, including VEGF, of EPAS1 were correlated with methylation of EPAS1. The relationship between EPAS1 and the downstream target genes was confirmed using siRNA to block EPAS1 expression in cell lines and RNASeq analysis to quantify gene expression changes [27]. This integrative analysis of genome-wide DNA methylation and gene expression profiles demonstrated that aberrant gene methylation play an important role in the development of COPD, of which EPAS1 may be the key regulator. However, Andresen et al. performed bisulfite sequencing in bronchial epithelium and BAL fluid cells from COPD patients and failed to reveal the relationship of aberrant methylation of DEFB1 promoter and upregulated expression of DEFB1 in COPD [28]. This suggests that there might be epigenetics factors, other than DNA methylation, regulate gene expression of COPD. Although a number of studies on role of gene methylation have been carried out in patients with COPD, roles of DNA methylation in the development of COPD need to be furthermore clarified and evidenced, especially due to the complexity of DNA methylation and the unknown etiology of COPD.

7.6 Gene Methylation and Mucus Hypersecretion in COPD

The excessive production of mucus and hypertrophy or metaplasia of goblet cells are critical characters of COPD. SAM-pointed domain-containing ETS-like factor (SPDEF) and forkhead box protein A2 (FOXA2) are key transcription factors to regulate the differentiation of goblet cells [29, 30]. Song et al. assessed DNA methylation and expression of SPDEF and FOXA2 in primary airway epithelial cells isolated from COPD patients and used IL-13 to induce differentiation of goblet cell [31]. They found that the hypermethylation of SPDEF promoter occurred during goblet cell differentiation, rather than FOXA2 promoter. During goblet cell differentiation, expression of SPDEF increased, whereas expression of FOXA2 decreased. In the absence of IL-13, both SPDEF promoter and FOXA2 promoter were hypomethylated, while only SPDEF expression increased. It implicates that differential methylation of SPDEF and FOXA2 promoters play a role in COPD mucus hypersecretion.

DNA methylation could not only affect airway mucus secretion but also affect mucus clearance. Tessema et al. found aberrant DNA methylation of coiled-coil domain containing 37 (CCDC37) in sputum samples of COPD patients [32]. Meanwhile, the expression level of CCDC37 in

patients with COPD was significantly lower than patients without COPD. CCDC37 encodes coiled-coil domain containing (CCDC) proteins to regulate ciliary motility [33]. The declined function of cilia motility could result in the obstacle of mucous clearing, as one of important pathophysiological characteristics of COPD. Studies on the relation between DNA methylation and mucus hypersecretion remain limited. Mucin 5AC (MUC5AC) and mucin 5B (MUC5B) are important regulators for production and secretion of mucins in COPD. Cells treated with 5-aza-2-′-deoxycytidine (5-AzadC) could downregulate DNA methylation of CpG islands near MUC5AC promoter and increase MUC5AC expression [33]. Studies found that demethylation of MUC5B promoter regions with 5-AzadC led to significant upregulation of their expression [34, 35]. It is questioned whether those effects are cell-line specific or regulatory function-specific, since such findings could not be observed in difference cell lines [36].

MUC5AC promoter has fewer CpG dinucleotides to probably contribute to MUC5AC expression regulation, in which it is questioned whether MUC5AC promoter DNA methylation may be involved [36]. The relationship between MUC5AC and MUC5B expression and DNA methylation still needs to be further defined, due to lack of solid evidence and conflicting results. In addition to MUC5AC promoter aberrant methylation, aberrant methylation of transcription factors could contribute to the regulation of MUC5AC expression. Song et al. explored the regulatory mechanism of MUC gene expression via silencing SPDEF gene in lung epithelial cells and found that fusing zinc finger proteins to DNMT3A and G9A induced targeted hypermethylation and decreased MUC5AC expression in mouse lung epithelial cells [37]. MUC5AC and MUC5B expression could also be regulated by the STAT6-STAT3 pathway. Src homology-2-containing protein-tyrosine phosphatase 1 (SHP-1) could downregulate the JAK/STAT pathway. Han et al. treated FLT3-ITD cells with 5-Aza to induce the occurrence of DNA hypomethylation and found that hypomethylation of SHP-1 promoter could downregulate the expression of STAT3 [38]. It was proposed that hypermethylation of SHP-1 gene and reduced SHP-1 protein may upregulate STAT3 expression, leading to over-expression of MUCIN genes in airway epithelial cells. Dysregulation of FOXA2 could cause mucus hypersecretion. Helling et al. found that expression of MUC5B was upregulated by hypermethylation of an enhancer region containing the rs35705950 variant that binds RNA polymerase II and FOXA2 [39]. Further studies need to define roles of DNA methylation in regulation of airway mucus hypersecretion in COPD.

7.7 Gene Methylation in COPD-Lung Cancer Transit

The incidence of lung cancer is five times higher in COPD patients [40]. The mechanism by which lung cancer develops in COPD patients is still unknown. Cigarette smoking is a common risk factor for both COPD and lung cancer and induces DNA methylation, which was found in both COPD and lung cancer. It has been supposed that DNA methylation might play a role in COPD-lung cancer transit. Suzuki et al. assessed methylation profiles of 12 genes from 229 non-small cell lung cancer (NSCLC) patients with or without COPD and found that DNA methylation of interleukin 12 receptor, beta 2 (IL-12Rb2), and WNT inhibitory factor 1 (Wif-1) in malignant lung tissues was significantly higher in the NSCLC with COPD patients than in NSCLC alone. The epigenetically silenced Wif-1 gene, a tumor suppressor gene, is associated with many cancers. IL-12Rb2 plays a role in Th1 cell differentiation. Hypermethylation of IL-12Rb2 and Wif-1 might not be related to pathogenesis of COPD, while it might be play a role in COPD-NSCLC transit [41]. Wauters et al. measured the methylation profiling on 49 pairs of tumor and adjacent lung tissue from patient with NSCLC with or without COPD [42]. Ontology analysis of genes with differential methylation

revealed that aberrant DNA methylation of immune genes was significantly enriched in NSCLC with COPD, rather than in NSCLC alone. It seems that aberrant DNA methylation of immune response genes in patients with COPD and NSCLC affects the tumor tissue specifically. Genes expressed particularly by immune cells, such as CD4-positive cells, exhibited promoter hypermethylation and expressions of transcriptional factor genes differed in COPD and NSCLC, indicating that DNA methylation of immune response genes may contribute to the COPD-NSCLC transit [42]. Tessema et al. found that aberrant methylation of CCDC37 and MAP1B occurred more frequently in NSCLC with COPD than in NSCLC alone, although relationship between CCDC37 and MAP1B with lung cancer remains unclear [32]. Bojesen et al. studied DNA methylation of AHRR (cg05575921) in blood sample harvested from 9234 individuals, and followed up for exacerbation rate of COPD, incidence of lung cancer, and mortality for up to 22 years [3]. The multifactorial adjusted hazard ratios for the lowest versus highest methylation quintiles were 4.58 (95% CI 2.83–7.42) for COPD exacerbations, 4.87 (2.31–10.3) for lung cancer, and 1.67 (1.48–1.88) for all-cause mortality. The expression of AHRR regulates cell growth and differentiation. Although some studies demonstrated that differential methylation of AHRR was associated with cigarette smoking, there is few evidences of the relationship of AHRR with lung cancer or COPD-lung Cancer transit. The role of AHRR in the development of COPD and lung cancer and their transit needs to be furthermore investigated. DNA methylation was found often in patients with COPD and lung cancer, indicating that aberrant methylation plays a role in COPD-lung cancer transit.

7.8 DNA Methylation in Different Races with COPD

Clinical characteristics of COPD differ among races. African-Americans may develop COPD at a younger age and have higher incidence of COPD hospitalizations [43]. Despite fewer pack-years of smoking, African-Americans present the same degree of airflow limitation as white and have lower quality of life scores, when COPD developed. Race is an important contributor to genetic variability of COPD and also to epigenomic of COPD. Sun et al. found that DNA methylation caused by cigarette smoking differs between racial groups [44]. Busch et al. measured methylation of DNA extracted from blood samples in 362 African-American cigarette smokers, and then compared the COPD-related aberrant DNA methylation loci discovered in the PA-SCOPE African-American cohort with ICGN white cohort [45]. Seven of 12 differential methylation genes were identified in the white cohort with significant difference. There was a difference in the methylation in those of African-American descent when compared to Anglo-Saxons, as indicated by the left (2.5 percentile) tail of the distribution. However, there is a conflicting finding in race-related differences of aberrant DNA methylation in COPD patients. Lee et al. performed an epigenome-wide association study on DNA extracted from blood sample in a Korean COPD cohort, and found aberrant DNA methylation associated with COPD and declined lung function, as identified previously in Caucasians or African-Americans populations [45, 46]. Although there are evidences that race factors may affect differential methylation in COPD, there is no clinical trial directly to compare DNA methylation of certain genes between races.

7.9 DNA Methylation and Airway Inflammation of COPD

The relationship between DNA methylation and airway inflammation in COPD is initially investigated, although airway inflammation is an essential feature of COPD. Wauters et al. reported differential DNA methylation of immune related genes in COPD with NSCLC, including CD4 encoding a membrane glycoprotein of T

lymphocytes to interact with major histocompatibility complex class II antigens, C-C motif chemokine ligand 5 (CCL5) encoding a chemoattractant, tumor necrosis factor receptor superfamily, member 21 (TNFRSF21) encoding a member of the tumor necrosis factor (TNF) receptor superfamily, and sushi domain containing 2 (SUSD2) influencing tumor infiltration by CD4-positive lymphocytes [42]. Although the specimens of this particular study were harvested from patients with COPD and lung cancer rather than COPD alone, the results implicated that DNA methylation may be play a role in the airway inflammation of COPD. Vucic et al. analyzed genome-wide methylation and gene expression on airway epithelial cells obtained from COPD patient during bronchoscopy and found that aberrant DNA methylation was associated with expression variation of genes and pathways involved in COPD, especially nuclear factor, erythroid 2 like 2 (Nrf2) pathway which mediated oxidative stress response [6].

In addition to the important role of oxidative stress in airway inflammation, Cytochrome P450 4F11 (CYP4F11) gene in the pathway expression was modulated by DNA methylation. CYP4F11 has a direct role in inhibition of inflammation through suppression of leukotriene and prostaglandin signals [47]. Silencing of CYP4f11 in small airways of patients with COPD increased airway inflammation or altered activation of inhaled steroids [48]. Although there is a need to directly evidence roles of DNA methylation in airway inflammation, DNA methylation was noticed to be associated with systemic inflammation. FOXP3 is a transcription factor to control the development and function of T-regulatory (Treg) cells, while aberrant methylation in the promoter region of FOXP3 affected the formation and activity of Treg [49, 50]. Those studies call more attentions and actions to clarify the relationship between gene methylation and airway inflammation in patients with COPD.

Loss of muscle mass and strength is an important comorbidity of COPD. DNA methylation may be associated with loss of muscle mass of COPD. Lewis et al. found that DNA hypermethylation of the H19 imprinting control region in COPD patients was associated with muscle weakness [51]. H19 gene as the host gene for miR-675 was associated with loss of muscle mass by inhibiting myoblast proliferation and inducing differentiation. Hypomethylation of H19 imprinting control region increased the expression of H19 and miR-675, leading to a low fat free mass index in patients with COPD.

7.10 Summary and Future Directions

Although the aberrant DNA methylation was observed in blood sample and lung tissue in patients with COPD, COPD per se is a systemic disease involving in multiple systems and organs detected by medical and molecular imaging and proteomics [52, 53]. With or without the association with DNA methylation, a large number factors contribute to the development and progression of COPD, e.g., fatty acid binding proteins [54], inflammatory responses, environmental complexity [55], signaling mechanisms of YPEL4 [56], disordered biorhythms [57], epithelial cell responses [58, 59], and mitochondrial dysfunction [60]. Cigarette smoking as one of risk factors of COPD is clearly associated with the occurrence of aberrant DNA methylations. In addition, aberrant DNA methylation is also associated with airway inflammation, mucus hypersecretion, muscle mass loss, as well as COPD-cancer transit (Fig. 7.4).

Although the mechanism by which DNA methylation contributes to the development of COPD remains uncertain, the clinical and preclinical evidence on roles of DNA methylation in COPD may provide new views to understand the pathogenesis of COPD, new opportunities to identify and develop diagnostic biomarkers, and therapeutic targets for COPD. It is questioned whether DNA methylation-associated biomarkers and targets have the specificity of disease severity, duration, staging, and response to therapy [61–71]. Values of DNA methylation-associated therapeutic targets are also to be furthermore

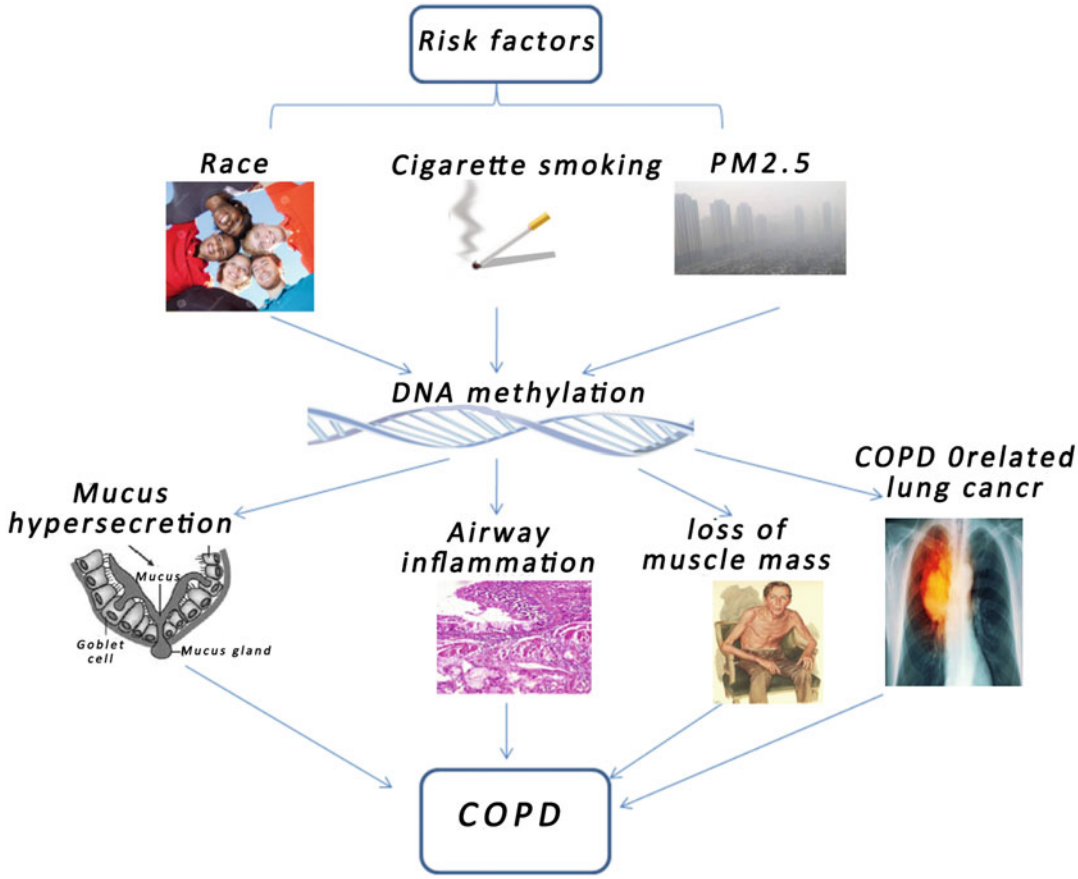


Fig. 7.4 The role of gene methylation in the pathogenesis of COPD. Risk factors of COPD, such as cigarette smoking and PM2.5, could induce aberrant DNA methylations. In addition, aberrant DNA methylation is also associated with airway inflammation, mucus hypersecretion, muscle mass loss, as well as COPD-related lung cancer

validated in diseases, especially in COPD with the contributions of multi-factors, since the on/off-target situation of DNA methylation-associated therapeutic targets in COPD is undiscovered, although there are some discussions recently [72–77]. Thus, we believe that the aberrant methylation of target genes plays the most important role in the initiation and progression of COPD and will provide new alternatives of therapies for COPD.

Funding: This works was supported by National key R&D program of China (2016YFC1304003)

References

1. Gros C, Fahy J, Halby L, Dufau I, Erdmann A, Gregoire JM, Ausseil F, Vispé S, Arimondo PB (2012) DNA methylation inhibitors in cancer: recent and future approaches. Biochimie 94(11):2280–2296
2. Liu S, Zhou Y, Liu S, Chen X, Zou W, Zhao D, Li X, Pu J, Huang L, Chen J, Li B, Liu S, Ran P (2017) Association between exposure to ambient particulate matter and chronic obstructive pulmonary disease: results from a cross-sectional study in China. Thorax 72(9):788–795
3. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG (2017) AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax 72(7):646–653
4. Cheng L, Liu J, Li B, Liu S, Li X, Tu H (2016) Cigarette smoke-induced hypermethylation of the GCLC gene is associated with COPD. Chest 149 (2):474–482

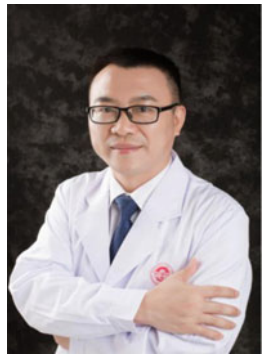
5. MacNee W (2000) Oxidants/antioxidants and COPD. Chest 117(5, Suppl 1):303S–317S
6. Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, Hokanson JE, Bowler RP, Crapo JD, DeMeo DL (2015) Smoking-associated site-specific differential methylation in buccal mucosa in the COPDGene study. Am J Respir Cell Mol Biol 53 (2):246–254
7. Shimada T, Hayes CL, Yamazaki H, Amin S, Hecht SS, Guengerich FP, Sutter TR (1996) Activation of chemically diverse procarcinogens by human cytochrome p-450 1b1. Cancer Res 56(13):2979–2984
8. Cote ML, Yoo W, Wenzlaff AS, Prysak GM, Santer SK, Claeys GB, Van Dyke AL, Land SJ, Schwartz AG (2009) Tobacco and estrogen metabolic polymorphisms and risk of non-small cell lung cancer in women. Carcinogenesis 30(4):626–635
9. Sundar IK, Yin Q, Baier BS, Yan L, Mazur W, Li D, Susiarjo M, Rahman I (2017) DNA methylation profiling in peripheral lung tissues of smokers and patients with COPD. Clin Epigenetics 9:38
10. Lee MK, Hong Y, Kim SY, London SJ, Kim WJ (2016) DNA methylation and smoking in Korean adults epigenome-wide association study. Clin Epigenetics 8:103
11. Sundar IK, Rahman I (2016) Gene expression profiling of epigenetic chromatin modification enzymes and histone marks by cigarette smoke implications for COPD and lung cancer. Am J Phys Lung Cell Mol Phys 311(6):L1245–L1258
12. Ladd-Acosta C, Shu C, Lee BK, Gidaya N, Singer A, Schieve LA, Schendel DE, Jones N, Daniels JL, Windham GC, Newschaffer CJ, Croen LA, Feinberg AP, Daniele FM (2016) Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. Environ Res 144(Pt A):139–148
13. Meyer KF, Krauss-Etschmann S, Kooistra W, Reinders-Luinge M, Timens W, Kobzik L, Plösch T, Hylkema MN (2017) Prenatal exposure to tobacco smoke sex dependently influences methylation and mRNA levels of the Igf axis in lungs of mouse offspring. Am J Phys Lung Cell Mol Phys 312(4):L542–L555
14. Moreno-Barriuso N, López-Malpartida AV, de Pablo F, Pichel JG (2006) Alterations in alveolar epithelium differentiation and vasculogenesis in lungs of LIF/IGF-I double deficient embryos. Dev Dyn 235 (8):2040–2050
15. Sood A, Petersen H, Blanchette CM, Meek P, Picchi MA, Belinsky SA, Tesfaigzi Y (2010) Wood smoke exposure and gene promoter methylation are associated with increased risk for COPD in smokers. Am J Respir Crit Care Med 182(9):1098–1104
16. Leclercq B, Platel A, Antherieu S, Alleman LY, Hardy EM, Perdrix E, Grova N, Riffault V, Appenzeller BM, Happillon M, Nesslany F, Coddeville P, Lo-Guidice JM, Garçon G (2017) Genetic and epigenetic alterations in normal and sensitive COPD-diseased human bronchial epithelial cells repeatedly exposed to air pollution-derived PM2.5. Environ Pollut 230:163–177
17. Soberanes S, Gonzalez A, Urich D, Chiarella SE, Radigan KA, Osornio-Vargas A, Joseph J, Kalyanaraman B, Ridge KM, Chandel NS, Mutlu GM, De Vizcaya-Ruiz A, Budinger GR (2012) Particulate matter air pollution induces hypermethylation of the P16 promoter via a mitochondrial ROS-JNK-DNMT1 pathway. Sci Rep 2:275. https://doi.org/10.1038/srep00275
18. Chen R, Qiao L, Li H, Zhao Y, Zhang Y, Xu W, Wang C, Wang H, Zhao Z, Xu X, Hu H, Kan H (2015) Fine particulate matter constituents, nitric oxide synthase DNA methylation and exhaled nitric oxide. Environ Sci Technol 49(19):11859–11865
19. Sundar IK, Yao H, Rahman I (2013) Oxidative stress and chromatin remodeling in chronic obstructive pulmonary disease and smoking-related diseases. Antioxid Redox Signal 18:1956–1971
20. Morrow JD, Glass K, Cho MH, Hersh CP, Pinto-Plata-V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Choi AMK, Quackenbush J, Silverman EK, DeMeo DL (2018) Human lung DNA methylation quantitative trait loci colocalize with COPD genome-wide association loci. Am J Respir Crit Care Med 197 (10):1275–1284
21. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, Rennard S, Agusti A, Anderson W, Lomas DA, DeMeo DL (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am J Respir Crit Care Med 185(4):373–381
22. Morrow JD, Cho MH, Hersh CP et al (2016) DNA methylation profiling in human lung tissue identifies genes associated with COPD. Epigenetics 11:730–739
23. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP, Lange C, Litonjua AA, Sparrow D, Regan EA, Make BJ, Hokanson JE, Murray T, Hetmanski JB, Pillai SG, Kong X, Anderson WH, Tal-Singer R, Lomas DA, Coxson HO, Edwards LD, MacNee W, Vestbo J, Yates JC, Agusti A, Calverley PM, Celli B, Crim C, Rennard S, Wouters E, Bakke P, Gulsvik A, Crapo JD, Beaty TH, Silverman EK, ICGN Investigators; ECLIPSE Investigators; COPDGene Investigators (2012) A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet 21(4):947–957
24. Nedeljkovic I, Lahousse L, Carnero-Montoro E, Faiz A, Vonk JM, de Jong K, van der Plaat DA, van Diemen CC, van den Berge M, Obeidat M, Bossé Y, Nickle DC, Consortium BIOS, Uitterlinden AG, van Meurs JBJ, Stricker BHC, Brusselle GG, Postma DS, Boezen HM, van Duijn CM, Amin N (2018) COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression. Hum Mol Genet 27 (2):396–405
25. Nedeljkovic I, Carnero-Montoro E, Lahousse L, van der Plaat DA, de Jong K, Vonk JM, van Diemen CC,

Faiz A, van den Berge M, Obeidat M, Bossé Y, Nickle DC, Consortium B, Uitterlinden AG, van Meurs JJB, Stricker BCH, Brusselle GG, Postma DS, Boezen HM, van Duijn CM, Amin N (2018) Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics. Eur J Hum Genet 26 (5):709–722

26. Li H, Yang T, Li FY, Ning Q, Sun ZM (2016) TLR4 Overexpression inhibits endothelial PAS domain-containing protein 1 expression in the lower respiratory tract of patients with chronic COPD. Cell Physiol Biochem 39(2):685–692
27. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A, Schadt EE, Powell CA, Zhu J (2015) Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. PLoS Genet 11(1):e1004898
28. Andresen E, Günther G, Bullwinkel J, Lange C, Heine H (2011) Increased expression of beta-defensin 1 (DEFB1) in chronic obstructive pulmonary disease. PLoS One 6(7):e21898
29. Rajavelu P, Chen G, Xu Y, Kitzmiller JA, Korfhagen TR, Whitsett JA (2015) Airway epithelial SPDEF integrates goblet cell differentiation and pulmonary Th2 inflammation. J Clin Invest 125:2021–2031
30. Wan H, Kaestner KH, Ang SL, Ikegami M, Finkelman FD, Stahlman MT, Fulkerson PC, Rothenberg ME, Whitsett JA (2004) Foxa2 regulates alveolarization and goblet cell hyperplasia. Development 131:953–964
31. Song J, Heijink IH, Kistemaker LEM, Reinders-Luinge M, Kooistra W, Noordhoek JA, Gosens R, Brandsma CA, Timens W, Hiemstra PS, Rots MG, Hylkema MN (2017) Aberrant DNA methylation and expression of SPDEF and FOXA2 in airway epithelium of patients with COPD. Clin Epigenetics 9:42
32. Tessema M, Yingling CM, Picchi MA, Wu G, Liu Y, Weissfeld JL, Siegfried JM, Tesfaigzi Y, Belinsky SA (2015) Epigenetic repression of CCDC37 and MAP1B links chronic obstructive pulmonary disease to lung cancer. J Thorac Oncol 10(8):1181–1188
33. Yamamoto R, Song K, Yanagisawa HA, Fox L, Yagi T, Wirschell M, Hirono M, Kamiya R, Nicastro D, Sale WS (2013) The MIA complex is a conserved and novel dynein regulator essential for normal ciliary motility. J Cell Biol 201:263–278
34. Yamada N, Nishida Y, Yokoyama S, Tsutsumida H, Houjou I, Kitamoto S, Goto M, Higashi M, Yonezawa S (2010) Expression of MUC5AC, an early marker of pancreatobiliary cancer, is regulated by DNA methylation in the distal promoter region in cancer cells. J Hepatobil Pancreat Sci 17(6):844–854
35. Renaud F, Vincent A, Mariette C, Crépin M, Stechly L, Truant S, Copin MC, Porchet N, Leteurtre E, Van Seuningen I, Buisine MP (2015) MUC5AC hypomethylation is a predictor of microsatellite instability independently of clinical factors associated with colorectal cancer. Int J Cancer 136 (12):2811–2821
36. Ho JJ, Han SW, Pan PL, Deng G, Kuan SF, Kim YS (2003) Methylation status of promoters and expression of MUC2 and MUC5AC mucins in pancreatic cancer cells. Int J Oncol 22(2):273–279
37. Song J, Cano-Rodriquez D, Winkle M, Gjaltema RA, Goubert D, Jurkowski TP, Heijink IH, Rots MG, Hylkema MN (2017) Targeted epigenetic editing of SPDEF reduces mucus production in lung epithelial cells. Am J Phys Lung Cell Mol Phys 312(3):L334–L347
38. Al-Jamal HAN, Mat Jusoh SA, Hassan R, Johan MF (2015) Enhancing SHP-1 expression with 5-azacytidine may inhibit STAT3 activation and confer sensitivity in lestaurtinib (CEP-701)-resistant FLT3-ITD positive acute myeloid leukemia. BMC Cancer 15:869
39. Helling BA, Gerber AN, Kadiyala V, Sasse SK, Pedersen BS, Sparks L, Nakano Y, Okamoto T, Evans CM, Yang IV, Schwartz DA (2017) Regulation of MUC5B expression in idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol 57(1):91–99
40. Skillrud DM, Offord KP, Miller RD (1986) Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. Ann Intern Med 105:503–507
41. Suzuki M, Wada H, Yoshino M, Tian L, Shigematsu H, Suzuki H, Alaa M, Tamura H, Fujiwara T, Nagato K, Motohashi S, Moriya Y, Hoshino H, Yoshida S, Shibuya K, Hiroshima K, Nakatani Y, Yoshino I (2010) Molecular characterization of chronic obstructive pulmonary disease-related non-small cell lung cancer through aberrant methylation and alterations of EGFR signaling. Ann Surg Oncol 17(3):878–888
42. Wauters E, Janssens W, Vansteenkiste J, Decaluwé H, Heulens N, Thienpont B, Zhao H, Smeets D, Sagaert X, Coolen J, Decramer M, Liston A, De Leyn P, Moisse M, Lambrechts D (2015) DNA methylation profiling of non-small cell lung cancer reveals a COPD-driven immune-related signature. Thorax 70 (12):1113–1122
43. Han MK, Curran-Everett D, Dransfield MT, Criner GJ, Zhang L, Murphy JR, Hansel NN, DL DM, Hanania NA, Regan EA, Make BJ, Martinez FJ, Westney GE, Foreman MG, COPD Gene Investigators (2011) Racial differences in quality of life in patients with COPD. Chest 140:1169–1176
44. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, Smith JA, Almli LM, Binder EB, Klengel T, Cross D, Turner ST, Ressler KJ, Kardia SL (2013) Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. Hum Genet 132:1027–1037
45. Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D (2016) Differential DNA methylation marks and gene comethylation of COPD in African-

Americans with COPD exacerbations. Respir Res 17 (1):143
46. Lee MK, Hong Y, Kim SY, Kim WJ, London SJ (2017) Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. Epigenomics 9(7):971–984
47. Wang Y, Bell JC, Keeney DS, Strobel HW (2010) Gene regulation of CYP4f11 in human keratinocyte hacat cells. Drug Metab Dispos 38:100–107
48. Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, Zhang M, Lonergan KM, Steiling K, Brown CJ, McWilliams A, Ohtani K, Lenburg ME, Sin DD, Spira A, Macaulay CE, Lam S, Lam WL (2014) DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. Am J Respir Cell Mol Biol 50(5):912–922
49. Kim HP, Leonard WJ (2007) CREB/ATF-dependent T cell receptor-induced FoxP3 gene expression: a role for DNA methylation. J Exp Med 204(7):1543–1551
50. Floess S, Freyer J, Siewert C, Baron U, Olek S, Polansky J, Schlawe K, Chang HD, Bopp T, Schmitt E, Klein-Hessling S, Serfling E, Hamann A, Huehn J (2007) Epigenetic control of the foxp3 locus in regulatory T cells. PLoS Biol 5(2):e38
51. Lewis A, Lee JY, Donaldson AV, Natanek SA, Vaidyanathan S, Man WD, Hopkinson NS, Sayer AA, Patel HP, Cooper C, Syddall H, Polkey MI, Kemp PR (2016) Increased expression of H19miR-675 is associated with a low fat-free mass index in patients with COPD. J Cachexia Sarcopenia Muscle 7 (3):330–344
52. Myc LA, Shim YM, Laubach VE, Dimastromatteo J (2019) Role of medical and molecular imaging in COPD. Clin Transl Med 8(1):12
53. Gil J, Betancourt LH, Pla I, Sanchez A, Appelqvist R, Miliotis T et al (2019) Clinical protein science in translational medicine targeting malignant melanoma. Cell Biol Toxicol 35(4):293–332
54. Rao DM, Phan DT, Choo MJ, Weaver MR, Oberley-Deegan RE, Bowler RP, Gally F (2019) Impact of fatty acid binding protein 5-deficiency on COPD exacerbations and cigarette smoke-induced inflammatory response to bacterial infection. Clin Transl Med 8 (1):7
55. Kc R, Shukla SD, Gautam SS, Hansbro PM, O'Toole RF (2018) The role of environmental exposure to non-cigarette smoke in lung disease. Clin Transl Med 7(1):39. https://doi.org/10.1186/s40169-018-0217-2
56. Truong L, Zheng YM, Song T, Tang Y, Wang YX (2018) Potential important roles and signaling mechanisms of YPEL4 in pulmonary diseases. Clin Transl Med 7(1):16
57. Khatami M (2018) Cancer; an induced disease of twentieth century! Induction of tolerance, increased entropy and 'Dark Energy': loss of biorhythms (Anabolism v. Catabolism). Clin Transl Med 7(1):20. https://doi.org/10.1186/s40169-018-0193-6
58. Roberts N, Al Mubarak R, Francisco D, Kraft M, Chu HW (2018) Comparison of paired human nasal and bronchial airway epithelial cell responses to rhinovirus infection and IL-13 treatment. Clin Transl Med 7 (1):13. https://doi.org/10.1186/s40169-018-0189-2
59. Colombo G, Garavaglia ML, Astori E, Giustarini D, Rossi R, Milzani A, Dalle-Donne I (2019) Protein carbonylation in human bronchial epithelial cells exposed to cigarette smoke extract. Cell Biol Toxicol 35(4):345–360
60. Fang T, Wang M, Xiao H, Wei X (2019) Mitochondrial dysfunction and chronic lung disease. Cell Biol Toxicol 35:493. https://doi.org/10.1007/s10565-019-09473-9
61. Ansari D, Torén W, Zhou Q, Hu D, Andersson R (2019) Proteomic and genomic profiling of pancreatic cancer. Cell Biol Toxicol 35(4):333–343
62. Qiao T, Wang X (2019) A new light of proteomics in cell biology and toxicology. Cell Biol Toxicol 35 (4):289–291
63. Wu D, Cheng Y, Wang X (2019) CSGT group. Definition of clinical gene tests. Cell Biol Toxicol 35 (2):83–87
64. Lan YL, Zou YJ, Lou JC, Xing JS, Wang X, Zou S, Ma BB, Ding Y, Zhang B (2019) The sodium pump α1 subunit regulates bufalin sensitivity of human glioblastoma cells through the p53 signaling pathway. Cell Biol Toxicol 35:521. https://doi.org/10.1007/s10565-019-09462-y
65. Kong S, Zhang Y (2019) Deciphering Hi-C: from 3D genome to function. Cell Biol Toxicol 35(1):15–32
66. Song D, Yang D, Powell CA, Wang X (2019) Cell-cell communication: old mystery and new opportunity. Cell Biol Toxicol 35(2):89–93
67. Song D, Tang L, Wang L, Huang J, Zeng T, Fang H, Wang X (2020) Roles of TGFβ1 in the expression of phosphoinositide 3-kinase isoform genes and sensitivity and response of lung telocytes to PI3K inhibitors. Cell Biol Toxicol 36:51. https://doi.org/10.1007/s10565-019-09487-3
68. Lopes-Pacheco M, Robba C, Rocco PRM, Pelosi P (2020) Current understanding of the therapeutic benefits of mesenchymal stem cells in acute respiratory distress syndrome. Cell Biol Toxicol 36:83. https://doi.org/10.1007/s10565-019-09493-5
69. Lee CH, Decker AM, Cackowski FC, Taichman RS (2019) Bone microenvironment signaling of cancer stem cells as a therapeutic target in metastatic prostate cancer. Cell Biol Toxicol. https://doi.org/10.1007/s10565-019-09483-7
70. Zahra D, Ramia S, Mélanie LL, Sara N, Hilary C, David M, Bertrand R, Luc F, Olivier J (2020) Protein and lipid homeostasis altered in rat macrophages after exposure to metallic oxide nanoparticles. Cell Biol Toxicol 36:65. https://doi.org/10.1007/s10565-019-09484-6
71. Jiang Y, Liang Y, Li L, Zhou L, Cheng W, Yang X et al (2019) Targeting neddylation inhibits intravascular survival and extravasation of cancer cells to prevent

lung-cancer metastasis. Cell Biol Toxicol 35 (3):233–245. https://doi.org/10.1007/s10565-019-09472-w
72. Giri AK, Ianevski A, Aittokallio T (2019) Genome-wide off-targets of drugs: risks and opportunities. Cell Biol Toxicol 35:485. https://doi.org/10.1007/s10565-019-09491-7
73. Li D, Zhou H, Zeng X (2019) Battling CRISPR-Cas9 off-target genome editing. Cell Biol Toxicol 35:403. https://doi.org/10.1007/s10565-019-09485-5
74. Wang DC, Wang X (2019) Off-target genome editing: a new discipline of gene science and a new class of medicine. Cell Biol Toxicol 35(3):179–183
75. Liang P, Huang J (2019) Off-target challenge for base editor-mediated genome editing. Cell Biol Toxicol 35 (3):185–187
76. Wang W, Hou J, Zheng N, Wang X, Zhang J (2019) Keeping our eyes on CRISPR: the "Atlas" of gene editing. Cell Biol Toxicol 35(4):285–288
77. Mills EM, Barlow VL, Luk LYP, Tsai YH (2019) Applying switchable Cas9 variants to in vivo gene editing for therapeutic applications. Cell Biol Toxicol. https://doi.org/10.1007/s10565-019-09488-2



Xiaoyang Chen is currently the Vice Director of Fujian Respiratory Medical Center. In addition, Dr. Chen is currently a member of the Tobacco Related Disease Research Group of the Chinese Thoracic Society, a member of the Chinese Society of Allergy, a standing committee member of the Allergy Society of Fujian Medical Association, and a member of Youth Committee of the Chinese Thoracic Society. His main research directions include chronic obstructive pulmonary disease, asthma, and sleep apnea syndrome. Dr. Chen has taken charge of two provincial and two ministerial scientific research topics and has obtained four invention patents and seven utility model patents. He is the author of six SCI articles.

Furong Yan, Master of Medicine, is a researcher at the Clinical Center for Molecular Diagnosis and Therapy of the Second Affiliated Hospital of Fujian Medical University. Her main research is focused on ion channel and signal transduction in cardiovascular diseases and lung diseases. She is the author of five scientific publications with the impact factor about 16.

Xiaoping Lin, is an associate chief physician of Pulmonary and Critical Care Medicine at the Second Affiliated Hospital of Fujian Medical University. She is a standing committee member both of the Allergy Prevention and Control of Fujian Medical Association and Respiratory Medicine Research Branch of Fujian Hospital Association. Her main research directions include chronic obstructive pulmonary disease, asthma, and other chronic airway inflammatory diseases. She is the author of more than ten scientific publications.

Liyong Shi, is an attending physician in the Pulmonary and Critical Care Medicine division of the Second Affiliated Hospital of Fujian Medical University. Her major research and clinical direction is chronic airway disease. She is the author of five scientific publications in China.

Xiangdong Wang, is a distinguished professor of medicine, director of Shanghai Institute of Clinical Bioinformatics, executive director of Clinical Science Institute of Fudan University Zhongshan Hospital, director of Fudan University Center of Clinical Bioinformatics, deputy director of Shanghai Respiratory Research Institute, and visiting professor of King's College London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 300 scientific publications with the impact factor about 900, citation number about 6920, h-index 48, i10-index 221, and cited journal impact factor about 8000.

Yiming Zeng, Professor, is a chairman of Academic Committee of the Second Affiliated Hospital of Fujian Medical University, expert of State Council Expert for Special Allowance, and director of Sleep Medicine Key Laboratory of Fujian Province. He achieved the Outstanding Contribution of Middle-aged Expert of National Health and Family Planning Commission of the People's Republic of China (NHFPC). His main research is focused on clinical, interventional pulmonology, sleep-breathing disorders, and noninvasive mechanical ventilation. He is the author of more than 150 scientific publications.

The Role of RASSF1 Methylation in Lung Carcinoma

8

Tingting Zhang, Yanjun Li, Hui Zhang, Xiangdong Wang, Xiaozhuan Liu, and Li Li

Abstract

Lung carcinoma is the most frequently diagnosed malignant neoplasms and mainly consists of small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). Large number of lung carcinoma patients have poor outcomes due to the late diagnosis and the limited therapeutic options. Previous attempts have proved that the evolution of lung carcinoma is a multistep molecular aberration which various genetic or epigenetic alterations may be take part in. Among these molecular aberrations, the inactivation of tumor suppressor gene has been widely observed in all types of carcinoma including lung carcinoma. As a vital inactivated mechanism, DNA methylation of tumor suppressor gene is frequently found in lung cancer. To gain exhaustive comprehension of the carcinogenesis of lung carcinoma, we summarize our current knowledge on DNA methylation of RASSF1 (RAS-Association Domain Family 1) and its clinical significance in lung carcinoma.

Keywords

Lung carcinoma · RASSF1 · Tumor-suppressor gene · DNA methylation

Abbreviations

SCLC	Small-cell lung carcinoma
NSCLC	Non-small-cell lung carcinoma
RASSF1	RAS-Association Domain Family 1
TEADs	Transcriptional enhancer factors
TAZ	WW domain-containing transcriptional coactivators
YAP	TAZ paralog
hMOB1	MOB kinase activator 1B
hSAV1	The adaptor/scaffold proteins Salvador homolog 1

Tingting Zhang, Yanjun Li, and Hui Zhang contributed equally to this work.

T. Zhang · Y. Li · H. Zhang · X. Liu (✉)
Clinical Center for Single Cell Biomedicine, Henan Provincial People's Hospita, Zhengzhou, Henan, China

People's Hospital of Zhengzhou University, Zhengzhou, Henan, China

People's Hospital of Henan University, Zhengzhou, Henan, China

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China

L. Li
Department of Scientific Research and Discipline Construction, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

Clinical Center for Single Cell Biomedicine, Henan Provincial People's Hospita, Zhengzhou, Henan, China
e-mail: lili@henu.edu.cn

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_8

LATS1 and LATS2	Large tumor suppressor 1 and 2
MST1 and MST2	Mammalian sterile 20-like kinase 1 and 2
LOH	Heterozygosity

8.1 Introduction

Lung carcinoma is the leading cause of cancer-related death, with an estimated 388,000 deaths in Europe in 2018 [1, 2]. Based upon the data of smoking prevalence from the population-based Adult Health Survey in 2003, the estimated lung cancer mortality was 15.0 and 7.1 per 100,000 among men and women in 2018 [1], respectively. In China, the incidence of lung carcinoma is also high, with the highest mortality rate as compared with other countries [2]. The uptake of tobacco among males and exposure to unventilated cooking fumes among females are the predominant non-genetic risk factors for lung carcinoma [3–7]. The 5-year survival rate of lung carcinoma is very low, especially in Eastern Asia, due to the large proportion of lung carcinoma patients present with advanced metastatic tumors when diagnosed [8, 9].

Lung carcinoma mainly consists of small-cell lung carcinoma (SCLC) as the most aggressive lung carcinoma accounted for about 25% of bronchogenic carcinomas and non-small-cell lung carcinoma (NSCLC) as the most common lung carcinoma subtype for approximately 85% of lung cancer cases [10, 11]. The major histological subtypes of NSCLC are represented by lung adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [12], of which the resection is performed in the early stage and chemo-combination in the late stage, with the mean 5-year survival of 15%. The large number of lung cancer patients have poor outcomes due to the late diagnosis, acquired multidrug resistance, and complex mechanisms [13]. This chapter aims at exploring the comprehensive mechanisms on the carcinogenesis of lung cancer by furthermore understanding DNA methylation of RAS-Association Domain Family 1 (RASSF1) and its clinical significance in lung carcinoma. We pay more specific attention on the potential mechanisms and new specific molecular markers of lung cancer, especially DNA methylation of tumor suppressor genes and inactivated genes in the development of lung carcinoma.

8.2 The RASSF1 Gene

RASSF1 is one of the key tumor-suppressor genes allocated in chromosome 3p21.3 and spans about 11,151 bp [14, 10]. RASSF1 promotes apoptosis, microtubule stability and polymerization, and mitotic progression [15]. The protein encoded by RASSF1 can participate in RAS-related cellular signal pathways and regulate oncogenesis, cell proliferation, differentiation, and apoptosis in a wide variety of cancer types [16]. Eight transcripts, i.e., RASSF1A, B, C, D, E, F, G, and H, are generated by RASSF1 gene and contain a Ras-Association (RA) domain in the carboxyterminal segments, except for RASSF1F-H which is similar to the RAS effector proteins, Raf1. Raf1 is associated with Ras-GTP to activate Ras proteins, suppress cell growth, and promote proapoptotic effects.

RASSF1A and RASSF1C are two predominant common isoforms and encode an ATM-kinase phosphorylation site and a conserved carboxyterminal SARAH (Sav/RASSF/Hpo) domain as a key component of the Hippo signaling pathway, except for the RA domain. RASSF1A has a diacylglycerol/phorbol ester-binding (DAG) domain containing a central zinc finger which is also known as the protein kinase C conserved domain (C1 domain). RASSF1C variant is shorter than RASSF1A and lacks the amino terminal C1 domain. RASSF1D and E have the RA, SARAH, C1 domains and ATM-kinase phosphorylation site similar to RASSF1A in structure. RASSF1B contains one RA and SARAH domain, respectively. Isoforms

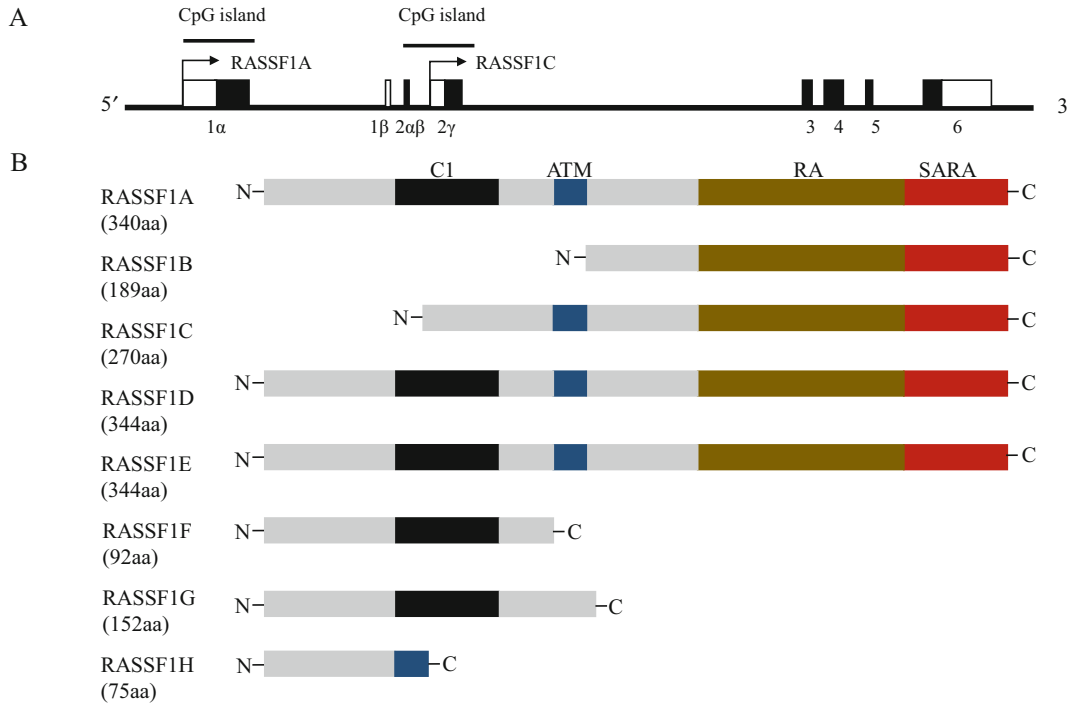


Fig. 8.1 (**a**) Schematic map of RASSF1 locus. Black boxes indicate exons and open boxes indicate untranslated regions, respectively. Two CpG islands are shown by black lines. The transcription start sites of isoform A and isoform C are indicated with black arrows. (**b**) The domain structures of polypeptides encoded by RASSF1. C1, DAG/diacylglycerol binding domain (black) putative ATM kinase phosphorylation consensus sequence motif (blue) RA, Ras-association domain (brown) and SARAH, Sav/RASSF/Hpo interaction domain (red)

F/G and H have a C1 domain and an ATM-kinase phosphorylation site, respectively (Fig. 8.1).

RASSF1C appears to share many of the biological characteristics of RASSF1A. On basis of the similar structure to Ras effector, RASSF1 gene regulates cell proliferation, differentiation, and apoptosis. RASSF1A functions as a negative regulator of cell proliferation by blocking the cell cycle progression at the level of G1/S-phase [17] and has the dual role in the coordination of p53 and p73 responses [18], while RASSF1C exhibits growth inhibitory potency [19], although there is little known on functions of other variants. RASSF1B, D, and E are found poorly expressed in hemopoietic, cardiac, and pancreatic cells, respectively. RASSF1F, E, D, and G share the same promoter region with RASSF1A, although the biological significance remains unclear.

RASSF2, RASSF3, NORE1, and RASSF6 were identified as the homolog of RASSF1, which share similar Ras-association domain with RASSF1. These genes have SARAH domain and code multiple transcripts. RASSF2 shares a lower homology (29% identity) with RASSF1, while acting as a tumor suppressor gene and undergoing promoter methylation at high frequency similar to RASSF1 [20]. The inactivation of RASSF2 may be associated with tumor progression [20], and RASSF3 suppresses tumor formation through interacting with MDM2 and inducing NSCLC cell apoptosis [21]. NORE1 shares about 50% sequence identities with RASSF1 and has similar pattern of mRNA transcript expression and function as a tumor-suppressor gene [22–24]. RASSF6 is found frequently suppressed in several human cancers [20, 25–27].

8.3 Inactivation of RASSF1A by DNA Methylation in Lung Carcinoma

The genomic mutational landscape provided evidences that genetic alterations taken part in the tumorigenesis of lung cancer. The epigenetic regulation provides a novel insight in the progression and evolution of lung cancer [28, 29]. Of the epigenetic modifications, DNA methylation mainly occurs in C-G dinucleotide-rich regions, also named CpG islands [30], where the methyl group is added by DNA methyltransferase at the 5-position cytosine and erased by demethylase. DNA methylation mainly occurs at the cytosine--phosphate-guanine (CpG) island which locates in promoter region of a gene and regulates the expression of gene, which plays a vital role in genomic imprint erasure, instability of chromatin structure, and X-chromosome inactivation. The aberrant DNA methylation interacts with gene expression in the early stage of human cancers and dynamically during lung carcinogenesis. A lot of methylated genes have been identified in lung carcinoma, including RASSF1, major tumor suppressor 1, fragile histidine triad, methylguanine-DNA methyltransferase, and adenomatosis polyposis coli tumor suppressor.

The loss of heterozygosity (LOH) is the most frequent event during lung tumorigenesis [31], while rarely attributed to somatic mutations, except for one frame-shift and missense mutation identified in nasopharyngeal carcinomas [32]. RASSF1 is inactivated frequently by the hypermethylation of the promoter CpG island in cancers [33–37]. RASSF1 methylation was originally reported in lung cancer and then shown as the common event in cancers [33, 38]. RASSF1A was methylation-inactivated in SCLC, while aberrant methylation of the RASSF1C CpG island promoter was not observed in lung cancer [33].

8.4 Signaling Pathway Involving RASSF1 in Lung Cancer

Among signaling pathways, RASSF1A contributes to the carcinogenesis of lung cancer mainly through Hippo signaling pathways. The Hippo pathway (i.e., Salvador-Warts-Hippo pathway) in a kinase cascade regulates the organ size through regulating cell proliferation, differentiation, and apoptosis [39–41]. The core components of the pathway encompass the mammalian sterile 20-like kinase 1 and 2 (MST1 and MST2) and the large tumor suppressor 1 and 2 (LATS1 and LATS2), and cooperate with the adaptor/scaffold proteins, Salvador homolog 1 (hSAV1), and MOB kinase activator 1A and 1B (hMOB1). The downstream effectors of Hippo pathway are two WW domain-containing transcriptional coactivators TAZ and its paralog YAP. Mst1/2 phosphorylates hSAV1 and forms the activated Mst1/2-hSAV1 complex which cooperates with hMOB1 and activates LATS1/2. After then LATS1/2 phosphorylates YAP/TAZ which is prevented from entering to the nucleus. Then the complex with transcriptional enhancer factors (TEADs) is formed and the expression of anti-apoptotic and pro-proliferative genes are activated [42–44].

During DNA damage, RASSF1A activated by ATM can induce apoptosis through the interaction of Hippo pathway with MST1/MST2 via the C-terminus to prevent the autophosphorylation of those protein kinases [18, 44, 45]. The components of the Hippo pathway are intimately involved in lung morphogenesis and tumorigenesis [46–48]. The abnormal expression of those components is associated with the clinical classification, poor differentiation, metastasis, and poor prognosis and survival in lung cancer [47, 49–52]. The DNA methylation of promoter results in the inactivation of RASSF1A, RASSF1A-MST1/MST2 complex, and dysfunction of the Hippo pathway. RASSF1A can enhance the transcription of proapoptotic genes

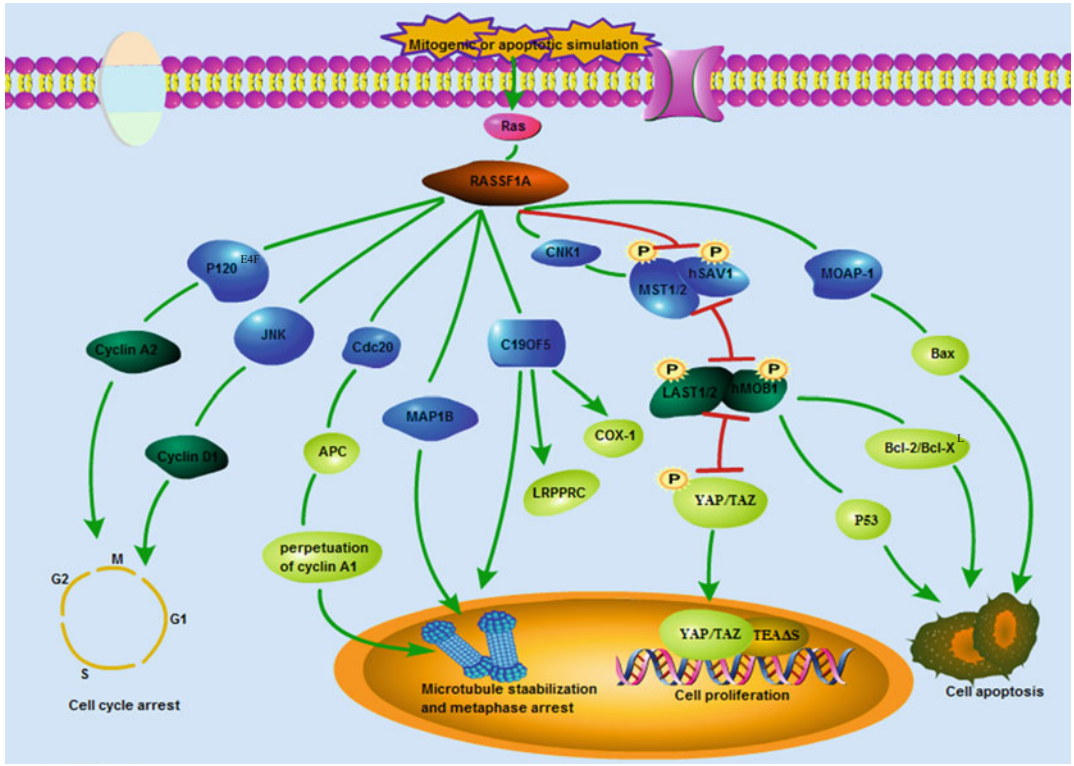


Fig. 8.2 A summary of RASSF1A pathways in cancergenesis of lung cancer. RASSF1A regulates cell apoptosis through its interactions with the connector enhancer of KSR (CNK1), the proapoptotic kinase MST1, and the modulator of apoptosis-1 (MAP-1). The CNK1–MST1 complex is also thought to play important role in cell proliferation. RASSF1A can regulate the microtubule network by recruiting effectors of the microtubule-associated protein 1B (MAP1B), C19ORF5, and the Cdc20. RASSF1A also induced G1 and S-phase cell cycle arrest through inhibiting the transcription factor $p120^{E4F}$ (RASSF1A was enhanced by $p120^{E4F}$) and JNK

through the formation of a complex with YAP and p73 in the nucleus (Fig. 8.2).

RASSF1A may contribute to the carcinogenesis of lung cancer through microtubules and binding of Cdc20 via an N-terminal region. Cdc20 cannot bind with APC and fail to form the complex Cdc20-APC for the spindle assembly checkpoint during mitosis [53]. RASSF1A is required for stabilizing the microtubule. RASSF1A controls the motility and invasion of lung cancer cells through the modulation of tubulin dynamics [54, 55]. The promoter hypermethylation of RASSF1A activates premature APC, following by accelerated cell division, mitotic spindle abnormalities, and chromosome misalignment [53] (Fig. 8.2). The exogenous expression of RASSF1A modulates levels of cyclin D1 and induces cell cycle arrest in lung carcinoma cells [17]. RASSF1A inhibits lung cancer cell growth through reducing the phosphorylation of JNK [56] (Fig. 8.2).

8.5 Clinical Significance of RASSF1 in Lung Carcinoma

RASSF1 methylation in cancer may serve an important role in clinical utilities, especially in lung cancer. For example, the aberrant RASSF1A methylation may be an ideal biomarker for early diagnostic and prognostic due to the non-invasive, high sensitivity, and high

specificity characteristics. It is questionable whether RASSF1A methylation can be a powerful marker for patient prognosis at early stage of lung cancer. RASSF1A exhibited lung cancer-specific methylation pattern, with the hypermethylation level up to 100% in SCLC and 63% in NSCLC [57, 58]. RASSF1A methylation can be detected in body fluids including blood, urine, sputum, and bronchial alveolar lavages [56–59]. For example, RASSF1A methylation is observed in the blood of patients with NSCLC [59]. The RASSF1A methylation of bronchial aspirates was 21% in patients with lung cancer and smoking and 1% in patients with lung cancer alone, respectively. The methylation level of RASSF1A was associated with the number of cigarette packs and smoking years during the lifetime of patients with lung cancer [60]. The RASSF1 methylation of bronchial washings was found to have diagnostic sensitivity [61], which has the great potential to screen risk populations of patients with lung cancer. DNA methylation of RASSF1A is correlated with poor clinicopathological characteristics in nearly all solid tumors [62], which also includes lung cancer. RASSF1 promoter methylation was found in poorly differentiated tumors [63–65], associated with tumor grades, stages, and survival. For example, RASSF1A methylation was associated with patient survival time in lung adenocarcinoma [66]. Decreased survival time was observed in NSCLC patients with RASSF1A methylation, irrespective of whether patients have received adjuvant radio therapy or surgical treatment [58, 64, 67, 68]. On basis of those evidence, RASSF1 and isoforms as disease biomarkers should be furthermore evaluated, since disease biomarkers are expected to have the clear specificity for disease per se, disease stage, phase, severity, duration, or response to therapy [69–76]. Several natural compounds can regulate DNMT activity or expression to re-activate RASSF1A [77]. Peperomin E, as a natural bioactive secolignan polyphenol extracted from the plant peperomia dindygulensis, could demethylate RASSF1A and upregulate the expression of RASSF1A by reducing the level of DNMT1 in lung cancer cells [78].

8.6 Conclusion

Epigenetics changes especially DNA methylation has been proved to take part in the carcinogenesis of cancers. The DNA methylation of the tumor suppressor genes may be exploitable for the biologic and clinical significance of cancers. Overall, as the common tumor suppressor gene of lung cancer, evidence have suggested the DNA methylation of RASSF1 can be an essential potential clinic diagnostic or prognostic marker and may provide new therapeutic strategies for future successful treatment of lung cancer. It will be very interesting to further explore how to develop non-invasive, rapid and less cost detection methods for DNA methylation and to confirm the reliability and sensitivity of DNA methylation.

Acknowledgement This work is supported by the project of "Henan province clinical biomedical big data" by Department of science and technology of Henan Province (CXJD2019002).

References

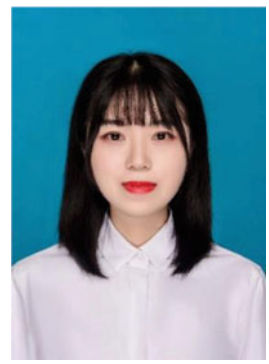
1. Torre LA, Siegel RL, Jemal A (2016) Lung cancer statistics. Adv Exp Med Biol 893:1–19
2. Ferlay J, Colombet M, Soerjomataram I et al (2018) Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 103:356
3. Winkler V, Mangolo NJ, Becher H (2015) Lung cancer in South Africa: a forecast to 2025 based on smoking prevalence data. BMJ Open 5:e006993
4. Wang L, Yu C, Liu Y et al (2016) Lung cancer mortality trends in China from 1988 to 2013: new challenges and opportunities for the government. Int J Environ Res Public Health 13:1052
5. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. CA Cancer J Clin 68:7–30
6. Hong S, Mok Y, Jeon C et al (2016) Tuberculosis, smoking and risk for lung cancer incidence and mortality. Int J Cancer 139:2447–2455
7. Raaschou-Nielsen O, Andersen ZJ, Beelen R et al (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol 14:813–822
8. Mirtavoos-Mahyari H, Ghafouri-Fard S, Khosravi A, Motevaseli E, Esfahani-Monfared Z, Seifi S, Salimi B et al (2019) Circulating free DNA concentration as a

marker of disease recurrence and metastatic potential in lung cancer. Clin Transl Med 8(1):14. https://doi.org/10.1186/s40169-019-0229-6
9. Phan TT, Tran BT, Nguyen ST, Ho TT, Hang Nguyen HT, Le VT, Le AT (2019) EGFR plasma mutation in prediction models for resistance with EGFR TKI and survival of non-small cell lung cancer. Clin Transl Med 8(1):4. https://doi.org/10.1186/s40169-019-0219-8
10. Virmani AK, Gazdar AF (2003) Tumor suppressor genes in lung cancer. Methods Mol Biol 222:97–115
11. Devarakonda S, Morgensztern D, Govindan R (2015) Genomic alterations in lung adenocarcinoma. Lancet Oncol 16:e342–e351
12. The World Health Organization (1982) Histological typing of lung tumours. Neoplasma 29:111–123
13. Solyanik GI (2010) Multifactorial nature of tumor drug resistance. Exp Oncol 32:181–185
14. Hesson LB, Cooper WN, Latif F (2007) Evaluation of the 3p21.3 tumour-suppressor gene cluster. Oncogene 26:7283–7301
15. Vos MD, Ellis CA, Bell A et al (2000) Ras uses the novel tumor suppressor RASSF1 as an effector to mediate apoptosis. J Biol Chem 275:35669–35672
16. Zhou Y, Zhang X, Klibanski A (2014) Genetic and epigenetic mutations of tumor suppressive genes in sporadic pituitary adenoma. Mol Cell Endocrinol 386:16–33
17. Shivakumar L, Minna J, Sakamaki T et al (2002) The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol Cell Biol 22:4309–4318
18. Hamilton G, Yee KS, Scrace S et al (2009) ATM regulates a RASSF1A-dependent DNA damage response. Curr Biol 19:2020–2025
19. Li J, Wang F, Protopopov A et al (2004) Inactivation of RASSF1C during in vivo tumor growth identifies it as a tumor suppressor gene. Oncogene 23:5941–5949
20. Guo W, Dong Z, Guo Y et al (2016) Decreased expression and frequent promoter hypermethylation of RASSF2 and RASSF6 correlate with malignant progression and poor prognosis of gastric cardia adenocarcinoma. Mol Carcinog 55:1655–1666
21. Fukatsu A, Ishiguro F, Tanaka I et al (2014) RASSF3 downregulation increases malignant phenotypes of non-small cell lung cancer. Lung Cancer 83:23–29
22. Hesson L, Dallol A, Minna JD et al (2003) NORE1A, a homologue of RASSF1A tumour suppressor gene is inactivated in human cancers. Oncogene 22:947–954
23. Aoyama Y, Avruch J, Zhang XF (2004) Nore1 inhibits tumor cell growth independent of Ras or the MST1/2 kinases. Oncogene 23:3426–3433
24. Vos MD, Martinez A, Ellis CA et al (2003) The pro-apoptotic Ras effector Nore1 may serve as a Ras-regulated tumor suppressor in the lung. J Biol Chem 278:21938–21943
25. Liang YY, Deng XB, Zeng LS et al (2018) RASSF6-mediated inhibition of Mcl-1 through JNK activation improves the anti-tumor effects of sorafenib in renal cell carcinoma. Cancer Lett 432:75–83
26. Mi Y, Zhang D, Jiang W et al (2017) miR-181a-5p promotes the progression of gastric cancer via RASSF6-mediated MAPK signalling activation. Cancer Lett 389:11–22
27. Chen E, Yang F, He H et al (2016) Decreased level of RASSF6 in sporadic colorectal cancer and its anti-tumor effects both in vitro and in vivo. Oncotarget 7:19813–19823
28. Ma H, Chen X, Hu H et al (2018) Hypermethylation of MDFI promoter with NSCLC is specific for females, non-smokers and people younger than 65. Oncol Lett 15:9017–9024
29. Grasse S, Lienhard M, Frese S et al (2018) Epigenomic profiling of non-small cell lung cancer xenografts uncover LRP12 DNA methylation as predictive biomarker for carboplatin resistance. Genome Med 10:55
30. Toh TB, Lim JJ, Chow EK-H (2019) Epigenetics of hepatocellular carcinoma. Clin Transl Med 8(1):13. https://doi.org/10.1186/s40169-019-0230-0.
31. Wistuba II, Behrens C, Virmani AK et al (2000) High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res 60:1949–1960
32. Lo KW, Kwong J, Hui AB et al (2001) High frequency of promoter hypermethylation of RASSF1A in nasopharyngeal carcinoma. Cancer Res 61:3877–3881
33. Dammann R, Li C, Yoon JH et al (2000) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet 25:315–319
34. Byun DS, Lee MG, Chae KS et al (2001) Frequent epigenetic inactivation of RASSF1A by aberrant promoter hypermethylation in human gastric adenocarcinoma. Cancer Res 61:7034–7038
35. Kuroki T, Trapasso F, Yendamuri S et al (2003) Allele loss and promoter hypermethylation of VHL, RAR-beta, RASSF1A, and FHIT tumor suppressor genes on chromosome 3p in esophageal squamous cell carcinoma. Cancer Res 63:3724–3728
36. Lee MG, Kim HY, Byun DS et al (2001) Frequent epigenetic inactivation of RASSF1A in human bladder carcinoma. Cancer Res 61:6688–6692
37. Wong N, Li L, Tsang K et al (2002) Frequent loss of chromosome 3p and hypermethylation of RASSF1A in cholangiocarcinoma. J Hepatol 37:633–639
38. Dammann R, Takahashi T, Pfeifer GP (2001) The CpG island of the novel tumor suppressor gene RASSF1A is intensely methylated in primary small cell lung carcinomas. Oncogene 20:3563–3567
39. Yu FX, Zhao B, Guan KL (2015) Hippo pathway in organ size control, tissue homeostasis, and cancer. Cell 163:811–828
40. Grusche FA, Degoutin JL, Richardson HE et al (2011) The Salvador/Warts/Hippo pathway controls

regenerative tissue growth in Drosophila melanogaster. Dev Biol 350:255–266
41. Dobrokhotov O, Samsonov M, Sokabe M, Hirata H (2018) Mechanoregulation and pathology of YAP/TAZ via Hippo and non-Hippo mechanisms. Clin Transl Med 7(1):23. https://doi.org/10.1186/s40169-018-0202-9
42. Hong JH, Hwang ES, McManus MT et al (2005) TAZ, a transcriptional modulator of mesenchymal stem cell differentiation. Science 309:1074–1078
43. Bora-Singhal N, Nguyen J, Schaal C et al (2015) YAP1 regulates OCT4 activity and SOX2 expression to facilitate self-renewal and vascular mimicry of stem-like cells. Stem Cells 33:1705–1718
44. Lehmann W, Mossmann D, Kleemann J et al (2016) ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. Nat Commun 7:10498
45. Pefani DE, Latusek R, Pires I et al (2014) RASSF1A-LATS1 signalling stabilizes replication forks by restricting CDK2-mediated phosphorylation of BRCA2. Nat Cell Biol 16:962–971. 1–8
46. Lange AW, Sridharan A, Xu Y et al (2015) Hippo/Yap signaling controls epithelial progenitor cell proliferation and differentiation in the embryonic and adult lung. J Mol Cell Biol 7:35–47
47. Lau AN, Curtis SJ, Fillmore CM et al (2014) Tumor-propagating cells and Yap/Taz activity contribute to lung tumor progression and metastasis. EMBO J 33:468–481
48. Luo SY, Sit KY, Sihoe AD et al (2014) Aberrant large tumor suppressor 2 (LATS2) gene expression correlates with EGFR mutation and survival in lung adenocarcinomas. Lung Cancer 85:282–292
49. Dhanasekaran SM, Balbin OA, Chen G et al (2014) Transcriptome meta-analysis of lung cancer reveals recurrent aberrations in NRG1 and Hippo pathway genes. Nat Commun 5:5893
50. Su LL, Ma WX, Yuan JF et al (2012) Expression of Yes-associated protein in non-small cell lung cancer and its relationship with clinical pathological factors. Chin Med J 125:4003–4008
51. Kim JM, Kang DW, Long LZ et al (2011) Differential expression of Yes-associated protein is correlated with expression of cell cycle markers and pathologic TNM staging in non-small-cell lung carcinoma. Hum Pathol 42:315–323
52. Noguchi S, Saito A, Horie M et al (2014) An integrative analysis of the tumorigenic role of TAZ in human non-small cell lung cancer. Clin Cancer Res 20:4660–4672
53. Song MS, Song SJ, Ayad NG et al (2004) The tumour suppressor RASSF1A regulates mitosis by inhibiting the APC-Cdc20 complex. Nat Cell Biol 6:129–137
54. Liu L, Tommasi S, Lee DH et al (2003) Control of microtubule stability by the RASSF1A tumor suppressor. Oncogene 22:8125–8136
55. Dallol A, Agathanggelou A, Tommasi S et al (2005) Involvement of the RASSF1A tumor suppressor gene in controlling cell migration. Cancer Res 65:7653–7659
56. Whang YM, Kim YH, Kim JS et al (2005) RASSF1A suppresses the c-Jun-NH2-kinase pathway and inhibits cell cycle progression. Cancer Res 65:3682–3690
57. Jimenez AP, Traum A, Boettger T et al (2017) The tumor suppressor RASSF1A induces the YAP1 target gene ANKRD1 that is epigenetically inactivated in human cancers and inhibits tumor growth. Oncotarget 8:88437–88452
58. Burbee DG, Forgacs E, Zochbauer-Muller S et al (2001) Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst 93:691–699
59. Ramirez JL, Taron M, Balana C et al (2003) Serum DNA as a tool for cancer patient management. Rocz Akad Med Bialymst 48:34–41
60. Schmiemann V, Bocking A, Kazimirek M et al (2005) Methylation assay for the diagnosis of lung cancer on bronchial aspirates: a cohort study. Clin Cancer Res 11:7728–7734
61. van der Drift MA, Prinsen CFM, Knuiman GJ et al (2012) Diagnosing peripheral lung cancer: the additional value of the Ras-association domain family 1A gene methylation and Kirsten rat sarcoma 2 viral oncogene homolog mutation analyses in washings in nondiagnostic bronchoscopy. Chest 141:169–175
62. Grawenda AM, O'Neill E (2015) Clinical utility of RASSF1A methylation in human malignancies. Br J Cancer 113:372–381
63. Drilon A, Sugita H, Sima CS et al (2014) A prospective study of tumor suppressor gene methylation as a prognostic biomarker in surgically resected stage I to IIIA non-small-cell lung cancers. J Thorac Oncol 9:1272–1277
64. Wang J, Lee JJ, Wang L et al (2004) Value of p16INK4a and RASSF1A promoter hypermethylation in prognosis of patients with resectable non-small cell lung cancer. Clin Cancer Res 10:6119–6125
65. Pelosi G, Fumagalli C, Trubia M et al (2010) Dual role of RASSF1 as a tumor suppressor and an oncogene in neuroendocrine tumors of the lung. Anticancer Res 30:4269–4281
66. Tomizawa Y, Kohno T, Kondo H et al (2002) Clinicopathological significance of epigenetic inactivation of RASSF1A at 3p21.3 in stage I lung adenocarcinoma. Clin Cancer Res 8:2362–2368
67. Kim DH, Kim JS, Ji YI et al (2003) Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. Cancer Res 63:3743–3746
68. Wang J, Wang B, Chen X et al (2011) The prognostic value of RASSF1A promoter hypermethylation in non-small cell lung carcinoma: a systematic review and meta-analysis. Carcinogenesis 32:411–416

69. Wang X (2018) Clinical trans-omics: an integration of clinical phenomes with molecular multiomics. Cell Biol Toxicol 34(3):163–166. https://doi.org/10.1007/s10565-018-9431-3
70. Qiao T, Wang X (2019) A new light of proteomics in cell biology and toxicology. Cell Biol Toxicol 35 (4):289–291. https://doi.org/10.1007/s10565-019-09492-6
71. Wu D, Cheng Y, Wang X (2019) CSGT group. Definition of clinical gene tests. Cell Biol Toxicol 35 (2):83–87. https://doi.org/10.1007/s10565-019-09464-w
72. Song D, Yang D, Powell CA, Wang X (2019) Cell-cell communication: old mystery and new opportunity. Cell Biol Toxicol 35(2):89–93. https://doi.org/10.1007/s10565-019-09470-y
73. Chirshev E, Oberg KC, Ioffe YJ, Unternaehrer JJ (2019) Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. Clin Transl Med 8(1):24. https://doi.org/10.1186/s40169-019-0240-y
74. Malm J, Sugihara Y, Szasz M, Kwon HJ, Lindberg H, Appelqvist R, Marko-Varga G (2018) Biobank integration of large-scale clinical and histopathology melanoma studies within the European Cancer Moonshot Lund Center. Clin Transl Med 7(1):28. https://doi.org/10.1186/s40169-018-0203-8
75. Omrani S, Taheri M, Omrani MD, Arsang-Jang S, Ghafouri-Fard S (2019) The effect of omega-3 fatty acids on clinical and paraclinical features of intractable epileptic patients: a triple blind randomized clinical trial. Clin Transl Med 8(3):3. https://doi.org/10.1186/s40169-019-0220-2
76. Qi X, Yu C, Wang Y, Lin Y, Shen B (2019) Network vulnerability-based and knowledge-guided identification of microRNA biomarkers indicating platinum resistance in high-grade serous ovarian cancer. Clin Transl Med 8(1):28. https://doi.org/10.1186/s40169-019-0245-6
77. Fang M, Chen D, Yang CS (2007) Dietary polyphenols may affect DNA methylation. J Nutr 137:223S–228S
78. Wang XZ, Cheng Y, Wang KL et al (2016) Peperomin E reactivates silenced tumor suppressor genes in lung cancer cells by inhibition of DNA methyltransferase. Cancer Sci 107:1506–1519

Tingting Zhang received her postgraduate degree in People's Hospital of Zhengzhou University. She has strong background of epigenetic modification of oocytes. Currently, Tingting Zhang is interested in next generation sequencing data analysis and interpretation, especially on single-cell analysis and cancer research. She has published three SCI papers in the past 3 years.



Yanjun Li did her Master's degree in Cell Biology at Zhengzhou University. Now, she works as a research assistant at Center for Clinical Single Cell Biomedicine in Henan Provincial People's Hospital. Her main research direction is focused on lung cancer. Currently, Yanjun Li is interested in the epigenetic alterations in lung cancer, especially in DNA methylation.

Hui Zhang is an editor-in-chief assistant at Henan Provincial People's Hospital and a postgraduate in Zhengzhou University. She has read a large number of scientific articles on topics such as single-cell sequencing, gene editing, molecular drug targeting, stem cells, immune therapy, and heterogeneity during her editing work. In addition, years of experience in clinical surgery and ethics committee gave her a deep understanding of medical practice ethics and scientific research ethics.

Xiangdong Wang is a Distinguished Professor of Medicine, Director of Shanghai Institute of Clinical Bioinformatics, Executive Director of Clinical Science Institute of Fudan University Zhongshan Hospital, Director of Fudan University Center of Clinical Bioinformatics, Deputy Director of Shanghai Respiratory Research Institute, and visiting professor of King's College of London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 200 scientific publications with the impact factor about 900, citation number about 6918, h-index 46, i10-index 181, and cited journal impact factor about 7000.

Xiaozhuan Liu received her PhD on Epidemiology and Health Statistics at Zhengzhou University, and work as Researcher in the Center for Clinical Single Cell Biomedicine of Henan Provincial People's Hospital. She mainly engaged in the research of tumor pathogenesis. Currently, Xiaozhuan Liu is interested in the role of gene methylation in cancers.

Li Li is a Director of Department of Science Research and Discipline Construction, Henan Provincial people's Hospital. She is Member of clinical research group of Chinese Medical Association's Scientific Research Management Branch, Standing Committee member of Chinese Medical Association's Henan Research and Management Branch, and Vice-chairman of Henan Discipline Management Branch of Chinese Hospital Management Society. She has engaged in the management of medical scientific research for 30 years and her main research is focused on health management scientific research big data, laboratory biosafety, and medical ethics. She published more than 20 scientific papers.

Single Cell Sequencing: A New Dimension in Cancer Diagnosis and Treatment

9

Fengying Wu, Jue Fan, Jingwen Fang, Priya S. Dalvi, Margarete Odenthal, and Nan Fang

Abstract

Cancer is one of the leading causes of death worldwide and well known for its complexity. Cancer cells within the same tumor or from different tumors are highly heterogeneous. Furthermore, stromal and immune cells within tumor microenvironment interact with cancer cells to play important roles in how tumors progress and respond to different treatments. Recent advances in single cell technologies, especially massively parallel single cell sequencing, have made it possible to analyze cancer cells and cells in its tumor microenvironment in parallel with unprecedented high resolution. In this chapter, we will review recent developments in single cell sequencing technologies and their applications in cancer research. We will also explain how insights generated from single cell sequencing can be used to develop novel diagnostic and therapeutic approaches to conquer cancer.

Keywords

Single cell sequencing · Tumor microenvironment · Diagnosis · Therapy

9.1 Background

Cancer is one of the leading causes of death in the world, accounting for 18.1 million new cases and 9.6 million deaths in 2018 [1]. Over the last decades, the treatment of malignant tumors has undergone rapid improvements with the use of targeted therapy and immunotherapy and the overall survival rate of cancer patients has increased significantly. For example, the 5-year survival rate of advanced lung cancer patients increased from 3% to current 16% after use of anti-PD1/PD-L1 immunotherapy [2]. However, big challenges remain. For targeted therapy, there are patients with actionable mutations who cannot benefit from tyrosine kinase inhibitors (TKIs), while the underlining mechanism for such primary resistance is still unclear. Besides primary resistance, almost all patients will eventually develop resistance to the TKIs. For

Fengying Wu and Jue Fan contributed equally to this work.

F. Wu
Department of Medical Oncology, Shanghai Pulmonary Hospital, Tongji University, Shanghai, China

J. Fan · J. Fang
Singleron Biotechnologies, Nanjing, Jiangsu, China

P. S. Dalvi · M. Odenthal
Institute for Pathology, University Hospital of Cologne, Cologne, Germany

Centre for Molecular Medicine Cologne, University of Cologne (UoC), Cologne, Germany

N. Fang (✉)
Singleron Biotechnologies, Nanjing, Jiangsu, China

Department of Pathology, Anhui Medical University, Hefei, Anhui, China
e-mail: nan@singleron.com

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_9

example, non-small cell lung cancer (NSCLC) patients with EGFR mutations who are treated with EGFR-TKIs often develop resistance by gain of additional mutations in the EGFR locus, such as the common gatekeeper T790M mutation (50–60%) [3] or mutations at EGFR codons D761 [4], L747 [5], or C797 [6]. Additionally, other genetic alterations such as amplification of MET or Her2 genes also contribute to acquired resistance [7].

Immune therapy has changed the paradigm of cancer treatment. Immune checkpoint inhibitors (ICIs) such as PD1 or PD-L1 antibodies have been approved in clinical practice for many cancer types [8, 9]. Despite substantial improvement in duration of response (DOR), the overall response rate (ORR) is only about 20% when used as monotherapy in unselected patients, and 45% in patients with high PD-L1 expression or a high tumor mutation burden (TMB) [8], which accounts for only 30% of all patients. There remains a large number of NSCLC patients who cannot derive benefit from ICIs. Furthermore, some PD-L1-negative patients can also benefit from ICIs, which suggests that PD-L1 expression alone is not sufficient as a reliable predictive biomarker [10]. Are there better predictive biomarkers or models to stratify patients? Are there better strategies to improve the efficacy of immune therapy in current non-responders? Are there other effective targets for immune therapy? To answer these questions and develop better treatment options, a comprehensive understanding of tumor heterogeneity and tumor microenvironment (TME) is necessary.

With the application of molecular biology and "omics" tools, the process of tumor genesis is now understood to originate from DNA mutations that influence the gene regulatory network of critical cellular processes. However, there is no fixed mutation pattern that leads to tumor development and progression. A wide variety of genetic alterations in the regulatory network of cell cycle, cell apoptosis, cell migration, and angiogenesis are suggested to be involved in the conversion of a normal cell to a malignant cell [11]. The bulk tumor commonly comprises cells with distinct molecular signatures across different regions (spatial heterogeneity), and temporal variations during the course of disease progression (temporal heterogeneity) [12]. In addition, tumors of the same histological type might have different pathogenesis due to differences in the germline background, the somatic mutation pattern, and environmental factors. The intertumoral and intratumoral heterogeneity contributes to different levels of drug sensitivity and resistance. Patients with high levels of intratumoral heterogeneity might suffer from a poor clinical outcome and therapeutic resistance, due to subclones predestined to evolve tumor driving mutations or drug resistance. Until now, no standard procedures have been used in clinical care to identify heterogeneity of tumor and then personalize treatment for patients.

Tumor microenvironment (TME) is comprised of cancer cells, infiltrating immune cells, stromal cells, and other cell types together with non-cellular tissue components, which determine disease progression and response to treatment (Fig. 9.1) [13]. The immune cell components in tumor microenvironments are complex and diverse, including T cells, macrophages, natural killer (NK) cells, granulocytes, and dendritic cells. Among tumor infiltrating immune cells, regulatory T cells (Tregs), tumor associated macrophages (TAMs), and myeloid-derived suppressor cells (MDSCs) are mainly responsible for immune suppression and inflammation, while NK cells, effector T cells, and dendritic cells are responsible for attacking tumor cells. The process of tumor development is shaped by interaction of tumor cells and immune cells in the TME, normally undergoing the following three stages:

1. Elimination. Tumor cells are recognized and eliminated by immune cells.
2. Equilibrium. The quantity and strength of tumor and immune system are in equilibrium.
3. Escape [14]. The tumor cells break through the immune barrier, start to proliferate without control, and spread to distant sites.

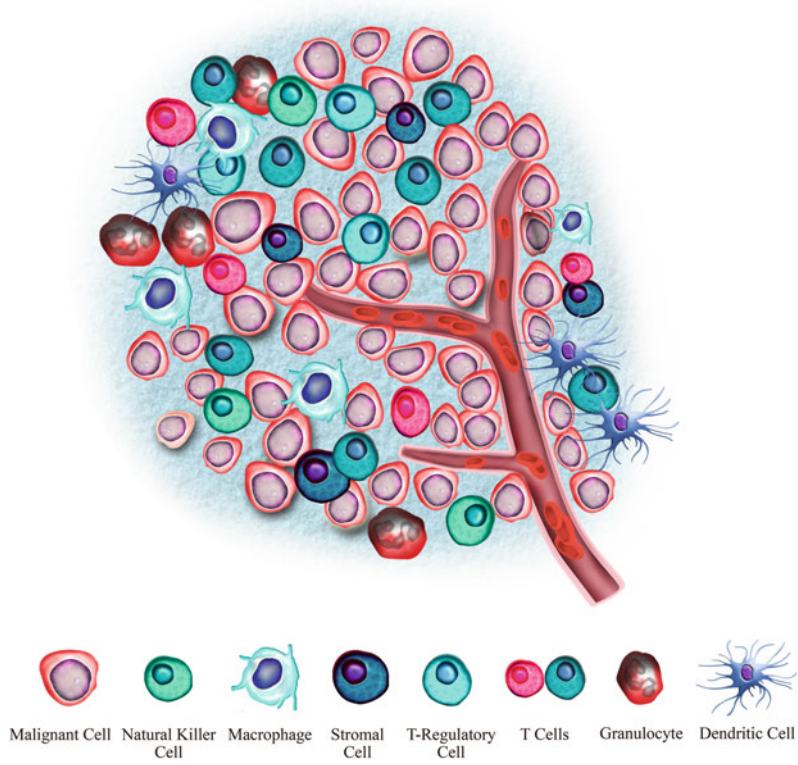


Fig. 9.1 Cancer cells interact with other cell types to form tumor microenvironment (TME). A typical TME has cancer cells, stromal cells, and infiltrating immune cells such as different types of T cells, natural killer (NK) cells, macrophages, granulocytes, and dendritic cells. The interactions between cancer cells and other cells in the TME play an important role in tumor progression and response to treatment

Briefly, the mechanisms of tumor immune escape can be summarized as (1) defective tumor antigen presentation; (2) immunosuppressive tumor microenvironment. The expression of major histocompatibility complex (MHC)-class I molecules on the surface of most tumor cells is decreased or absent, including complete loss of MHC molecules and loss of haplotype. As a result, T cells cannot be effectively activated by such tumor cells. In addition, immune inhibitory molecules can be upregulated in tumor cells. For example, tumor cells can directly contact T cells or natural killer cells and downregulate their killing capability by binding to cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) and programmed death-ligand 1 (PD-1) on the cell surface. Tumor cells can also secrete a series of immunosuppressive cytokines and chemokines into the tumor microenvironment. Besides, with the growth of tumor, new blood vessels must develop from pre-existing vascular network to satisfy tumors' demand for oxygen and nutrients, and hypoxia is a crucial driver of this event. VEGFA (Vascular Endothelial Growth Factor A) is secreted by hypoxic cancer cells and recruits angiogenesis-supporting TAMs into TME, which engages VEGFR2 expressed on the endothelial cells of neighboring blood vessels to initiate angiogenesis [15, 16].

Cancer cells and immune cells in the TME can be characterized by various cellular and molecular biology techniques such as FACS, immunohistology, PCR, and sequencing. High-throughput sequencing has gradually become a routine method applied in both cancer research and molecular diagnosis. Large-scale studies such as TCGA [17] have provided new insights into the inter-patient tumor heterogeneity at genomic, transcriptomic, and epigenetic levels. Molecular diagnostic tests based on mutation detection or gene expression profiles have been successfully commercialized for tumor subtype classification [18].

Although cellular and molecular mechanisms of tumor and TME heterogeneity and their interactions have been proposed to explain patient prognosis and drug responses, current clinical investigations are largely based on conventional bulk sequencing methods and cannot accurately detect genomic and genetic differences among different cell types, nor different tumor subclones, in the TME. Technologies with high resolution are required to decipher the complex tumor ecosystem and their dynamic changes. Recently developed high-throughput single cell sequencing methods, where genetic material of thousands of single cells can be sequenced in parallel, allow simultaneous characterization of cancer cells, immune cells, and stromal cells in the TME to shed light on intratumoral genomic diversity and intertumoral heterogeneity. In the following sections, we will describe recent technical advances of single cell sequencing, its clinical applications, as well as remaining challenges. We will also give our perspectives on the potential utilities of this revolutionizing technology as a new frontier in cancer diagnosis and treatment.

9.2 Single Cell Sequencing to Characterize Cancer

Advances in nucleic acid amplification and high-throughput sequencing techniques make it feasible to sequence genome or transcriptome at single cell level. The first mammalian single cell RNA sequencing procedure was developed in 2009 [19] and the first human cell genome was sequenced in 2011 [20]. Single cell sequencing is now gaining traction in cancer research as shown in Fig. 9.2.

A typical single cell sequencing experiment has the following steps: single cell isolation; whole genome amplification (WGA), or whole transcriptome amplification (WTA); library construction; sequencing; and data analysis. To ensure that genetic information from each single cell can be accurately captured, physical separation of each individual cell from a sample is required. This can be achieved by manual retrieval of cells under a microscope, or by using special instrumentations such as FACS sorters or laser capture microdissection (LCM) instruments. More recently, microfluidic systems of various designs have been developed to isolate single cells [21]. Following cell isolation, the next critical step is amplification of nucleic acids. A typical human cell contains about 6 pg of genomic DNA and 10–20 pg of RNA—too little to be sequenced directly on current sequencing platforms. There are diverse strategies to amplify DNA or RNA from a single cell. For single cell WGA, the mostly commonly used methods are based either on PCR or multiple displacement amplification (MDA) [22]. Single cell WTA normally starts with a cDNA synthesis step, followed by PCR or in vitro transcription (IvT) for cDNA amplification [23]. Optimized enzymatic reactions developed in recent years enable efficient amplification of genomic DNA or cDNA from single cells, with some WGA methods achieving close to 100% genome coverage of single cells [24]. Using microfluidic systems to conduct single cell WGA or WTA in miniature volumes can further improve reaction efficiency by increasing local concentration of substrates [25]. Sequencing libraries can be constructed from amplified single cell gDNA or cDNA and sequenced with standard protocols. To analyze single cell sequencing data, artifacts specifically related to single cell amplification, such as allele drop out (ADO), amplification bias, and batch effect, should be taken into consideration [26].

The first generation of single cell sequencing methods is quite low throughput regarding number of cells analyzed simultaneously. This is largely due to the limitation by low-throughput cell isolation methods at the time, as well as relatively high sequencing costs. The main application of low-throughput single cell sequencing methods in cancer research is to detect mutations in individually selected cancer cells or circulating tumor cells (CTCs) [27, 28]. Low-throughput single cell sequencing technologies are valuable tools to uncover the mechanisms of cancer genesis and were also able to shed light on tumor heterogeneity and evolution. However, sequencing information from a dozen of single cells lacks sufficient granularity and statistical power to

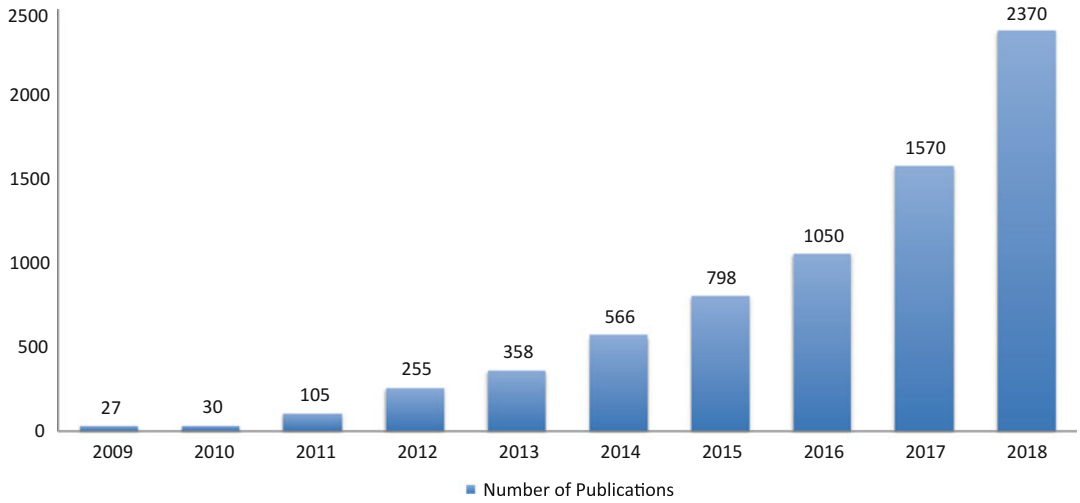


Fig. 9.2 Number of publications on using single cell sequencing for cancer research, from year 2009 to 2018. Data retrieved from Google Scholar with search keywords "single cell sequencing" AND "cancer"

interrogate biological systems as complex and heterogeneous as cancer.

Several independently developed high-throughput single cell sequencing methods, all based on microfluidic systems and cell-barcoding techniques, were published in 2015. These innovative technologies finally made it technically and financially feasible to simultaneously sequence thousands or more single cells from the same sample [29–31], opening up a new era in cancer research. The principle of high-throughput single cell sequencing is relatively simple. First, single cells are partitioned into either droplets or micro-wells together with barcoding beads that carry oligos with unique barcode sequences. In those micro-compartments, each single is lysed and labeled with a cell barcode. For single cell RNA sequencing, cell-barcoding beads normally have oligo-dT following a stretch of cell-barcoding sequence, thereby labeling all cDNA strands from the same cell with a unique cell barcode sequence during cDNA synthesis. cDNA from all single cells can then be pooled together, amplified, and sequenced. Sequence information from each single cell can be obtained based on their unique cell barcode sequences. More recently, methods have also been developed to forgo microfluidics-based partition of single cells. Instead, cells are first made permeable by fixation, so that barcoding oligos can enter cells to label DNA or RNA. Sequential splitting, barcoding, and pooling of cells is used to give genetic material from each single cell a unique barcode after undergoing several rounds of such split-and-pool procedures [32, 33].

By barcoding genetic material individually for thousands of single cells and pooling them together in subsequent amplification and library construction steps, high-throughput single cell sequencing methods significantly reduce the time and cost to produce single cell libraries at per cell basis, making it possible to decipher complex systems such as cancer at high resolution and efficiency. More importantly, the ability to simultaneously process, sequence, and analyze thousands of single cells from the same sample effectively reduces the impact of technical and biological variability intrinsic to single cell analysis, making the overall results more reliable and accurate. High-throughput single cell sequencing methods have been rapidly adopted in almost all research fields in life sciences, as shown by drastic increase in number of cells in each published study where single cell sequencing is used (Fig. 9.3). They are currently acting as essential workhorses for the human cell atlas (HCA)

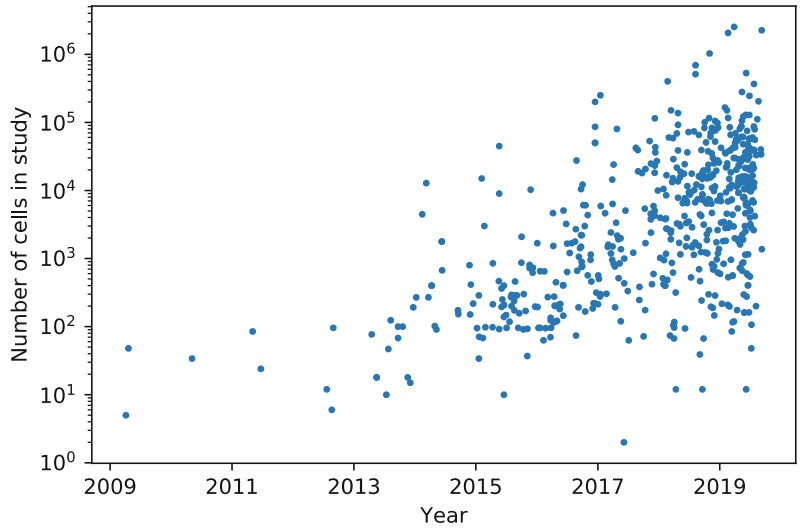


Fig. 9.3 High-throughput single cell sequencing is applied in increased number of studies. A summary of total single cell numbers per study in published single cell sequencing projects, 2009 to the first half year of 2019. Figure adopted from "A curated database reveals trends in single cell transcriptomics," Svensson V. et al., doi: https://doi.org/10.1101/742304

project (https://www.humancellatlas.org) to make a comprehensive reference map of all human cells.

9.2.1 Single Cell Sequencing Data Analysis and Interpretation

Dedicated bioinformatics algorithms and analysis pipelines are required to address the technical and biological variations in single cell sequencing data. High-throughput single cell sequencing generates large datasets with unprecedented complexity and granularity. Effective data analysis and interpretation tools are required to extract useful information and insights from such datasets. Here, we will use high-throughput single cell RNA (scRNA-seq) sequencing as an example to illustrate commonly used primary and secondary analysis algorithms to generate insights from the single cell gene expression data (Fig. 9.4).

Primary analysis steps process raw data to gene expression matrix. Following filtering and trimming of low quality reads, the remaining reads are mapped to a reference genome and annotated. A gene expression matrix containing gene expression information for each single cell is then generated. For second analysis, quality control is usually the first step to remove cells with low viability, such as low number of expressed genes and high mitochondrial contents. Due to the high level of dropouts and technical noise, high-throughput scRNA-seq data are zero-inflated, and thus normalization must be performed to remove cell-specific biases. Typically, normalization methods estimate a scale factor per cell to adjust the total number of molecules detected. Next, unsupervised clustering is required to partition cells into groups, followed by differential expression analysis to identify marker genes in each cell group. To visualize the whole transcriptome expression patterns of thousands of cells, popular algorithms such as t-distributed stochastic neighbor embedding (tSNE) and uniform manifold approximation and projection (UMAP) can be implemented to project the high-dimensional single cell data onto a two-dimensional space [34, 35].

Besides cell subpopulation identification, advanced analysis algorithms have been developed to extract more comprehensive information from high-throughput scRNA-seq data. For example, pseudotime trajectory inference method utilizes the unsynchronized individual cells at an instantaneous time point to reconstruct dynamic trajectories of cell differentiation or progression [36]. This approach could be applied to dissect the tumor progression dynamics and also shed

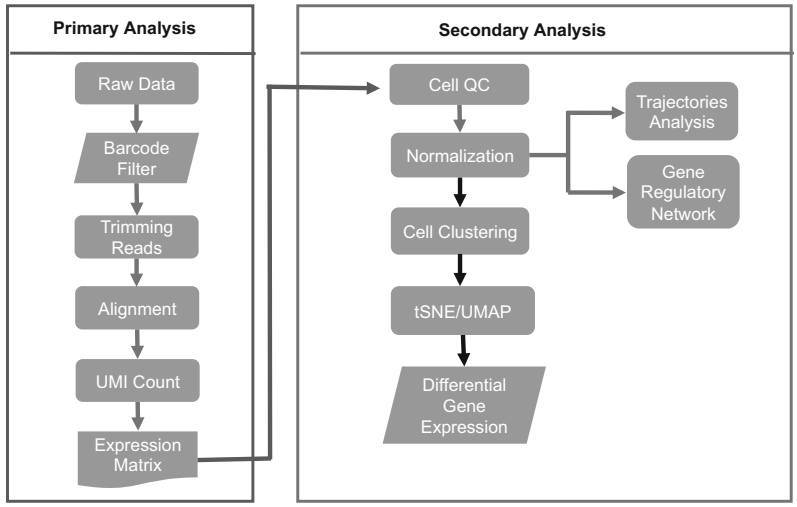


Fig. 9.4 A typical scRNA-seq data analysis pipeline. Following sequencing, alignment and gene annotation are performed to generate a gene expression matrix (Cell ∗ Gene). To reduce bias in qualitative data analysis, cell filtering and normalization are performed with different statistical approaches. Subsequent analysis is performed to explore the biology insight from dataset, including cell clustering to identify subtypes, cell trajectories to describe complex differentiation processes, and detecting differentially expressed genes and gene regulatory network to find biomarkers. tSNE: t-distributed stochastic neighbor embedding; UMAP: Uniform Manifold Approximation and Projection, both non-linear dimensionality reduction techniques for embedding high-dimensional single cell sequencing data for visualization in a low-dimensional space

light on mechanisms determining the fate of tumor infiltrating immune cells.

Another example is construction of gene regulatory networks (GRNs). This is inherently powerful in combination with scRNA-seq, since gene expression patterns of thousands of cells are captured under the same condition. Several methods have been developed to identify GRN from single cell data, and these methods have been successfully applied to T cell biology, providing new insights from co-expression data analysis [37, 38]. Elucidation of GRNs within the complex tumor environment of different cell types could facilitate the identification of potential drug targets to reverse the abnormal activities of tumor cell proliferation and immune system suppression.

9.3 Single Cell Sequencing in Cancer Research

Rapidly developing high-throughput single cell sequencing technologies are revolutionizing the field of cancer research. The complex genomic, transcriptomic, and epigenetic patterns of individual cells in tumor and in the TME can be effectively deciphered by such methods, while advanced bioinformatics algorithms can be used to infer cellular dynamics and intercellular interactions [39]. This makes high-throughput single cell sequencing a powerful tool to study molecular mechanisms of cancer initiation and evolution, as well as immune evasion; to discover biomarkers that can be used for prognosis and prediction of treatment responses; and to identify potential therapeutic targets or combinational treatment strategies [40].

9.3.1 Cellular and Molecular Mechanisms of Cancer

Cancer originates from normal cells that have acquired driver mutations through their life cycle and also have evaded immune surveillance. The origin and evolution of cancer are tightly shaped by the interaction of cancer cells and

other non-cancerous cells in the TME. Single cell sequencing technology is able to generate comprehensive genetic information on all cell types within a tumor. By analyzing gene expression patterns of different tumor subpopulations and comparing them to those of the normal tissue, either from the same patient or from a model system, the origin of tumor can be accurately determined. Further analysis of differentially expressed genes and pathways in cancer cells can also help to identify genes and pathways crucial for cancer development and discover potential drug targets for therapy. Advanced bioinformatics analysis such as pseudotime analysis can further determine the evolution trajectory of tumor by positioning subpopulations of a tumor on a time axis based on their gene expression patterns, further illustrating potential therapeutic targets. Analysis of gene expression patterns in different cancer cell subclones can also shed light on the mechanisms for drug resistance. Furthermore, single cell sequencing can also be used to detect different stromal and immune cell populations in the TME, while specific bioinformatics tools can be used to infer their interactions. For example, a recent scRNA-seq study by Lambrechts et al. [41] on stromal cells in the lung cancer microenvironment was able to detect 52 different cell types and analyze their potential roles in cancer development. Such information can help to understand how TME shapes the initiation and growth of cancer cells, as well as how cancer cells might influence other cells in the TME.

9.3.2 Biomarker Discovery

Cancer biomarkers can serve as a useful tool for diagnosis, prognosis, and prediction of drug responses and cancer recurrence. Traditional cancer biomarkers are biomolecules with distinct features, such as cell free DNA or FFPE DNA harboring EGFR mutations commonly used in companion diagnosis for lung cancer treatment. With the application of single cell sequencing, cells with defined gene expression patterns could become a new class of biomarkers that can predict cancer prognosis [42–44] or responses to treatment [42, 45]. Several recent studies have identified tumor-specific cell populations as biomarkers that can be used for prognosis or prediction of drug responses. For example, a tissue-resident memory T cell subset has been shown to correlate with breast cancer prognosis [44], while a NK-dendritic cell composition aids in determining response of melanomas to PD-1 inhibitor treatment [42]. We expect more and more such biomarkers in the form of specific cell population to be discovered with the wider application of single cell sequencing in cancer research. And it is only logical that for a disease as complex as cancer, the biomarkers that can be used to accurately predict disease progression and treatment responses cannot be limited to the genomic DNA mutation information derived from averaged bulk sequencing.

9.3.3 Precision Medicine

Similar to the bulk-sequencing-based biomarkers such as DNA mutations and gene expression signatures, cell-specific DNA mutations and expression signatures could be employed in patient stratification for different treatment options. However, unlike bulk sequencing which only gives information on the presence or absence of certain mutations, high-throughput single cell sequencing approach can simultaneously obtain information at both molecular and cellular level. It can not only detect different cellular groups in a tumor sample, but also obtain information on the characteristic gene expression patterns of each cell type and infer cellular interactions, making it possible to discover potential therapeutic targets that can be used to select treatment options more precisely. For immunotherapy, levels of immune cell infiltration could be an indicator of checkpoint inhibitor efficacy. Relative levels of exhausted T cells and CD8+ effector T cells in tumor tissue would help evaluate if suppressed T cells could be re-activated and effectively attack cancer by checkpoint inhibitor.

Single cell sequencing could also reveal novel targets for efficient drug design. Gene expression

and pathway activation patterns for multiple cancer cell subpopulations from the same patient could be simultaneously detected through high-throughput single cell measurements, making it possible to select combinatorial therapy targeting different cancer subclones within the same tumor. Due to the highly heterogeneous nature of tumor, such information-based combinatorial therapy is theoretically more effective and sustainable than therapies that target only one driver mutation. There has been evidence that such approach could also be used to combine immunotherapy with existing targeted therapy. Jerby-Arnon et al. used scRNA-seq to identify characteristic activation of CDK4/6 signaling pathway in tumors from melanoma patients resistant to anti-PD1 immunotherapy [45]. They also demonstrated that tumors not responding to immunotherapy alone could be controlled using a combination of CDK4/6 inhibitor and immunotherapy in a mouse model [45].

High-throughput single cell sequencing gives a comprehensive picture of cancer cells, infiltrating immune cells, and their interactions. This would facilitate drug discovery and screening process. For example, the molecular signatures and functional roles of TAMs and MDSCs and their subtypes are still not completely clear till this date, which makes it difficult to design drugs targeting TAMs. Single cell sequencing can characterize gene expression patterns in different TAM or MDSC subgroups to understand their functions and discover potential new therapeutic targets.

Taken together, single cell sequencing can be used to understand precise molecular features of cancer, to provide high-dimensional profiles of both cancer and immune cells, and to identify qualified therapeutic targets. Multi-dimensional cancer information delivered by single cell sequencing can be especially useful in developing and optimizing novel cancer therapies such as immunotherapy, cell therapy, and gene editing.

9.4 Outlook

To apply single cell sequencing in clinical settings, robust standard operation procedures (SOPs) need to be established, verified, and validated for each step along the workflow. This include sample acquisition and processing, single cell isolation, amplification, sequencing library construction, sequencing, data analysis pipelines, data annotation, and data interpretation.

Clinical samples come in a variety of forms and status, not always optimal for single cell sequencing, which normally requires single cell suspension with high cellular viability. For clinical applications involving longitudinal studies such as drug response and minimal residue disease monitoring, the ability to process archived samples would be critical. Methods have been developed to dissociate single cells from cryopreserved and chemically fixed samples that can be used for single cell RNA sequencing [46, 47]. For example, single nucleus sequencing (snRNA-seq) methods have been used to profile the gene expression of single nuclei in archived or otherwise challenging samples [48, 49]. Although comparative studies have been carried out to confirm the consistencies between transcriptome of single cells and single nuclei for tissues such as breast, kidney, and brain [50–52], it needs to be studied whether this conclusion can be extended to other tissue and cell types, as well as tissues preserved with different protocols and for different durations.

Typical clinical trials involve multicenter study design and thus require minimal batch effects while conducted by different operators at different sites. Current experimental procedures for single cell sequencing are still relatively long and tedious. Streamlined and automated protocols are needed to facilitate broader adoption of single cell sequencing in clinical settings. Furthermore, different single cell sequencing protocols have system-specific bias in their gene expression quantifications [53]. On the data analysis side, different technology platforms also have their own accompanying pipelines. Biases introduced by both experiment and analysis steps make meta-analysis across different studies challenging. Computational batch effect correction methods [54, 55] showed promises in removing technical biases among different experimental platforms. However, prior to their integration

into large-scale meta-analysis pipeline, a significant amount of validation studies is still required to assess if real biological variations between samples are retained while technical variations are removed.

Besides standardization of sample handling and library preparation protocols, the standardization of the computational analysis pipelines is also critical in a controlled clinical trial setting. The primary analysis turning raw sequencing data to the gene expression matrices is one source of variation in the data analysis pipelines. Some third-party analysis tools aimed to solve this problem by unifying the primary analysis steps for various platforms [56]. Secondary analysis pipeline of scRNA-seq usually involves cell quality control, noise reduction, and expression profile normalization, followed by clustering and then cluster-level cell type identification through manual annotation. To determine the optimal parameters for single cell sequencing data QC, comprehensive benchmark efforts are needed to evaluate the effects of different tools and parameter settings. For scRNA-seq clustering methods, benchmark studies showed that different analysis tools seem to have low consistencies [55, 56]. In addition, their performances are dependent on multiple factors, such as complexity of the data, number of cells, and parameter settings of the tool. On the other hand, cell types and subtypes are hierarchically structured and sometimes don't have a discrete separation between them, which makes the accuracy of cell clusters difficult to define. Cell ontology, an effort to define cell types and their hierarchical structures based on scRNA-seq data, could be extremely useful to connect cell types to their molecular characteristics and functions [57]. Large international consortiums, such as Human Cell Atlas (HCA) and the Life-Time Initiative, are joining forces to comprehensively profile single cells from different organs, in health and diseases.

As the number of cells profiled in single cell studies continues to grow, the clustering and annotation approach would face a hurdle of large computational resource requirements and tedious manual workload. As an alternative, supervised methods use existing reference datasets and their validated labels to predict the label of new datasets. Once the models have been trained, the prediction step generates the label for each cell independently and thus could be parallelized infinitely. The performance of current supervised methods designed for scRNA-seq showed promise for common sample types such as PBMC and human pancreas [58]. To routinely adopt such methods in scRNA-seq analysis, cell atlas type of studies and validated cell type labels have to be generated to construct a comprehensive reference dataset. However, more validation work is still required to apply these methods to cancer samples, due to the highly heterogeneous nature of cancer cells [59, 60].

To facilitate application of single cell sequencing in cancer diagnosis and drug development, a standard pipeline based on comprehensive single cell knowledgebase is required for data interpretation. In terms of cell types in the cancer microenvironment, tremendous efforts have been made to study their normal and pathological functions in cancer through low-dimensional single cell studies such as flow cytometry and CYTOF [61]. The functions of many important genes such as cytokines and transcription factors have also been studied extensively. To translate the existing knowledge of immunology and immune-oncology into the interpretation of single cell sequencing data of patients, a comprehensive knowledgebase, similar to the genome variation databases such as COSMIC, Clinvar, and HGMD, is required to link scRNA-seq data to the clinical phenotypes of patients. Single cell sequencing data with information on both cellular and molecular levels can potentially act as biomarkers for cancer diagnosis and needs to be included in such database. In addition, the database should also incorporate information on disease-associated genes or signaling pathways, drugs targeting such genes or pathways, and ongoing clinical trials to maximize the clinical insights that can be derived from single cell sequencing datasets.

Emerging new technologies could also greatly speed up the knowledge discovery and in turn the application of single cell sequencing in clinics.

For example, the location of infiltrating immune cells within a tumor could be important in the immunotherapy response [62]. Furthermore, directly linking genomic and phenotypic information within single cells could be extremely useful for understanding the mechanisms of disease as well as drug responses. Therefore, spatially resolved single cell sequencing and multi-omics single cell sequencing are both exciting new technologies that can bring new insights to cancer research and clinical applications. High-throughput in-situ single cell technologies could be a convenient new way to measure single-cell level phenotypes without the need for extracting nucleic acids and proteins from the cells, while retaining important spatial information [63].

In summary, single cell sequencing has demonstrated tremendous potential to impact not only basic cancer research, but also the way cancer is diagnosed and treated in the upcoming decade. However, extensive work has to be done to optimize and standardize the workflow in a controlled clinical environment, to build up comprehensive single cell database and knowledgebase, and to carry out large-scale clinical studies through joint efforts of different research groups. Technology developers, basic and translational researchers, clinical practitioners, regulatory agencies, and commercial companies developing products and assays will have to work together to bring this exciting new technology to its full potential for next-generation precision medicine and more effective drug development.

Acknowledgement *Disclosure*: Nan Fang is Founder and Chief Executive Officer of Singleron Biotechnologies, a company that develops single-cell sequencing products. Jue Fan and Jingwen Fang are employees at Singleron Biotechnologies.

References

1. Bray F et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424
2. Pabani A, Butts CA (2018) Current landscape of immunotherapy for the treatment of metastatic non-small-cell lung cancer. Curr Oncol 25:S94–S102
3. Kobayashi S et al (2005) EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 352:786–792
4. Balak MN et al (2006) Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. Clin Cancer Res 12:6494–6501
5. Yamaguchi F et al (2014) Acquired resistance L747S mutation in an epidermal growth factor receptor-tyrosine kinase inhibitor-naïve patient: a report of three cases. Oncol Lett 7:357–360
6. Thress KS et al (2015) Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med 2018 (21):560–562
7. Wu S-G, Shih J-Y (2018) Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. Mol Cancer 17:38–14
8. Martinez P, Peters S, Stammers T, Soria J-C (2019) Immunotherapy for the first-line treatment of patients with metastatic non-small cell lung cancer. Clin Cancer Res 25:2691–2698
9. Singh SS, Dahal A, Shrestha L, Jois SD (2019) Genotype driven therapy for non-small cell lung cancer: resistance, pan inhibitors and immunotherapy. Curr Med Chem 26:1–39
10. Roviello G, Corona SP, Nesi G, Mini E (2019) Results from a meta-analysis of immune checkpoint inhibitors in first-line renal cancer patients: does PD-L1 matter? Ther Adv Med Oncol 11:1758835919861905
11. Kandoth C et al (2013) Mutational landscape and significance across 12 major cancer types. Nature 502:333–339
12. Dagogo-Jack I, Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 15:81–94
13. Spill F, Reynolds DS, Kamm RD, Zaman MH (2016) Impact of the physical microenvironment on tumor progression and metastasis. Curr Opin Biotechnol 40:41–48
14. Beatty GL, Gladney WL (2015) Immune escape mechanisms as a guide for cancer immunotherapy. Clin Cancer Res 21:687–692
15. Hirata E, Sahai E (2017) Tumor microenvironment and differential responses to therapy. Cold Spring Harb Perspect Med 7:a026781
16. Wu T, Dai Y (2017) Tumor microenvironment and therapeutic response. Cancer Lett 387:61–68
17. Cancer Genome Atlas Research Network et al (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45:1113–1120
18. Bedard PL, Hansen AR, Ratain MJ, Siu LL (2013) Tumour heterogeneity in the clinic. Nature 501:355–364

19. Tang F et al (2009) mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382
20. Navin N et al (2011) Tumour evolution inferred by single-cell sequencing. Nature 472:90–94
21. Shinde P et al (2018) Current trends of microfluidic single-cell technologies. Int J Mol Sci 19:3143
22. Borgström E, Paterlini M, Mold JE, Frisen J, Lundeberg J (2017) Comparison of whole genome amplification techniques for human single cell exome sequencing. PLoS One 12:e0171566
23. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58:610–620
24. Zhang C-Z et al (2015) Calibrating genomic and allelic coverage bias in single-cell sequencing. Nat Commun 6:6822–6810
25. Reece A et al (2016) Microfluidic techniques for high throughput single cell analysis. Curr Opin Biotechnol 40:90–96
26. Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17:175–188
27. De Luca F et al (2016) Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. Oncotarget 7:26107–26119
28. Zhang C-Z et al (2015) Chromothripsis from DNA damage in micronuclei. Nature 522:179–184
29. Macosko EZ et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214
30. Klein AM et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161:1187–1201
31. Bose S et al (2015) Scalable microfluidics for single-cell RNA printing and sequencing. Genome Biol 16:120–116
32. Rosenberg AB et al (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360:176–182
33. Cao J et al (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357:661–667
34. Maaten LVD, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
35. McInnes L, Healy J, Melville J (2018) UMAP: Uniform manifold approximation and projection for dimension reduction
36. Bendall SC et al (2014) Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157:714–725
37. Fiers MWEJ et al (2018) Mapping gene regulatory networks from single-cell omics data. Brief Funct Genomics 17:246–254
38. Goodell MA, Bodine DM (2019) Single-cell technology meets hematology: introduction to a review series. Blood 133:1387–1388
39. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA (2017) Single-cell transcriptomics to explore the immune system in health and disease. Science 358:58–63
40. Shalek AK, Benson M (2017) Single-cell analyses to tailor treatments. Sci Transl Med 9:eaan4730
41. Lambrechts D et al (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med 24:1277–1289
42. Barry KC et al (2018) A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. Nat Med 24:1178–1191
43. Zheng H et al (2018) Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. Hepatology 68:127–140
44. Savas P et al (2018) Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med 2018 (24):986–993
45. Jerby-Arnon L et al (2018) A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell 175:984–997.e24
46. Alles J et al (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol 15:44–14
47. Guillaumet-Adkins A et al (2017) Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol 18:45–15
48. Habib N et al (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods 14:955–958
49. Lake BB et al (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36:70–80
50. Kim C et al (2018) Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. Cell 173:879–893.e13
51. Lake BB et al (2019) A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat Commun 10:2832–2815
52. Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36:421–427
53. Ziegenhain C et al (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65:631–643.e4
54. Smith T, Heger A, Sudbery I (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res 27:491–499
55. Freytag S, Tian L, Lonnstedt I, Ng M, Bahlo M (2018) Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. F1000Res 7:1297
56. Duò A, Robinson MD, Soneson C (2018) A systematic performance evaluation of clustering methods for single-cell RNA-seq data. F1000Res 7:1141
57. Meehan TF et al (2011) Logical development of the cell ontology. BMC Bioinformatics 12:6–12

58. Abdelaal T et al (2019) A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol 20:194–119
59. Tirosh I et al (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352:189–196
60. Patel AP et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344:1396–1401
61. Kourelis TV et al (2019) Mass cytometry dissects T cell heterogeneity in the immune tumor microenvironment of common dysproteinemias at diagnosis and after first line therapies. Blood Cancer J 9:72–13
62. Mansfield AS et al (2016) Temporal and spatial discordance of programmed cell death-ligand 1 expression and lymphocyte tumor infiltration between paired primary lesions and brain metastases in lung cancer. Ann Oncol 27:1953–1958
63. Crosetto N, Bienko M, van Oudenaarden A (2014) Spatially resolved transcriptomics and beyond. Nat Rev Genet 16:57–66

Fengying Wu is Associate Professor at Department of Medical Oncology, Shanghai Pulmonary Hospital, Tongji University, Shanghai, China.

Jue Fan is Vice President of Bioinformatics at Singleron Biotechnologies, Nanjing, China.

Jingwen Fang is a bioinformatician at Singleron Biotechnologies, Nanjing, China.

Priya S. Dalvi is a postdoctoral researcher at the Centre for Molecular Medicine Cologne, University of Cologne (UoC), Cologne, Germany.

Margarete Odenthal is Professor at the Institute for Pathology and Head of the Laboratory of Translational Molecular Pathology, University of Cologne (UoC). She is project leader at the Centre for Molecular Medicine Cologne, University of Cologne (UoC), Cologne, Germany.

Nan Fang is Founder and CEO at Singleron Biotechnologies, Nanjing, China; and Adjunct Professor at Anhui Medical University, Anhui, China.

The Role of Methylation in the CpG Island of the ARHI Promoter Region in Cancers

10

Xiaozhuan Liu, Tingting Zhang, Yanjun Li, Yuwei Zhang, Hui Zhang, Xiangdong Wang, and Li Li

Abstract

Hypermethylation can downregulate many tumor suppressor gene expressions. Aplasia Ras homologue member I (ARHI, DIRAS3) is one of the maternally imprinted tumor suppressors in the RAS superfamily. This chapter overviewed the importance of ARHI methylation and expression phenomes in various types of cancers, although the exact mechanisms remain unclear. As an imprinted gene, aberrant DNA methylation of the paternal allele of ARHI was identified as a primary inhibitor of ARHI expression. The role of methylation in the CpG islands of the ARHI promoter region vary among ovarian cancers, breast cancers, hepatocellular carcinoma, colon cancers, pancreatic cancer osteosarcoma, glial tumors, follicular thyroid carcinoma, or lung cancers. The methylation of ARHI provides a new insight to understand molecular mechanisms of tumorigenesis and progression of cancers.

Keywords

ARHI · Methylation · Cancer · Disease · Inhibitor

10.1 Introduction

Aplasia Ras homologue member I (ARHI, DIRAS3) is the first tumor suppressor gene identified in the Ras superfamily [1] and allocated in chromosome 1p31 where there is loss of heterozygosity. ARHI has a distinctive N terminal extension for the suppression of tumor growth and is one of 40 genes to be imprinted in the human genome. ARHI is expressed in cells from the paternal allele during the process of embryonic development [1]. The protein coding region is located within exon 2 and encodes a 229-residue small GTP binding protein belonging

Xiaozhuan Liu, Tingting Zhang, and Yanjun Li contributed equally to this work.

X. Liu · T. Zhang · Y. Li · Y. Zhang · H. Zhang
Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

X. Wang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China

L. Li (✉)
Department of Scientific Research and Discipline Construction, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China
e-mail: lili@henu.edu.cn

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_10

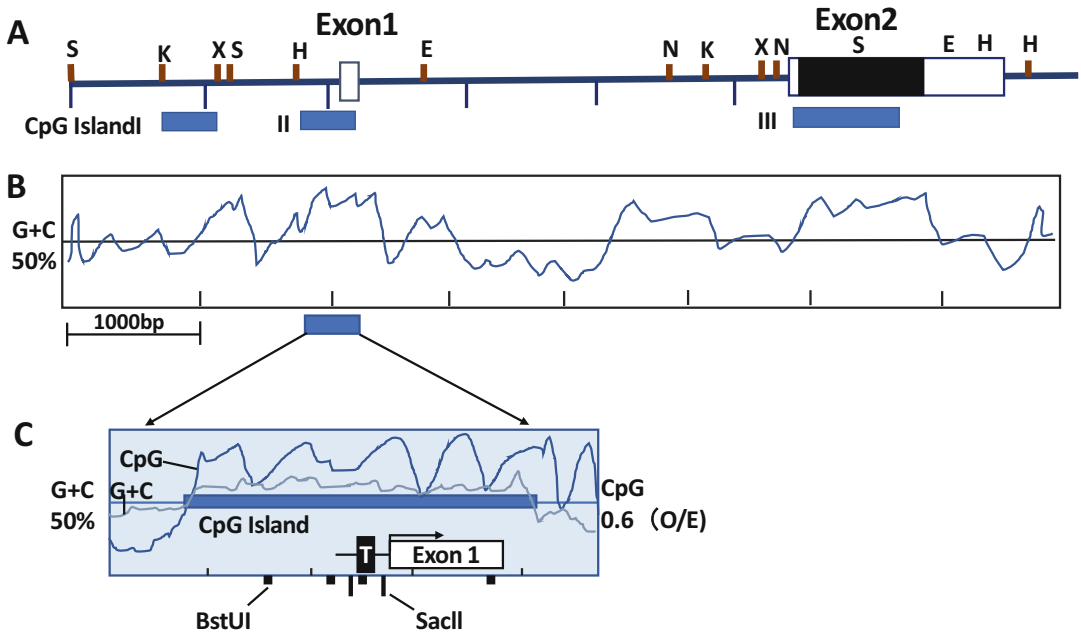


Fig. 10.1 Structure of the human ARHI gene. (**a**) Structural organization of the ARHI gene and schematic drawing of the ARHI cDNA. The ARHI gene contains two exons interrupted by large intron. The blocked and opened boxes represent the coding and noncoding regions. Shadowed boxes are CpG island regions. Restriction enzyme sites are designated as: *S* SmaI, *K* KpnI, *X* XbaI, *H* HindIII, *E* EcorI, *N* NcoI. (**b**) The GC content per 100 bp across the entire ARHI locus. (**c**) The G + C content per 100 bp and CpG density per 100 bp for the CpG island II spanning the region upstream of and encompassing the entire exon1

to the Ras superfamily [2]. Three potential CpG islands about 300 base pairs were found in the promoter and exons of the ARHI gene (Fig. 10.1). CpG island I, II, and III are located about 1 kb upstream of the transcription initiation site, and in the region of exon 2, respectively (Fig. 10.1b). Of those, CpG island II spans the 5′-up-stream region of ARHI, including the transcription initiation site and a portion of exon 1 (Fig. 10.1b) [3].

The imprinted gene ARHI undergoes the dysfunction with a "single hit" during carcinogenesis by inactivating single functional allele [4]. ARHI silencing in cancers can be caused by multiple mechanisms, including LOH, DNA methylated, histone deacetylation, histone methylation, and transcriptional regulation. The acetylation and methylation of chromatin lead to the downregulation of ARHI expression and ability to suppress tumor growth [4]. The histone deacetylation and H3 lysine 9 methylation contribute to the silence of ARHI by DNA methylation-dependent pathway (Fig. 10.2) and the binding of transcriptional repressors to recruit relevant enzymes onto chromatin (Fig. 10.3). Human oncogenesis may be due to the change of DNA methylation. About 50% of human genes have clusters of CpG islands in the 5′-regulatory sequences, of which the most are not methylated. In human cancers, the aberrant methylation includes hypomethylation, hypermethylation, and increased DNA methyltransferase activity [5, 6].

Aberrant methylation of CpG islands acts as a distinct molecular mechanism, leading to malignant transformation and providing the epigenetic equivalent of mutation/deletion during oncogenesis [7, 8]. Such DNA methylation is also recognized as potential driver of carcinogenesis [9]. CpG methylation lead to gene transcription declining in the promoter region in ARHI genes [10]. The downregulation of ARHI is found in many types of cancer, including ovarian cancer, hepatocellular carcinoma, and others [11]. This

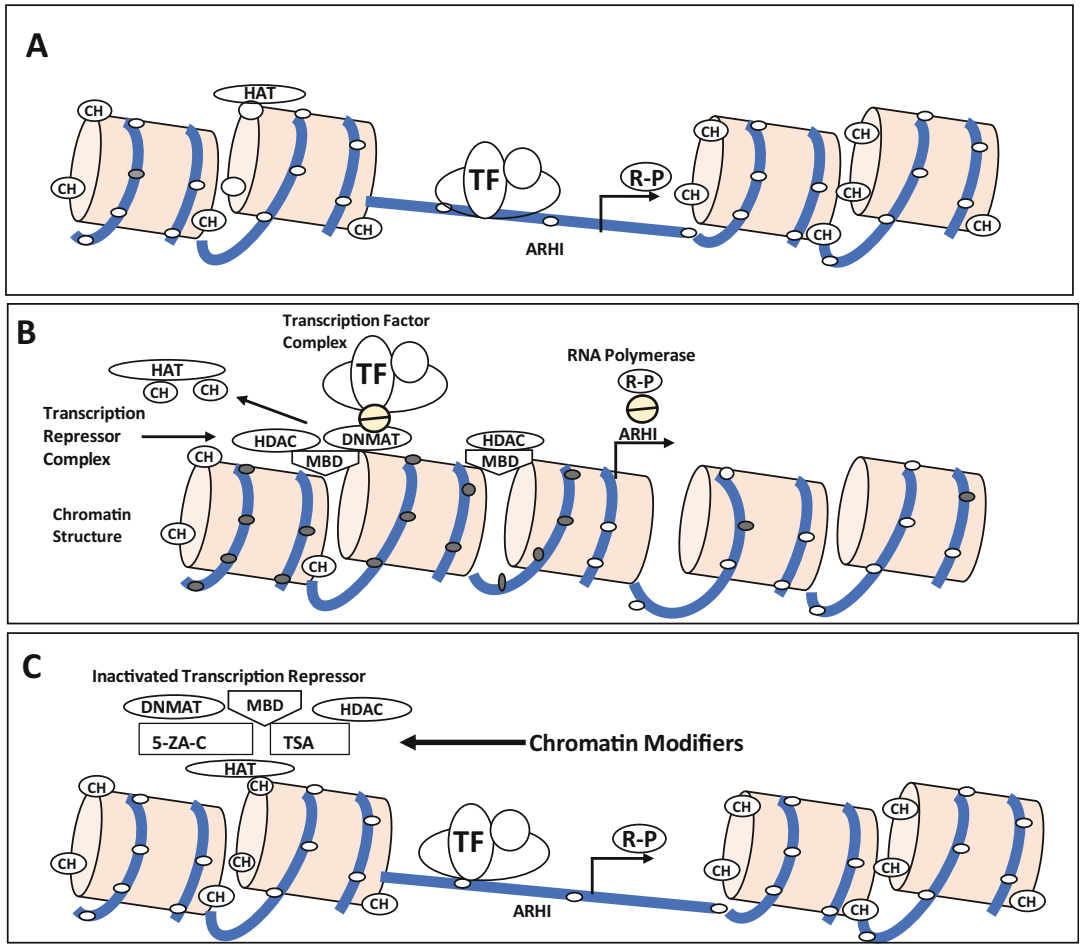


Fig. 10.2 Methylation-dependent model for the silencing of the ARHI gene in cancer cell. (**a**) A transcriptionally active CpG island promoter is depicted with positioned nucleosomes, consisting of acetylated (CH) histone and unmethylated CpG residue (white circle). Histone acetyltransferase (HAT) creates an accessible chromatin configuration that facilitates transcriptional activity. (**b**) Silenced ARHI gene. Transcriptional repressor complex including methyl-CpG binding domain (MBD) protein, DNA methyltransferase (DNMT), histone deacetylase (HDAC), and other repressors binds to methylated CpG (gray circles) and inactivates the ARHI gene. (**c**) Chemical inhibitors such as 5-AZA and TSA can inhibit DNMT and reactivate the ARHI gene

chapter aims at overviewing the correlation between ARHI CpG methylation and the tumor in the development of cancer.

10.2 ARHI and Ovarian Cancer

Of malignancies, the highest expression ARHI is expressed in ovarian tissues [1]. The ARHI expression was downregulated in ovarian tumor tissues, as compared with the normal ovarian tissues [12, 13]. The ARHI expression was reduced in ovarian serous papillary carcinomas [14] and ARHI protein consistently expressed in epithelial cells of ovarian surface [4]. The levels of ARHI expression were correlated with the malignancy of tumors [14], of which ARHI was reduced in 88% of ovarian cancer tissues. The overexpression of ARHI can inhibit the proliferation of ovarian tumor cells and induced autophagy and tumor dormancy and other phenomena [15].

The ARHI expression is regulated by CpG island methylation in the ARHI promoter region

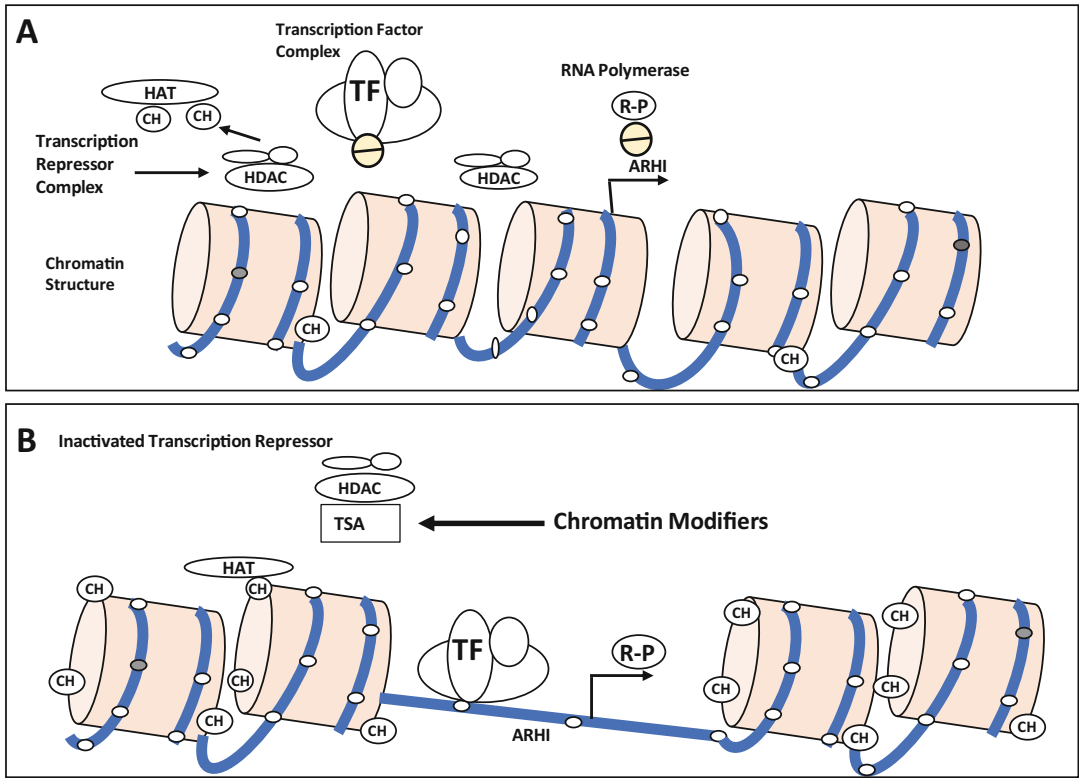


Fig. 10.3 Methylation-independ model for silencing the ARHI gene in cancer. (**a**) Silence ARHI gene. Transcriptional repressor complex, including histone deacetylase (HDAC) and other transcription repressors, inhibits HAT and inactivates the ARHI gene. (**b**) TSA can inhibit HDAC and reactive the ARHI gene

and other way. ARHI CpG islands I and II were hypermethylated in 31% and 12% of ovarian cancers, respectively, associated with reduced ARHI expression [16]. ARHI expression reduced in ovarian cancer epithelial and modified cancer cells (SKOV-3 and HO-8910), where CpG islands I and II were partially methylated or hypermethylated, enhancing the proliferation of tumor cells. Such proliferation was reversed by the administration of 5-aza-2′-deoxycytidine [17].

10.3 ARHI and Breast Cancer

ARHI expression is lost or downregulated in most breast cancers, while the ARHI overexpression inhibits the growth of tumor cells and induces the apoptosis of tumor cells [18]. Transcriptional repression of ARHI is closely related to breast cancer progression [19]. The expressions of ARHI were detected in normal breast epithelia, downregulated in 41% of ductal carcinoma in situ (DCIS) and 70% of invasive carcinomas [20]. Compared with DCIS in the same sample, ARHI was further downregulated in 26% of invasive carcinomas. About 17% of invasive carcinoma lost ARHI protein expression. Other investigators reported that ARHI mRNA expression decreased in 46–48% of human breast cancer specimens [20, 21], correlated with lymph node metastases [21] and involved with the progression of breast tumor.

ARHI expression can be downregulated by various mechanisms. For example, hypermethylation of both alleles in the CpG island II of the ARHI promoter region was closely correlated with silencing of ARHI expression in 10–15% of patients with breast cancers [4, 10]. Aberrant methylation was accompanied with decreased ARHI expression in breast cancer

cells. Hypermethylation was detected at CpG island I of 67% breast cancer cells, 33% at CpG island II, and 56% at CpG island III, while hypomethylation at CpG island II of 44% breast cancer cells. Treatment with 5-aza-2deoxycytidine, a methyltransferase inhibitor, can demethylate and partially restore ARHI expression with hypermethylation of CpG islands [10]. ARHI expression was partially upregulated in cells with hypermethylation of CpG islands. CpG islands methylation was studied in 20 human tissues. On the other hand, no hypermethylation was found in CpG island I of surgical specimen, 15% hypermethylation in CpG island II, and 20% in CpG island III [10]. During imprinting, CpG islands are consistently methylated and silenced in the maternal allele of normal cells, whereas not in paternal alleles. CpG island II hypermethylation of both alleles completely eliminated ARHI promoter activity. The degree of ARHI methylation is related to the survival of patients [22], which provides a new mechanism for the breast tumors [10].

10.4 ARHI and Colon Cancer

ARHI expression was also downregulated in colon cancer cells, while overexpression could reduce the number of invaded cells and the adhesive ability [23] and promote colon cancer cell apoptosis [24]. ARHI expression was downregulated in 62% of colon cancer specimens, associated with worse differentiation degree and Dukes' stage. Methylation-specific PCR assay revealed that the methylation rates of ARHI were 53% and 47% in CpG Island I and CpG Island II, respectively. The promoter methylation may downregulate ARHI expression in colon cancer, which can be a therapeutic potential for the disease [25].

10.5 ARHI and Hepatocellular Carcinoma (HCC)

ARHI gene expression was found to be related to hepatocellular carcinoma, evidenced by the fact that ARHI expression was downregulated in 78.6% HCC specimens, accompanied by reduced levels of ARHI protein [26]. The overexpression of ARHI inhibited HCC growth and colony formation, while the silencing of endogenous ARHI promoted cell growth [26]. Upregulated ARHI expression inhibited tumor growth and angiogenesis in hepatocellular carcinoma, which were prevented by 5-aza-20-deoxycytidine [27, 28]. ARHI hypermethylation occurred in 47% of patients with HCC without ARHI expression. The downregulated expression of ARHI in HCCs acts as a tumor suppressor role, which was mainly stimulated by the epigenetic modification in HCC [26].

10.6 ARHI and Pancreatic Cancer

Overexpression of ARHI can inhibit the cell cycle and apoptosis in pancreatic tumor cells [29]. Compared with normal pancreatic tissues, ARHI is downregulated in approximately 50% in pancreatic cancer tissues. The immediate reason for this downregulation or loss of ARHI expression in pancreatic cancer cells was due to the aberrant methylation of ARHI locus. Hypermethylation was detected at CpG island I of 100% pancreatic cancer cells, at CpG island II of 40%, and at CpG island III of 80%, respectively. The growth of pancreatic cancer cells can be suppressed by the overexpression of ARHI which is involved with the apoptosis of cancer cells. The upregulation of ARHI mRNA expression induced by the demethylation of ARHI can obviously inhibit cell growth and increase apoptosis in human pancreatic cancer cells. It was evidenced that ARHI serves as a gene that inhibits growth in pancreatic cancers [30].

10.7 ARHI and Osteosarcoma (OS)

ARHI protein and RNA levels were downregulated in OS cells [31]. The knockdown of ARHI could promote OS cell proliferation and attenuate apoptosis. Zebularine may upregulate the tumor suppressor genes through a

demethylation function, which inhibits the growth and promotes apoptosis in OS cells. The ARHI expression was upregulated by Zebularine due to the downregulation of ARHI methylation and the function of DNA methyltransferase 1 (DNMT1) and histone methyltransferase G9a. The distinct reduction of ARHI methylation can be induced by knockdown of DNMT1 or G9a. Zebularine may directly repress DNMT1 alone, while G9a through regulating DNMT1 function on ARHI methylation, which were restored by knockdown of ARHI [32].

10.8 ARHI and Glial Tumors

Experimental studies demonstrated that expression of ARHI was downregulated in human glioma tumors as compared with normal brain tissue as well as four different glioma cells [33]. The proliferation and invasion of glioma cell can be suppressed by up-expression of ARHI [33]. The expression and methylation status of ARHI were evaluated in tissue and peripheral blood [34]. The expression of ARHI RNA increased in 67% of patients with glial tumor and decreased in 33% [34]. Methylation of the CpG island at ARHI was detected using the combined bisulfite restriction analysis and the restriction fragment length polymorphism in glial tumors as compared with hypermethylated healthy volunteers. Hypermethylation was detected at CpG island I in two glial tumors, indicating that the progression of glial tumor may be due to the downregulation of ARHI [34].

ARHI can be influenced by a large number of genetic events and epigenetic mechanisms [3, 22, 35, 36], while ARHI expression may be firstly silenced by the aberrant DNA methylation of ARHI, varying among cell types [10].

10.9 ARHI and Follicular Thyroid Carcinoma (FTC)

The global gene expression analysis showed that ARHI expression was low in **FTC**. Studies revealed that a complete methylation pattern was exist in ARHI in **FTC** shows [37]. The silencing of ARHI, primarily by large genomic deletion is involved with hypermethylation of the genomically imprinted allele, which may be an important early event in FTC [37].

10.10 ARHI and Lung Cancer

Studies demonstrated that overexpression of ARHI gene can inhibit the growth, proliferation and invasion of lung cancer cells, and promote the apoptosis of lung cancer cells [38]. Aberrant DNA methylation was observed in non-small cell lung cancers. The methylation status of 245 CpG positions in 59 candidate genes was examined in different types of lung cancer and normal adjacent lung tissues from smokers, which found that the DNA-methylation levels were different among different histological types of tumor tissues and normal adjacent tissue [39]. The highest degree of DNA methylations in squamous cell carcinoma was observed in ARHI, GP1Bbeta, RAR beta genes, etc. It was proposed that methylation profiles of specific genes may be used to distinguish histological types of lung cancer [39].

10.11 Conclusion and Perspectives

This chapter overviewed the importance of ARHI methylation and expression phenomes in various types of cancers, although the exact mechanisms remain unclear. As an imprinted gene, aberrant DNA methylation of the paternal allele of ARHI was identified as a primary inhibitor of ARHI expression. The role of methylation in the CpG islands of the ARHI promoter region vary among ovarian cancers, breast cancers, hepatocellular carcinoma, colon cancers, pancreatic cancer osteosarcoma, glial tumors, follicular thyroid carcinoma, or lung cancers. The methylation of ARHI provides a new insight to understand molecular mechanisms of tumorigenesis and progression of cancers.

There are further needs to explore whether ARHI methylation and expression can be defined

as disease-specific biomarkers with the specificity of disease duration, severity, stage, phase, phenome, and response to therapy as requested [40–46]. It is questioned whether the heterogeneity of ARHI methylations exists among cells of the same cancer. The single-cell sequencing was widely applied for the identification of the intra- and inter-heterogeneity among cancer locations, types, and durations within the cancer [47, 48]. Dynamic three-dimensional chromatin conformation and the potential association between cell-type specific chromatin conformation and differential DNA methylations should be considered in the understanding of ARHI methylation, since altered 3D genome controls gene regulation during development and disease [49–51]. Roles of ARHI methylation and expression in the development and diseases are furthermore specifically clarified by gene editing technologies, e.g., CRISPR [52–55]. Thus, we believe that the deep understanding of ARHI methylation and expression will provide new opportunities for future diagnosis and therapy.

References

1. Yu Y et al (1999) NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. Proc Natl Acad Sci U S A 96:214–219
2. Luo RZ et al (2001) Genomic structure and promoter characterization of an imprinted tumor suppressor gene ARHI. Biochim Biophys Acta 1519:216–222
3. Yu Y et al (2003) Epigenetic regulation of ARHI in breast and ovarian cancer cells. Ann N Y Acad Sci 983:268–277
4. Yu Y et al (2006) Biochemistry and biology of ARHI (DIRAS3), an imprinted tumor suppressor gene whose expression is lost in ovarian and breast cancers. Methods Enzymol 407:455–468
5. Jones P (1996) DNA methylation errors and cancer. Cancer Res 56:2463–2467
6. Toh T, Lim J, Chow E (2019) Epigenetics of hepatocellular carcinoma. Clin Transl Med 8(1):13. https://doi.org/10.1186/s40169-019-0230-0
7. Ghufran MS, Soni P, Kanade SR (2019) Aflatoxin-induced upregulation of protein arginine methyltransferase 5 is mediated by protein kinase C and extracellular signal-regulated kinase. Cell Biol Toxicol 35(1):67–80. https://doi.org/10.1007/s10565-018-9439-8
8. Baylin S, Herman J (2000) DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 16:168–174
9. Klett H et al (2018) Robust prediction of gene regulation in colorectal cancer tissues from DNA methylation profiles. Epigenetics 13:386–397
10. Yuan J et al (2003) Aberrant methylation and silencing of ARHI, an imprinted tumor suppressor gene in which the function is lost in breast cancers. Cancer Res 63:4174–4180
11. Peng H et al (2000) ARHI is the center of allelic deletion on chromosome 1p31 in ovarian and breast cancers. Int J Cancer 86:690–694
12. Lu Z, Bast R (2013) The tumor suppressor gene ARHI (DIRAS3) inhibits ovarian cancer cell migration through multiple mechanisms. Cell Adhes Migr 7:232–236
13. Lu Z et al (2008) The tumor suppressor gene ARHI regulates autophagy and tumor dormancy in human ovarian cancer cells. J Clin Invest 118:3917–3929
14. Santin A et al (2005) Gene expression fingerprint of uterine serous papillary carcinoma: identification of novel molecular markers for uterine serous cancer diagnosis and therapy. Br J Cancer 92:1561–1573
15. Washington M et al (2015) ARHI (DIRAS3)-mediated autophagy-associated cell death enhances chemosensitivity to cisplatin in ovarian cancer cell lines and xenografts. Cell Death Dis 6:e1836
16. Feng W et al (2008) Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation. Cancer 112:1489–1502
17. Li J et al (2013) STAT3 acetylation-induced promoter methylation is associated with downregulation of the ARHI tumor-suppressor gene in ovarian cancer. Oncol Rep 30:165–170
18. Zuo X et al (2014) Breast cancer cells are arrested at different phases of the cell cycle following the re-expression of ARHI. Oncol Rep 31:2358–2364
19. Li L et al (2014) JMJD2A contributes to breast cancer progression through transcriptional repression of the tumor suppressor ARHI. Breast Cancer Res 16:R56
20. Hisatomi H, Nagao K, Wakita K, Kohno N (2002) ARHI/NOEY2 inactivation may be important in breast tumor pathogenesis. Oncology 62:136–140
21. Shi Z et al (2002) [NOEY2 gene mRNA expression in breast cancer tissue and its relation to clinicopathological parameters]. Zhonghua zhong liu za zhi [Chin J Oncol] 24:475–478
22. Widschwendter M et al (2004) Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. Cancer Res 64:3807–3813
23. Ouyang J, Pan X, Hu Z (2017) The role of aplysia ras homolog I in colon cancer cell invasion and adhesion. Exp Ther Med 14:5193–5199
24. Chen X, Jiang Z (2016) [Overexpression of aplysia Ras homolog I (ARHI) increases apoptosis in colon

cancer SW480 cells]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 32:1503–1506
25. Wang W et al (2014) Loss of ARHI expression in colon cancer and its clinical significance. Contemp Oncol 18:329–333
26. Huang J et al (2009) ARHI, as a novel suppressor of cell growth and downregulated in human hepatocellular carcinoma, could contribute to hepatocarcinogenesis. Mol Carcinog 48:130–140
27. Zhao X, Li J, Zhuo J, Cai L (2010) Reexpression of ARHI inhibits tumor growth and angiogenesis and impairs the mTOR/VEGF pathway in hepatocellular carcinoma. Biochem Biophys Res Commun 403:417–421
28. Pei X, Yang Z, Liu H, Qiao S (2011) Aplasia Ras homologue member I overexpression induces apoptosis through inhibition of survival pathways in human hepatocellular carcinoma cells in culture and in xenograft. Cell Biol Int 35:1019–1024
29. Hu Y, Si L, Ye Z, Lin Z, Zhou J (2013) Inhibitory effect of ARHI on pancreatic cancer cells and NF-κB activity. Mol Med Rep 7:1180–1184
30. Yang H et al (2010) Imprinted tumor suppressor gene ARHI induces apoptosis correlated with changes in DNA methylation in pancreatic cancer cells. Mol Med Rep 3:581–587
31. Lu Z et al (2006) E2F-HDAC complexes negatively regulate the tumor suppressor gene ARHI in breast cancer. Oncogene 25:230–239
32. Ye K et al (2016) Zebularine enhances apoptosis of human osteosarcoma cells by suppressing methylation of ARHI. Cancer Sci 107:1851–1857
33. Chen J, Shi S, Yang W, Chen C (2014) Overexpression of ARHI decreases tumor growth, migration, and invasion in human glioma. Med Oncol 31:846
34. Yakut S et al (2011) Aplasia ras homologous member I gene and development of glial tumors. Balkan J Med Genet 14:37–44
35. Janssen E et al (2009) LOH at 1p31 (ARHI) and proliferation in lymph node-negative breast cancer. Cell Oncol 31:335–343
36. Yang J et al (2009) NOEY2 mutations in primary breast cancers and breast hyperplasia. Breast 18:197–203
37. Weber F et al (2005) Silencing of the maternally imprinted tumor suppressor ARHI contributes to follicular thyroid carcinogenesis. J Clin Endocrinol Metab 90:1149–1155
38. Wu X, Liang L, Dong L, Yu Z, Fu X (2013) Effect of ARHI on lung cancer cell proliferation, apoptosis and invasion in vitro. Mol Biol Rep 40:2671–2678
39. Field J et al (2005) Methylation discriminators in NSCLC identified by a microarray based approach. Int J Oncol 27:105–111
40. Ansari D, Torén W, Zhou Q, Hu D, Andersson R (2019) Proteomic and genomic profiling of pancreatic cancer. Cell Biol Toxicol 35(4):333–343. https://doi.org/10.1007/s10565-019-09465-9
41. Qiao T, Wang X (2019) A new light of proteomics in cell biology and toxicology. Cell Biol Toxicol 35(4):289–291. https://doi.org/10.1007/s10565-019-09492-6
42. Zhang L, Han X, Wang X (2019) Correction to: Is the clinical lipidomics a potential goldmine? Cell Biol Toxicol. https://doi.org/10.1007/s10565-019-09490-8
43. Wu D, Cheng Y, Wang X (2019) CSGT group. Definition of clinical gene tests. Cell Biol Toxicol 35 (2):83–87. https://doi.org/10.1007/s10565-019-09464-w
44. Song D, Tang L, Wang L, Huang J, Zeng T, Fang H, Wang X (2020) Roles of TGFβ1 in the expression of phosphoinositide 3-kinase isoform genes and sensitivity and response of lung telocytes to PI3K inhibitors. Cell Biol Toxicol 36:51. https://doi.org/10.1007/s10565-019-09487-3
45. Yin J, Wang Z, Li G, Lin F, Shao K, Cao B, Hou Y (2019) Characterization of circulating tumor cells in breast cancer patients by spiral microfluidics. Cell Biol Toxicol 35(1):59–66. https://doi.org/10.1007/s10565-018-09454-4
46. Qi X, Yu C, Wang Y, Lin Y, Shen B (2019) Network vulnerability-based and knowledge-guided identification of microRNA biomarkers indicating platinum resistance in high-grade serous ovarian cancer. Clin Transl Med 8(1):28. https://doi.org/10.1186/s40169-019-0245-6
47. Zeng Y, Chen X, Gao H, Wang X (2018) An artificial intelligent single cell is part of the cell dream world. Cell Biol Toxicol 34(4):247–249. https://doi.org/10.1007/s10565-018-9433-1
48. Busch S, Talamini M, Brenner S, Abdulazim A, Hänggi D, Neumaier M, Seiz-Rosenhagen M, Fuchs T (2019) Circulating monocytes and tumor-associated macrophages express recombined immunoglobulins in glioblastoma patients. Clin Transl Med 8(1):18. https://doi.org/10.1186/s40169-019-0235-8
49. Kong S, Zhang Y (2019) Deciphering Hi-C: from 3D genome to function. Cell Biol Toxicol 35(1):15–32. https://doi.org/10.1007/s10565-018-09456-2
50. Wang DC, Wang X (2018) Genome dimensions control biological and toxicological functions; myth or reality? Cell Biol Toxicol 34(5):333–336. https://doi.org/10.1007/s10565-018-9440-2
51. Sanchez A, Kuras M, Murillo JR, Pla I, Pawlowski K, Szasz AM et al (2019) Novel functional proteins coded by the human genome discovered in metastases of melanoma patients. Cell Biol Toxicol. https://doi.org/10.1007/s10565-019-09494-4
52. Schacker M, Seimetz D (2019) From fiction to science: clinical potentials and regulatory considerations of gene editing. Clin Transl Med 8(1):27. https://doi.org/10.1186/s40169-019-0244-7
53. Yan F, Wang W, Zhang J (2019) CRISPR-Cas12 and Cas13: the lesser known siblings of CRISPR-Cas9. Cell Biol Toxicol 35(6):489–492. https://doi.org/10.1007/s10565-019-09489-1
54. Li D, Zhou H, Zeng X (2019) Battling CRISPR-Cas9 off-target genome editing. Cell Biol Toxicol 35 (5):403–406. https://doi.org/10.1007/s10565-019-09485-5

55. Mills EM, Barlow VL, Luk LYP, Tsai YH (2019) Applying switchable Cas9 variants to in vivo gene editing for therapeutic applications. Cell Biol Toxicol. https://doi.org/10.1007/s10565-019-09488-2

Xiaozhuan Liu received her PhD on Epidemiology and Health Statistics at Zhengzhou University, and works as Researcher in the Center for Clinical Single Cell Biomedicine of Henan Provincial People's Hospital. She mainly engaged in the research of tumor pathogenesis. Currently, Xiaozhuan Liu is interested in the role of gene methylation in cancers.

Tingting Zhang received her postgraduate degree in People's Hospital of Zhengzhou University. She has strong background of epigenetic modification of oocytes. Currently, Tingting Zhang is interested in next generation sequencing data analysis and interpretation, especially on single-cell analysis and cancer research. She has published three SCI papers in the past 3 years.

Yanjun Li did her Master's degree in Cell Biology at Zhengzhou University. Now, she works as a research assistant at Center for Clinical Single Cell Biomedicine in Henan Provincial People's Hospital. Her main research direction is focused on lung cancer. Currently, Yanjun Li is interested in the epigenetic alterations in lung cancer, especially in DNA methylation.

Yuwei Zhang did her PhD in organic chemistry at Zhengzhou University. Her PhD thesis focused on the signal pathway for the expression of CTR1 gene, a copper ion transporter in the body. Currently, Yuwei mainly engaged in the analysis of biological information and pathogenic mechanism research.

Hui Zhang is an editor-in-chief assistant at Henan Provincial People's Hospital and a postgraduate in Zhengzhou University. She has read a large number of scientific articles on topics such as single-cell sequencing, gene editing, molecular drug targeting, stem cells, immune therapy, and heterogeneity during her editing work. In addition, years of experience in clinical surgery and ethics committee gave her a deep understanding of medical practice ethics and scientific research ethics.

Xiangdong Wang is a Distinguished Professor of Medicine, Director of Shanghai Institute of Clinical Bioinformatics, Executive Director of Clinical Science Institute of Fudan University Zhongshan Hospital, Director of Fudan University Center of Clinical Bioinformatics, Deputy Director of Shanghai Respiratory Research Institute, and visiting professor of King's College of London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular and cellular therapies. He is the author of more than 200 scientific publications with the impact factor about 900, citation number about 6918, h-index 46, i10-index 181, and cited journal impact factor about 7000.

Li Li is a Director of Department of Scientific Research and Discipline Construction, Henan Provincial people's Hospital. She is Member of clinical research group of Chinese Medical Association's Scientific Research Management Branch, Standing Committee member of Chinese Medical Association's Henan Research and Management Branch, and Vice-chairman of Henan Discipline Management Branch of Chinese Hospital Management Society. She has engaged in the management of medical scientific research for 30 years and her main research is focused on health management scientific research big data, laboratory biosafety, and medical ethics. She published more than 20 scientific papers.

Clinical Significance of P16 Gene Methylation in Lung Cancer

11

Yanjun Li, Tingting Zhang, Hui Zhang, Xiangdong Wang, Xiaozhuan Liu, Qihong Huang, and Li Li

Abstract

Lung cancer is the leading cause of death from cancer in China. The lack of early screening technologies makes most patients to be diagnosed at advanced stages with a poor prognosis which often miss the best treatment opportunities. Thus, identifying biomarkers for minimally invasive detection and prognosis of early stage disease is urgently needed. Genetic and epigenetic alterations that promote tumorigenesis and metastasis exist in multiple cancers. These aberrant alterations usually represent early events in cancer progression suggesting their potential applications as a biomarker for cancer prediction. Studies have shown that DNA methylation is one of the key factors in progression of lung cancer. P16 promoter methylation is one of the most common epigenetic change plays a key role in lung cancer. In this review, we highlight the p16 gene methylation and its clinical significance in lung cancer.

Keywords

Lung cancer · p16 · Methylation · Diagnosis · Prognosis

Y. Li · T. Zhang · H. Zhang · X. Liu
Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

X. Wang · Q. Huang
Zhongshan Hospital, Fudan University, Shanghai, Shanghai, China

L. Li (✉)
Department of Scientific Research and Discipline Construction, Henan Provincial People's Hospital, Zhengzhou, Henan, China

Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Henan University People's Hospital, Zhengzhou, Henan, China

Center for Clinical Single Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou, Henan, China
e-mail: lili@henu.edu.cn

11.1 Introduction

As one of the most common malignancies in China, lung carcinoma has become to be the main cause of cancer death. There are two major types of lung cancer: non-small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). NSCLC constituting 80% of all lung cancers is further divided into three main subtypes: adenocarcinoma (AC), squamous cell lung carcinoma (SCC), and large-cell lung carcinoma [1, 2]. The 5 years survival rate for lung patients is still very low (approximate 20%) due to the complex biological process and the scarcity of effective biomarkers. The most common epigenetic

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_11

alteration leading to NSCLC is the missing function of tumor suppressor gene p16, occurring in about 70% patients [3]. Functioning as a tumor suppressor, the methylation of p16 is associated with poor prognosis and therapeutic resistance of NSCLC patients. It is the aim of this review to evaluate the p16 gene methylation and its clinical significance in lung cancer.

11.2 DNA Methylation

Epigenetic alterations in DNA methylation, microRNA expression, modifications of histones, chromatin remodeling, and posttranscriptional modifications are considered major pathogenesis of lung cancer [3–5]. And studies have shown that the DNA methylation regarded as a hallmark contributed to carcinogenesis [6, 7]. The conventional view of DNA methylation is methyl groups tacked onto the 5′-carbon of cytosine in the DNA strand (5-methylcytosine, 5-mC) and this process is catalyzed by DNA methyltransferases (DNMTs), including DNMT1, DNMT2, and DNMT3 [8]. Although these 5-methylcytosine was thought to be a stable and heritable gene silencing mechanism, recent evidence has revealed that DNA changes rapidly and reversibly in methylation and demethylation [9]. The 5-methylcytosine can be converted to 5-hydroxymethylcytosine (5-hmC) with the catalysis of ten-eleven translocation (TET) proteins [10] (Fig. 11.1). CpG islands are some regions rich in CpG with a length of 300–3000 bp, which mainly locate near the promoters or exon regions of those highly expressed genes. And the CpG islands are common methylation in human tumors and related with tumor initiation and progression [11]. In lung cancer, the methylation genes involved in key cellular functions, such as DNA repair (i.e., O6-methylguanine-DNA methyltransferase, MGMT), growth and development (i.e., Short Stature Homobox2, SHOX2), and cell cycle (i.e., cyclin-dependent kinase inhibitor 2A, CDKN2A) [12, 13]. The p16 gene which is a tumor suppressor gene encoded by CDKN2A is the first found aberrant hypermethylation in lung cancer [14]. In the early stage of lung cancer progression, CDKN2A is silenced by DNMT1 which gene expression is upregulated in lung cancer [15].

11.3 Structure and Function of P16 Gene

The p16 gene (also known as CDKN2A, Multiple Tumor Suppressor 1 and as several other synonyms) is an anti-oncogene involved in the regulation of cell cycle and is encoded by the CDKN2A gene [16–19]. The CDKN2A gene is located at chromosome 9p21, 8.5 kb full length [20] and is composed of two introns and three exons: exon 1 (126 bp), exon 2 (307 bp), and exon 3 (11 bp). The exon 1 has two subtypes with two transcription methods, Exon 1α and Exon 1β, which is transcribed from its own promoter and encodes different proteins. The p16 protein, encoded by exons 1α, 2 and 3, is a nuclear phosphor-protein composed of 156 amino acids with a molecular weight of 16 KD. The main function of p16 gene is to inhibit the cell cycle by binding to cyclin-dependent kinases (CDKs) [20–22]. When CDKs binded to cyclin D, the active complex of cyclin D/CDKs phosphorylates the pRb protein, resulting in the release of E2F transcription factors and the transcription of gene critical for G1/S -phase regulation. The specific binding of p16 protein to the CDKs (such as CDK4 or CDK6) inhibits the formation of the active complex of cyclin D/CDKs by changing the allosteric conformation in these proteins [22]. Due to lacking of cyclin D/CDKs active complex, the retinoblastoma protein (Rb) maintains in its hypo-phosphorylated and growth-suppressive states. Hypo-phosphorylated Rb binds to E2F resulting in the arrest of G1 phase (Fig. 11.2). Indeed, several studies have demonstrated that the p16 gene is inactivated in multiple tumors including lung cancer [14, 23, 24]. And the loss of p16 gene has been shown to lead to carcinogenesis and metastasis in cancer patients with worse prognosis [25–28].

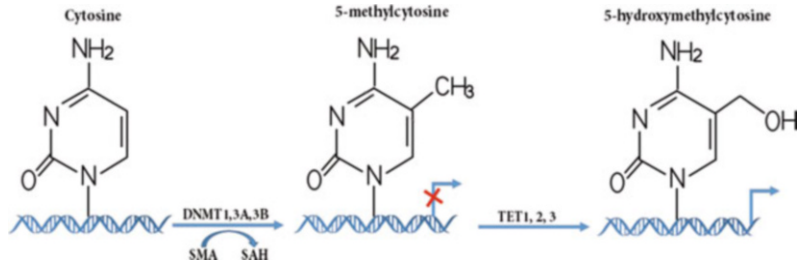


Fig. 11.1 Proposed DNA methylation and demethylation in the lung cancer. DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) catalyze DNA methylation at the 5′ cytosine of CpG sites (5mC) using S-adenosylmethionine (SAM) as a methyl donor. SAM is converted to S-adenosylhomocysteine (SAH). Members of the ten-eleven translocase family (TET1, TET2, and TET3) can catalyze the oxidation of the cytosine modification 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC)

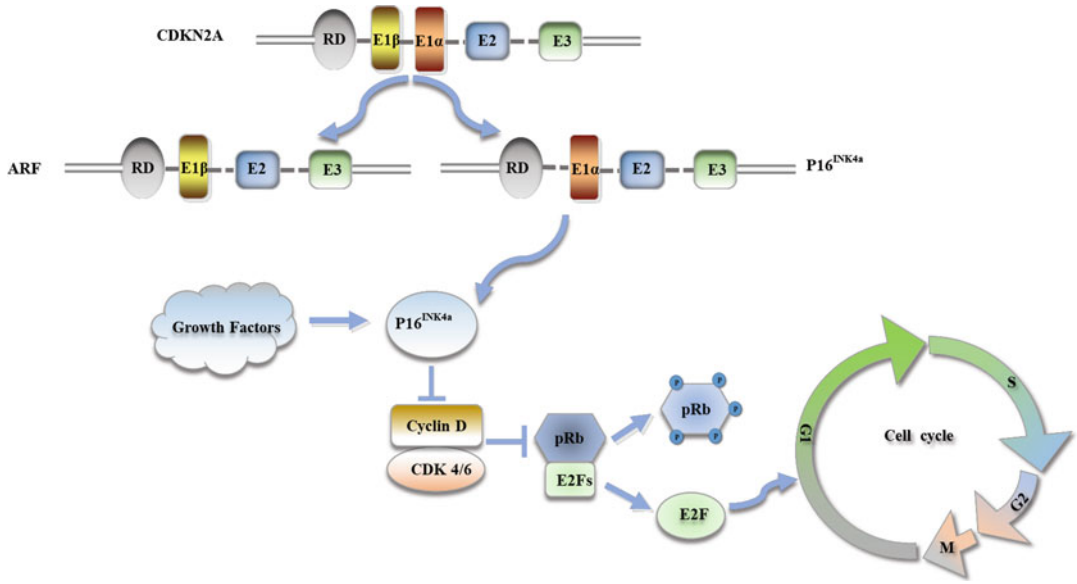


Fig. 11.2 Structure of the $p16^{INK4a}$/ARF locus and the role of $p16^{INK4a}$ in cell cycle. The $p16^{INK4a}$ protein binds to the cyclin D and CDK4/6 complexes and inhibits the activation of the transcription factors (E2Fs), which induces proteins to move from the G1 phase to S phase in the cell cycle

11.4 Methylation of P16 Gene and Lung Cancer

11.4.1 Frequency of P16 Methylation

DNA methylation plays a pivotal role in maintaining the genomic stability, which is dispensable for cells to maintain their physiological functions. Interestingly, alterations of DNA methylation are frequently observed in lung cancer. Although the p16 singling pathway regulates the G1/S phase, it is usually altered or mutated in many cancers. The p16 gene, cyclin D1, and CDKs, particularly p16 gene, are common abnormalities in NSCLC but is rare in SCLC. In SCLC, the Rb gene is usually inactivated [29]. In the detection of 78 paired NSCLC tumor and adjacent normal tissues, the p16 gene is demonstrated highly methylated in tumor tissues [30]. In another study, the methylation of p16 was in 34% of the 64 primary lung carcinomas. Further analysis indicated that the p16 gene is highly methylated in large-cell carcinoma (71%),

compared with adenocarcinoma (36%) and squamous cell carcinoma (33%) [31]. Therefore, it is believed that p16 methylation is a common event in lung cancer and Table 11.1 shows the frequency of p16 methylation in lung cancer. Compared with normal lung tissues, cancer tissues contain a higher occurrence rate in p16 methylation and according to literature, lung squamous cell carcinoma has a higher methylation level than that in lung adenocarcinoma whereas it rarely happens in small cell lung cancer.

11.4.2 Mechanisms of P16 Methylation in Lung Cancer

Inactivation of p16 is common in lung cancer and occurs via three major mechanisms: homozygous deletion (HD), point mutation, and the most common methylation in the promoter CpG island [32, 33]. Evidence has suggested that there exists a marked correlation of methylation changes after smoking. In a study analysis the correlation of p16 methylation and smoking, researcher detected the methylation of p16 in 81 NSCLC patients with smoking and 41 never-smoking patients. The result proved that p16 gene methylated in 52.6% cases and was significantly highly methylated among smokers [34]. The same results were also founded in other studies [35, 36]. Furthermore, there could be other inducements, such as air pollution, that can lead to p16 methylation. PM2.5 is a type of pollution less than 2.5 μm in diameter. Previous studies have shown that long-term exposure to PM2.5 contributes to the lung cancer risk [36–38]. Indeed, the frequency of p16 methylation increased in mice which exposed to PM2.5 for 8 h daily and last for 3 weeks. The same result has showed in the primary murine alveolar epithelial cells exposed to PM2.5 for daily doses of fine urban PM. Interestingly, chronically elevated levels of PM2.5 has been associated with an increase in lung cancer incidence further correlates indirectly the p16 methylation and lung cancer [39]. Despite the above potential mechanism by which smoking or air pollution could increase the p16 methylation thus eventually cancer, it should be noted that p16 methylation rarely occurs in SCLC. Considering smoking is a major risk factor for SCLC, other mechanisms of the p16 methylation in lung cancer have yet to be identified.

11.4.3 P16 Methylation for Early Detection and Diagnosis in Lung Cancer

Epigenetic changes were shown to be one of the most important mechanisms contributed to the oncogenesis and development of cancer [40]. Moreover, aberrant DNA methylation occurs in the early stage of cancer, thus representing excellent biomarkers for cancer early detection [41, 42]. This hypothesis can be further evidenced when considering frequent detection of methylated DNA in cancer tissue and body fluid in malignant carcinomas such as lung cancer [43, 44], liver cancer [45], and breast cancer [46, 47]. In a meta-analysis included all the published articles advocates that methylated p16 and other eight genes (SOX17, CDO1, ZFP42, TAC1, FAM19A4, RASSF1A, FHIT, MGMT) are useful biomarkers in the screening and auxiliary detection of lung cancer. And the p16 methylation is associated with ethnicity and sample size [48]. In order to investigate the association between promoter methylation of RASSF1A and p16 and the clinicopathological features in lung cancer, researchers retrieved 389 studies and analyses of 1259 lung cancer patients. The results provide convincing evidence that the promoter methylation ratio of p16 is associated with histological types and smoking status in lung cancer, indicating that p16 could be used as an effective biomarker in early diagnosis [49]. The same results were also reported in other studies. In a study, researcher detected the methylation of candidate genes (p16 and other seven genes) in 42 plasma samples from primary NSCLC patients and 10 plasma samples from healthy donors. The results proved that p16 gene showed no methylation in all healthy plasma, while is methylated in 45% cases of plasma from lung cancers [50]. In another study,

Table 11.1 P16 gene methylation in lung cancer (compared to corresponding normal lung tissue; listed papers with >50 cases)

Author	Cases (tumor tissue)	Histologic type	Methylation in tumor tissue	Methylation in corresponding normal lung tissue	Ref.
Lin Q, et al.	124	NSCLC	22.58%	7.69% ($N = 26$)	[58]
Jin M, et al.	72	NSCLC	29%	8% ($N = 72$)	[36]
Brock MV, et al.	104	NSCLC	52%	26% ($N = 50$)	[42]
Safar AM, et al.	105	NSCLC	39%	25% ($N = 25$)	[59]
Grote HJ, et al.	75	25 AC 25 SCC 25 SCLC	12% AC 56% SCC 4% SCLC	0% ($N = 64$)	[60]
Kim YT, et al.	61	42 AC 17 SCC	67% AC 67% SCC	46% ($N = 61$)	[61]
Tanaka R, et al.	57	AC	40% All	17% ($N = 53$)	[62]
Kim H, et al.	74	31 AC 43 SCC	28% All	6% ($N = 127$)	[63]
Toyooka S, et al.	514	299 AC 194 SCC 21 others	25% All 19% AC 35% SCC	5% ($N = 84$)	[64]
Jarmalaite S, et al.	64	NSCLC and SCLC	34% All	9% ($N = 58$)	[31]
Yanagawa N, et al.	75	43 AC 29 SCC 3 others	31% All 16% AC 48% SCC	0% ($N = 75$)	[65]
Zochbauer-Muller S, et al.	88	45 AC 43 SCC	27% All 13% AC 37% SCC	0% ($N = 104$)	[66]

180 samples including cancer tissues, adjacent normal lung tissues, blood plasma, and exhaled breath condensate (EBC) from 30 NSCLC patients and 30 healthy controls were analyzed for aberrant promoter methylation of p16. Results demonstrated that the frequency of p16 promoter methylation in tumor tissues, blood plasma, and EBC from tumor patients was 86.66%, 50%, and 40%, respectively, whereas it was not observed in the samples from the healthy controls. It was found that the detection of p16 promoter methylation in EBC was feasible, thus presenting a potential noninvasive biomarker for NSCLC diagnosis [51]. However, it should be noted that the positive rate was still low compared with that in tissues thus requesting more researches to develop it as a NSCLC diagnostic biomarker. In summary, these findings indicated that p16 methylation could be an effective biomarker for NSCLC diagnosis.

11.4.4 The P16 Methylation in Prognosis

Aberrant DNA methylation of tumor suppressor genes has been linked to cancer development and clinical outcome in non-small cell lung cancer (NSCLC) [43, 52, 53]. A study analyzed the promoter methylation of p16 in tumor tissues from 193 surgically treated NSCLC patients of stage I, Ib, IIa, and IIb (127 were older than 60 year old; 66 were 60 year old and younger). Patients were dichotomized according to the age using the cut-points of 40, 50, and 60 years of age at diagnosis. The results showed that lower frequency of p16 methylation was observed in specimens from 60 years or younger compared with older than 60 years and was significantly lower in specimens from 40 years. Further analyzed of the p16 methylation status for the survival rate, the result indicated patients with p16 non-methylated have a significantly extended survival time in patients of 60-year or younger, but no such association with clinical outcome in

patients older than 60-year age group [54]. This result was similar to other studies, which found p16 promoter hypermethylation is associated with worse survival time in lung cancer patients with an early stage [55–57]. Investigators analyzed the methylation status of p16 in 155 lung tumor tissues revealed that the patients with hypermethylated p16 had significantly shortened survival time than patients without p16 hypermethylation [57].

11.5 Conclusion

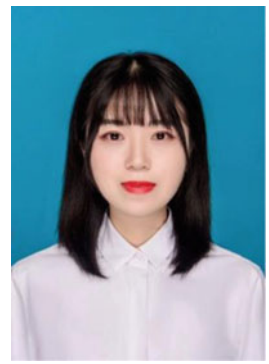
The relationship between DNA methylation and prognostic significance continues to be an area of interest for investigation. Previous studies have revealed that the p16 gene methylated is significantly associated with clinical and pathological features in NSCLC. Although evidence has showed the methylation of p16 gene could serve as a potential diagnostic biomarker that may facilitate the early detection and maybe useful to identify high-risk patients with lung cancer presenting at early stage once identified, the application of its methylation in the clinical has yet to be developed. Current methylation studies were mostly in tumor tissue, whereas that in blood and other samples has lower sensitivity and specificity. In addition, tumor markers in blood are not organ specific and cannot be used for the diagnosis of judged lung cancer. More research to support the use of p16 methylation as an early diagnostic marker of NSCLC is needed.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F et al (2016) Cancer statistics in China, 2015. CA Cancer J Clin 66:115–132
2. Jiang Y, Liang Y, Li L, Zhou L, Cheng W, Yang X et al (2019) Targeting neddylation inhibits intravascular survival and extravasation of cancer cells to prevent lung-cancer metastasis. Cell Biol Toxicol 35:233–245
3. Toh TB, Lim JJ, Chow EK (2019) Epigenetics of hepatocellular carcinoma. Clin Transl Med 8(1):13
4. Wu D, Cheng Y, Wang X, CSGT group (2019) Definition of clinical gene tests. Cell Biol Toxicol 35:83–87
5. Gao D, Zhu B, Sun H, Wang X (2017) Mitochondrial DNA methylation and related disease. Adv Exp Med Biol 1038:117–132
6. Gu J, Wen Y, Zhu S, Hua F, Zhao H, Xu H et al (2013) Association between P(16INK4a) promoter methylation and non-small cell lung cancer: a meta-analysis. PLoS One 8:e60107
7. Schacker M, Seimetz D (2019) From fiction to science: clinical potentials and regulatory considerations of gene editing. Clin Transl Med 8(1):27
8. Ghufran MS, Soni P, Kanade SR (2019) Aflatoxin-induced upregulation of protein arginine methyltransferase 5 is mediated by protein kinase C and extracellular signal-regulated kinase. Cell Biol Toxicol 35(1):67–80
9. Shen K, Cao Z, Zhu R, You L, Zhang T (2019) The dual functional role of MicroRNA-18a (miR-18a) in cancer development. Clin Transl Med 8(1):32
10. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–935
11. Rhee YY, Lee TH, Song YS, Wen X, Kim H, Jheon S et al (2015) Prognostic significance of promoter CpG island hypermethylation and repetitive DNA hypomethylation in stage I lung adenocarcinoma. Virchows Arch 466:675–683
12. Gong J, Hendifar A, Tuli R, Chuang J, Cho M, Chung V et al (2018) Combination systemic therapies with immune checkpoint inhibitors in pancreatic cancer: overcoming resistance to single-agent checkpoint blockade. Clin Transl Med 7:32
13. Selamat SA, Galler JS, Joshi AD, Fyfe MN, Campan M, Siegmund KD et al (2011) DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma. PLoS One 6:e21443
14. West J, Beck S, Wang X, Teschendorff AE (2013) An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. Sci Rep 3:1630
15. Liu B, Song J, Luan J, Sun X, Bai J, Wang H et al (2016) Promoter methylation status of tumor suppressor genes and inhibition of expression of DNA methyltransferase 1 in non-small cell lung cancer. Exp Biol Med 241:1531–1539
16. Lukas J, Parry D, Aagaard L, Mann DJ, Bartkova J, Strauss M et al (1995) Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. Nature 375:503–506
17. Zhao R, Choi BY, Lee MH, Bode AM, Dong Z (2016) Implications of genetic and epigenetic alterations of CDKN2A (p16(INK4a)) in cancer. EBioMedicine 8:30–39
18. Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV et al (1994) A cell cycle regulator potentially involved in genesis of many tumor types. Science 264:436–440

19. Witcher M, Emerson BM (2009) Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. Mol Cell 34:271–284
20. Gil J, Peters G (2006) Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nat Rev Mol Cell Biol 7:667–677
21. Quelle DE, Zindy F, Ashmun RA, Sherr CJ (1995) Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell 83:993–1000
22. Serrano M, Hannon GJ, Beach D (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature 366:704–707
23. Kim JC, Choi JS, Roh SA, Cho DH, Kim TW, Kim YS (2010) Promoter methylation of specific genes is associated with the phenotype and progression of colorectal adenocarcinomas. Ann Surg Oncol 17:1767–1776
24. Lin J, Zeng RM, Li RN, Cao WH (2014) Aberrant DNA methylation of the P16, MGMT, and hMLH1 genes in combination with the methylenetetrahydrofolate reductase C677T genetic polymorphism and folate intake in gastric cancer. Genet Mol Res 13:2060–2068
25. Cui C, Gan Y, Gu L, Wilson J, Liu Z, Zhang B et al (2015) P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. Genome Biol 16:252
26. Fakhry C, Ferris RL (2018) P16 as a prognostic biomarker for nonoropharyngeal squamous cell cancers: avatar or mirage? J Natl Cancer Inst 110:1290–1291
27. Bryant AK, Sojourner EJ, Vitzthum LK, Zakeri K, Shen H, Nguyen C et al (2018) Prognostic role of p16 in nonoropharyngeal head and neck cancer. J Natl Cancer Inst 110:1393–1399
28. Liu Z, Lin H, Gan Y, Cui C, Zhang B, Gu L et al (2019) P16 methylation leads to paclitaxel resistance of advanced non-small cell lung cancer. J Cancer 10:1726–1733
29. Sekido Y, Fong KM, Minna JD (1998) Progress in understanding the molecular pathogenesis of human lung cancer. Biochim Biophys Acta 1378:F21–F59
30. Zhang Y, Wang R, Song H, Huang G, Yi J, Zheng Y et al (2011) Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer Lett 303:21–28
31. Jarmalaite S, Kannio A, Anttila S, Lazutka JR, Husgafvel-Pursiainen K (2003) Aberrant p16 promoter methylation in smokers and former smokers with nonsmall cell lung cancer. Int J Cancer 106:913–918
32. Hauptman N, Glavač D (2013) Long non-coding RNA in cancer. Int J Mol Sci 14:4655–4669
33. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28:1057–1068
34. Chan V, Tong TM, Chan TP, Tang M, Wan CW, Chan FY et al (2015) Meta-analyses of gene methylation and smoking behavior in non-small cell lung cancer patients. Sci Rep 5:8897
35. Liu Y, Lan Q, Siegfried JM, Luketich JD, Keohavong P (2006) Aberrant promoter methylation of p16 and MGMT genes in lung tumors from smoking and never-smoking lung cancer patients. Neoplasia 8:46–51
36. Jin M, Kawakami K, Fukui Y, Tsukioka S, Oda M, Watanabe G et al (2009) Different histological types of non-small cell lung cancer have distinct folate and DNA methylation levels. Cancer Sci 100:2325–2330
37. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G et al (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol 14:813–822
38. Hystad P, Demers PA, Johnson KC, Carpiano RM, Brauer M (2013) Long-term residential exposure to air pollution and lung cancer risk. Epidemiology 24:762–772
39. Soberanes S, Gonzalez A, Urich D, Chiarella SE, Radigan KA, Osornio-Vargas A et al (2012) Particulate matter air pollution induces hypermethylation of the p16 promoter via a mitochondrial ROS-JNK-DNMT1 pathway. Sci Rep 2:275
40. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674
41. Tuo L, Sha S, Huayu Z, Du K (2018) P16 gene promoter methylation as a biomarker for the diagnosis of non-small cell lung cancer: an updated meta-analysis. Thorac Cancer 9:1032–1040
42. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S et al (2008) DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 358:1118–1128
43. Shen N, Du J, Zhou H, Chen N, Pan Y, Hoheisel JD et al (2019) A diagnostic panel of DNA methylation biomarkers for lung adenocarcinoma. Front Oncol 9:1281
44. Ooki A, Maleki Z, Tsay JJ, Goparaju C, Brait M, Turaga N et al (2017) A panel of novel detection and prognostic methylated DNA markers in primary non-small cell lung cancer and serum DNA. Clin Cancer Res 23:7141–7152
45. Zhou Y, Wang X, Qiu X, Shuai Z, Wang C, Zheng F (2018) CDKN2A promoter methylation and hepatocellular carcinoma risk: a meta-analysis. Clin Res Hepatol Gastroenterol 42:529–541
46. Bai Y, Shen Y, Yuan Q, Lv C, Xing Q (2019) Evaluation of relationship between occurrence of liver cancer and methylation of fragile histidine triad (FHIT) and P16 genes. Med Sci Monit 25:1301–1306
47. Rivandi M, Khorrami MS, Fiuji H, Shahidsales S, Hasanzadeh M, Jazayeri MH et al (2018) The 9p21 locus: a potential therapeutic target and prognostic marker in breast cancer. J Cell Physiol 233:5170–5179
48. Liu D, Peng H, Sun Q, Zhao Z, Yu X, Wang H et al (2017) The indirect efficacy comparison of DNA methylation in sputum for early screening and

auxiliary detection of lung cancer: a meta-analysis. Int J Environ Res Public Health 14(7):679
49. Han J, Xu F, Chen N, Qi G, Wei Y, Li H et al (2016) Promoter methylations of RASSF1A and p16 is associated with clinicopathological features in lung cancers. J Cancer Res Ther 12:340–349
50. Huang X, Wu C, Fu Y, Guo L, Kong X, Cai H (2018) Methylation analysis for multiple gene promoters in non-small cell lung cancers in high indoor air pollution region in China. Bull Cancer 105:746–754
51. Xiao P, Chen J, Zhou F, Lu C, Yang Q, Tao G et al (2014) Methylation of P16 in exhaled breath condensate for diagnosis of non-small cell lung cancer. Lung Cancer 83:56–60
52. El-Sherif WT, Sayed SK, Galal SH, Makhlouf HA, Hassan AT, Yousef HA et al (2016) Diagnostic role of RASSF1A and p16INK4a promoter gene hypermethylation in serum DNA of lung cancer patients: clinicopathological significance. Egypt J Immunol 23:1–16
53. Yang ZP, Qi WB, Sun L, Zhou H, Zhou B, Hu Y (2019) DNA methylation analysis of selected genes for the detection of early-stage lung cancer using circulating cell-free DNA. Adv Clin Exp Med 28:355–360
54. Bradly DP, Gattuso P, Pool M, Basu S, Liptay M, Bonomi P et al (2012) CDKN2A (p16) promoter hypermethylation influences the outcome in young lung cancer patients. Diagn Mol Pathol 21:207–213
55. Kim DH, Nelson HH, Wiencke JK, Zheng S, Christiani DC, Wain JC et al (2001) p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res 61:3419–3424
56. Wang J, Lee JJ, Wang L, Liu DD, Lu C, Fan YH et al (2004) Value of p16INK4a and RASSF1A promoter hypermethylation in prognosis of patients with resectable non-small cell lung cancer. Clin Cancer Res 10:6119–6125
57. Gu J, Berman D, Lu C, Wistuba II, Roth JA, Frazier M et al (2006) Aberrant promoter methylation profile and association with survival in patients with non-small cell lung cancer. Clin Cancer Res 12:7329–7338
58. Lin Q, Geng J, Ma K, Yu J, Sun J, Shen Z et al (2009) RASSF1A, APC, ESR1, ABCB1 and HOXC9, but not p16INK4A, DAPK1, PTEN and MT1G genes were frequently methylated in the stage I non-small cell lung cancer in China. J Cancer Res Clin Oncol 135:1675–1684
59. Safar AM, Spencer H III, Su X, Coffey M, Cooney CA, Ratnasinghe LD et al (2005) Methylation profiling of archived non-small cell lung cancer: a promising prognostic system. Clin Cancer Res 11:4400–4405
60. Grote HJ, Schmiemann V, Geddert H, Rohr UP, Kappes R, Gabbert HE et al (2005) Aberrant promoter methylation of p16(INK4a), RARB2 and SEMA3B in bronchial aspirates from patients with suspected lung cancer. Int J Cancer 116:720–725
61. Kim YT, Lee SH, Sung SW, Kim JH (2005) Can aberrant promoter hypermethylation of CpG islands predict the clinical outcome of non-small cell lung cancer after curative resection? Ann Thorac Surg 79:1180–1188
62. Tanaka R, Wang D, Morishita Y, Inadome Y, Minami Y, Iijima T et al (2003) Loss of function of p16 gene and prognosis of pulmonary adenocarcinoma. Cancer 103(3):608–615
63. Kim H, Kwon YM, Kim JS, Lee H, Park JH et al (2004) Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer. Clin Oncol 22:2363–2370
64. Toyooka S, Maruyama R, Toyooka KO, McLerran D, Feng Z, Fukuyama Y et al (2003) Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer 103:153–160
65. Yanagawa N, Tamura G, Oizumi H, Takahashi N, Shimazaki Y, Motoyama T (2003) Promoter hypermethylation of tumor suppressor and tumor-related genes in non-small cell lung cancers. Cancer Sci 94(7):589–592
66. Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD (2001) Aberrant promoter methylation of multiple genes in non-small cell lung cancers. Cancer Res 61:249–255



Yanjun Li did her Master's degree in Cell Biology at Zhengzhou University. Now, she works as a research assistant at Center for Clinical Single Cell Biomedicine in Henan Provincial People's Hospital. Her main research direction is focused on lung cancer. Currently, Yanjun Li is interested in the epigenetic alterations in lung cancer, especially in DNA methylation.

Tingting Zhang received her postgraduate degree in People's Hospital of Zhengzhou University. She has strong background of epigenetic modification of oocytes. Currently, Tingting Zhang is interested in next-generation sequencing data analysis and interpretation, especially on single cell analysis and cancer research. She has published three SCI papers in the past 3 years.

Hui Zhang is an editor-in-chief assistant at Henan Provincial People's Hospital and a postgraduate in Zhengzhou University. She has read a large number of scientific articles on topics such as single-cell sequencing, gene editing, molecular drug targeting, stem cells, immune therapy, and heterogeneity during her editing work. In addition, years of experience in clinical surgery and ethics committee gave her a deep understanding of medical practice ethics and scientific research ethics.

Xiangdong Wang is a Distinguished Professor of Medicine, Director of Shanghai Institute of Clinical Bioinformatics, Executive Director of Clinical Science Institute of Fudan University Zhongshan Hospital, Director of Fudan University Center of Clinical Bioinformatics, Deputy Director of Shanghai Respiratory Research Institute, and visiting professor of King's College of London. His main research is focused on clinical bioinformatics, disease-specific biomarkers, lung chronic diseases, cancer immunology, and molecular & cellular therapies. He is the author of more than 200 scientific publications with the impact factor about 900, citation number about 6918, h-index 46, i10-index 181, and cited journal impact factor about 7000.

Xiaozhuan Liu received her PhD. on Epidemiology and Health Statistics at Zhengzhou University, and work as Researcher in the Center for Clinical Single Cell Biomedicine of Henan Provincial People's Hospital. She mainly engaged in the research of tumor pathogenesis. Currently, Xiaozhuan Liu is interested in the role of gene methylation in cancers.

Qihong Huang is a Professor of Medicine at the Department of Clinical Sciences of Fudan University Zhongshan Hospital. His research mainly focuses on lung cancer early detection, evolution, and treatment.

Li Li, Director of Department of Scientific Research and Discipline Construction, Henan Provincial people's Hospital. She is Member of clinical research group of Chinese Medical Association's Scientific Research Management Branch, Standing Committee member of Chinese Medical Association's Henan Research and Management Branch, Vice-chairman of Henan Discipline Management Branch of Chinese Hospital Management Society. She has engaged in the management of medical scientific research for 30 years and her main research is focused on health management scientific research big data, laboratory biosafety, and medical ethics. She published more than 20 scientific papers.

Application of Single-Cell RNA Sequencing in Pancreatic Cancer and the Endocrine Pancreas

12

Qiankun Luo, Qiang Fu, Xu Zhang, Hongwei Zhang, and Tao Qin

Abstract

The pancreas is a complex organ composed of an endocrine (pancreatic islets) and an exocrine portion. This mixed cell population has resulted in an implacable barrier to exploring the detailed mechanism and function of each cell type in previous investigative approaches. In recent years, single-cell RNA sequencing (scRNA-seq) technologies have provided in-depth analysis of cell heterogeneity in the pancreas and in pancreatic cancer. It is especially effective in cell-type-specific molecule identification and detection of interactions between cancer cells and the stromal microenvironment. To date, numerous reports have described the application of scRNA-seq in studies of pancreatic islets and pancreatic cancer. The aim of this paper is to review recent advances of pancreatic transcriptomics and pancreatic cancer using scRNA-seq strategies.

Keywords

Single cell · RNA sequencing · Transcriptomics · Pancreatic islet · Heterogeneity · Pancreatic ductal adenocarcinoma · Circulating tumor cells · Cancer stem cells · Stromal cell

Q. Luo · Q. Fu · X. Zhang · H. Zhang · T. Qin (✉)
Department of Hepato-Biliary-Pancreatic Surgery, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, Henan, China

12.1 Introduction

The pancreas is a crucial organ for human digestion and metabolism. The emergence of diabetes is associated with the destruction of pancreatic β cells (Type 1) and insulin resistance accompanied by β cell dysfunction (Type 2). The incidence of diabetes is increasing gradually and threatening global health [1]. Nevertheless, the molecular mechanism of diabetes remains to be elucidated. Current studies show a relationship between transcriptome variations of the pancreatic islets and diabetes [2, 3]. However, the islets have an abundance of distinct cell types, which increase difficulties in detecting cell-type-specific transcriptomes. Thus, the heterogeneity and specific markers for α, β, ε, δ, and PP cells have not been comprehensively elaborated upon until now.

The exocrine gland of the pancreas is composed of acinar and ductal cells. Previous studies believed pancreatic ductal adenocarcinoma (PDAC) was derived from ductal cells because the tumor histology resembled that of ductal morphology [4]. However, subsequent studies showed acinar-to-ductal metaplasia could be induced in Kras mutated mice, and precancerous lesions occurred [5]. New research has demonstrated that both acinar and ductal cells can generate PDAC with distinct biological features [6]. In addition, it has been confirmed that pancreatic stellate cells divide into cancer-associated fibroblasts (CAFs), which contributes to the stromal microenvironment of PDAC

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_12

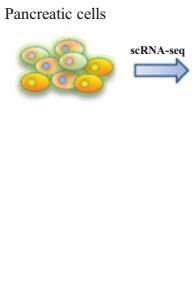
[7]. Cancer stem cells (CSCs) and circulating tumor cells (CTCs) are identified as having a vital role in therapeutic resistance and recurrence of malignancy [8–11]. However, the cellular characteristics and molecular mechanism are difficult to uncover using traditional technologies due to low counts of target cells and interference by heterogeneous cells.

Therefore, the advent of single-cell RNA sequencing (scRNA-seq) technologies truly offers an avenue to definitely understand the features, functions, and distinctions of these cell subtypes. In this review, we will summarize the discoveries of the endocrine pancreas and PDAC that have resulted from the implementation of scRNA-seq and discuss the significance and deficiencies of these findings.

12.2 Pancreatic Cell Type Identification and Transcriptome Analysis by scRNA-seq

To deeply understand pancreatic cell function and promote the study of diabetes development, Muraro MJ et al. developed a SORT-seq platform based on FACs, robotic liquid handling systems, and CEL-seq2 methodologies. With this scRNA-seq method, they improved single-cell transcripts by twofold over the conventional CEL-Seq2 approach [12]. They confirmed the specific factors of different pancreatic cell types in the human pancreas (Fig. 12.1). After that, they mapped the unique transcriptional markers of each cell type. According to the data, they identified distinct clusters that were characteristic for each cell type, and the specific genes expressed from the endocrine and exocrine cells by single-cell sequencing. Additionally, genes, including ITPR1, SLC6A4, and GHSR, that had been previously proposed to have a function in pancreatic diseases were verified by enriching α, PP, and δ cells [12]. Another research study investigated the original cells of the dorsal and ventral pancreas by single-cell transcriptomic sequencing and mapped their distinct developmental progression from pancreatic progenitors [13]. Similarly, Maayan Baron, et al. mapped transcriptomes of human and mouse pancreatic cells through scRNA-seq. Except for the above cell types validated by Muraro MJ et al., they also detected stellate cells, vascular cells, tissue-resident macrophages, mast cells, cytotoxic T cells, and Schwann cells. Genes were shown to play vital roles in cell drug therapy, cellular maturation, stellate cell activation, and Schwann cell dedifferentiation [14].

Type 2 diabetes (T2D) is thought to incapacitate pancreatic beta cells. Previous studies have demonstrated that transcriptome variations of the pancreatic islets play an important role in the development of T2D [15, 16]. However, the heterogeneity of the pancreatic islet cells, including alpha, beta, and delta cells, remains elusive in view of those indiscriminate bulk cell studies. The scRNA-seq approach conquered these shortages and has been applied for the detection of transcripts of the pancreatic islet cells [17–20]. The first application of scRNA-seq in the exploration of human pancreatic islet cells was performed by Jin Li et al. The research uncovered specific genes expressed in pancreatic endocrine cells (i.e., α cells, β cells, and PP cells) such as MORF4L1, IRX2, BMI1, MEIS1, and ETV1 [19]. Yurong Xin et al. provided a deep analysis of differentially expressed genes of the islet cells in T2D and normal pancreatic tissue, showing the relevance between the unique clusters and T2D. The 245 genes were shown to be modulated due to T2D, and the function of a large proportion of them in development of T2D remains to be explored. Genes enriched in α and β cells that were screened showed significant differentiation between the human and mouse pancreas [17]. However, specific gene expression of PP, ε and δ cells were detected by the same means. Acinar cells exhibiting MHC class II genes were demonstrated in this study [21]. These studies characterized different cell populations of the pancreas and have provided a resource for further analysis of pancreatic cell subtypes and biological function. However, most of the specific molecular properties of the cells have not been identified. Deep sequencing of each cell type should be



Cells	Genes	Transcription Factors (TF)
Alpha cells	GCG LOXL4 PLCE1 IRX2 GC KLHL41 CRYBA2 TTR TM4SF4 RGS4	IRX2 FEV ARX PTGER3 HMGB3 RFX6 MAFB SMARCA1 PGR LDB2
Beta cells	INS IAPP MAFA NPTX2 DLK1 ACYAP1 PFKFB2 PDX1 TGFBR3 SYT13	MAFA PDX1 SMAD9 CDKN1C TFCP2L1 SIX3 SIX2 MNX1 BMP5 PIR
Delta cells	SST PRG4 LEPR RBP4 BCHE HHEX FRZB PCSK1 RGS2 GABRG2	HHEX ERBB4 POU3F1 ISL1 PSIP1 BHLHE41 PDLIM4 EHF LCORL ETV1
Pancreatic polypeptide	PPY SERTM1 CARTPT SLITRK6 ETV1 THSD7A AQP3 ENTRD2 PTGFR CHN2	ETV1 MEIS2 ID2 EGR3 LMO3 MEIS1 ID4 ARX PAX6 ZNF503
Epsilon cells	GHRL ANXA13 PHGR1 ACSL1 FRZB SPTSSB ASGR1 HEPACAM2 VTN SERPINA1	VTN EBF1 BMP7 CDKN2A PROX1 ARX ZKSCAN1
Duct cells	SPP1 CFTR AQP1 ALDH1A3 KRT19 CRP DEFB1 CEACAM6 MMP7 TSPAN8	ONECUT2 LITAF SOX4 DAB2 CREB5 HLA-DQB1 WWTR1 PARGC1A PKHD1 NFIB
Acinar cells	PNLIP REG1B PRSS1 ALB PRSS3P2 CPA2 CTRB2 CEL PLA2G1B CELA3A	GATA4 MECOM NR5A2 ZFP36L1 CSDA CEBPD CREB3L1 XBP1 LGR4 NUPR1
Mesenchyma l cells	COL1A1 COL1A2 COL3A1 COL6A3 FN1 SFRP2 COL5A1 SPARC COL15A1 SERPINE1	WNT5A SNAI2 NOTCH3 FBN1 HEYL PRRX1 UACA AEBP1 TBX3 FOXF2
Endothelial cells	FLT1 KDR CD93 ESAM SOX18 PECAM1 ESM1 PASK SLCO2A1 PLVAP	SOX18 RGCC SMAD6 ERG PRDM1 TCF4 NOTCH4 SNAI1 NKX2-3 ETS1

Fig. 12.1 Special-expressed genes and enriched transcription factors in different cell subtypes of pancreas

implemented to describe detailed traits of the cells and establish a cell function map.

Another study illustrated cell type heterogeneity of the endocrine pancreas in children and adults by scRNA-seq. The data indicate that differentially expressed transcripts of alpha and beta cells in T2D adults have a certain amount of similarity with the pancreatic samples of children. It was confirmed that the sonic hedgehog signaling pathway contributes to the proliferation of alpha cells [18]. This study offers some important insights into the heterogeneity of pancreatic endocrine cells in adults and children and facilitates the investigation of pancreatic development. With regard to the influence of age, Martin Enge, et al. utilized scRNA-seq to identify transcriptomic features and variation traits in different ages [22]. The research verified that CDKN2A is a unique molecule related to aging at the single-cell level and that transcriptional noise there is a characteristic of increasing along with aging. The pattern of somatic mutation caused by aging was identified to have an obvious correlation with DNA damage induced by ROS. The scRNA-seq protocols provide a valuable insight into the character of CDKN2A and how its level of expression is impacted by aging but not by an initiation program [23]. While previous reports suggested the existence of aging-induced degeneration of cellular and organ functions, more research using scRNA-seq has identified that aging has no negative effect on the β-cell population or function in the mouse pancreas [24, 25]. This method could become a meaningful tool to advance an accurate understanding of aging and other factors that affect individual pancreatic cells.

12.3 Application of scRNA-seq to Gene Function in Pancreatic Cancer and CSCs

APE1 functions in the activation of many signal pathways as a DNA repair protein. It has an important role in the emergence of cancers of the lung, breast, colon, and pancreas [26–29]. However, the mechanism of its impacts on these cancers has not been illustrated. Several studies were implemented to detect the regulation of differential expression by APE1, but the results were not encouraging on account of the difficult goal of building a perfect APE1-knockout cell line [30]. Furthermore, although siRNA structure downregulated the expression of APE1, differentially expressed reads among cells brought notable bias. Thus, scRNA-seq technologies were used to cover these shortages. Fenil Shah et al. transfected pancreatic ductal adenocarcinoma (PDAC) cells with APE1 siRNA and scrambled siRNA. The group demonstrated that gene expression changes with the downregulation of APE1 [29]. They accurately analyzed the differentially expressed genes (DEGs) among

completely silenced APE1, inadequately silenced PDAC cells and control cells by scRNA-seq. Six genes (TMEM45A, TMEM126A, TMEM154, COMMD7, ISYNA1, and TNFAIP2) were found with overlapping expression relative to the APE1 knockdown [29]. Subsequent study indicated that scRNA-seq is a potent tool with high sensitivity for the investigation of transcriptome variation in cancer cells and a subtle way to dissect the signal pathways regulated by genes in PDAC.

Another research study combined scRNA-seq with spatial transcriptomics (ST), which remedied the shortage of weak cell resolution in ST. With this approach, they detected PDAC cells separated from resected tissue. The data noise was decreased by a *k*-nearest neighbor smoothing algorithm. Six cell types with different clusters were identified, including T cells, macrophages, fibroblasts, endothelial cells, and pericytes. They next completed the transcriptomic sequencing of PDAC tissue and spotted the spatial site of unique molecular markers and genes using ST. Finally, a spatial map of the cells was formed with corresponding spot transcripts in each cell type [31]. This study illustrated the cell heterogeneity and spatial locations within tumors utilizing scRNA-seq and ST protocols and provided a clear structure of the PDAC microenvironment. However, profound analysis of cellular spatial formation and functions are still awaiting clarification. Meanwhile, the small numerous transcriptome reads of single cells makes the establishment of spatial patterns challenging and introduce a higher risk of bias. Novel strategies should be developed to promote accurate and sensitive research [32].

Lytle NK et al. utilized scRNA-seq to confirm the unique molecular expression profiles in pancreatic cancer stem cells (CSCs) and to construct the molecular landscape of CSCs. Lytle, Ferguson [11] constructed a map of Msi+ cells, which are defined as characteristic of CSCs, showed obvious coherence with the expression levels of IL10Rβ, IL34, and Csf1r in stromal cells. These factors might be controlled by the retinoic-acid-receptor-related orphan receptor gamma (RORγ). scRNA-seq showed that RORγ expression was obviously upregulated in pancreatic CSCs, and RORγ proved to be a latent target and prognostic marker of pancreatic cancer. In response to the small quantity and unique molecular features of CSCs, scRNA-seq techniques could be significantly beneficial for the dissection of CSCs and provide a novel insight for PDAC metastasis and therapeutic resistance.

12.4 Exploration of the Pancreatic Cancer Stromal Microenvironment with scRNA-seq

With the exception of tumor cells, fibroblasts, immune cells, endothelial cells, and the extracellular matrix (ECM) comprise the microenvironment of pancreatic cancer. The microenvironment could provide tumors a protective barrier for drug or immune resistance, or secrete cytokines to promote cell proliferation and metastasis. The stromal cells of pancreatic cancer have been identified as playing an important role in tumor progression, growth, and metastasis [33]. Cancer-associated fibroblasts (CAFs), which make up the main part of the tumor stroma, are thought to make a considerable contribution to tumorigenesis, proliferation, and invasion [34, 35]. For instance, a PDAC mouse model with depletion of α-SMA myofibroblasts showed poor survival. The degrees of tumor differentiation were reduced significantly in myofibroblast-depleted mice. Moreover, mice treated with the smoothened inhibitor antagonized Hedgehog signaling, which resulted in a similar tumor pattern [36, 37]. CAFs have proved to be abundant in many cancers [38]. However, their molecular peculiarity has not been extensively explored, although α-smooth muscle actin (α-SMA), FSP-1 and fibroblast-activation protein α (FAPα) have been demonstrated to be a potential marker [39, 40].

In recent years, scRNA-seq was extensively applied in the investigation of cell heterogeneity involved in the cancer field. This technology allows the study of the PDAC microenvironment at the individual cell level and reduces the

sequencing noise caused by bulk signals. Giulia Biff et al. utilized scRNA-seq in detecting the heterogeneity of CAFs in PDAC tissue in a mouse model. CAFs and PDAC cells underwent scRNA-seq after coculturing. According to previous investigations, they found that CAFs expressed two molecular subtypes [41]. Myofibroblastic CAFs (myCAFs) were defined by high expression levels of α-SMA, which reportedly have a strong correlation with PDAC [42]. On the other hand, the population that overexpressed inflammatory cytokines and chemokines was named inflammatory CAFs (iCAFs). Then, scRNA-seq was performed to verify that the same molecular markers that identified myCAFs and iCAFs were present in mouse tumors. IL1/JAK/STAT3 and TGFβ/SMAD3 signaling were demonstrated to be key modulation pathways in controlling the heterogeneity and functions of myCAFs and iCAFs (Fig. 12.2) [7]. In consideration that there might be other populations in the PDAC stromal microenvironment, the team proceeded to explore different CAF subtypes using a scRNA-seq strategy. With an analysis of the sequence data from individual tumor cells, they mapped a cluster-enriched plot. It is noticeable that a novel CAF population was found in both human PDAC tissue and the mouse model. This new CAF subtype has specific molecular markers consistent with the Major Histocompatibility Complex (MHC) class II family and can induce CD25 and CD69 activation of T cells. Thus, these cells possessing the capacity of antigen presentation were confirmed to be antigen-presenting CAFs (apCAFs) [43]. Another paper presents new evidence for microenvironment degradation in precursor lesions of PDAC. This group detected transcriptomic heterogeneity in intraductal papillary mucinous neoplasms (IPMNs) and PDAC by scRNA-seq. They found that expression profiles of low-grade IPMNs (LGD-IPMN) showed high correlation with invasiveness and that the immune cells infiltrating the surrounding area were prone to be exhausted. Additionally, stromal CAFs transform into iCAF and myCAF subtypes and secrete factors that suppress immunoreaction and facilitate tumor progression, although there are some differences in the amounts of subtype cells between LGD-IPMN and high-grade IPMNs (HGD-IPMN) [44]. This deteriorating phenomenon of the microenvironment in IPMNs may represent a hypothesis that stromal malignancy in precursor lesions precedes that in tumors. The specific stromal molecules expressed in precursor lesions perhaps accelerate the declaration of new biomarker and prevention of IPMN transition to PDAC.

Now that stromal cell heterogeneity has been illustrated, the subpopulations of tumor cells and their regulated deserve continuing study. The current study described distinct PDAC cell types from solid tumor and circulating tumor cells. Epithelial-to-mesenchymal transition (EMT), proliferative (PRO) phenotypes and coexpression subtypes are proposed exist in the PDAC cell population [45]. Meanwhile, CAFs are not only modulated by cancer cells; they could reverse to promote or inhibit cancer cells in a complex mechanism [46, 47]. Subsequently, scRNA-seq and single-cell proteomics were applied to deeply explore how CAFs impact PDAC primary tumor cell variation. Matteo Ligorio et al. validated that CAFs could induce PDAC cell transformations to EMT, PRO, and double positive (DP) subpopulations by the interaction of CAFs and PDAC cells. TGFβ, MAPK/ERK, and STAT3 signaling pathways are contributors to the formation of PDAC subtypes (Fig. 12.3). They confirmed that the development of these specific PDAC cell phenotypes was determined by different ratios of CAFs and tumor cell coculturing patterns. Furthermore, the composition of cell types and the intratumoral spatial structure have different effects on tumor growth and progression [48].

Overall, scRNA-seq exhibited significant advantages in distinguishing cell subtype and intratumoral heterogeneity. The application of scRNA-seq in the investigation of tumor microenvironments will bring remarkable progress in the detection of the biological features and function of stromal cells.

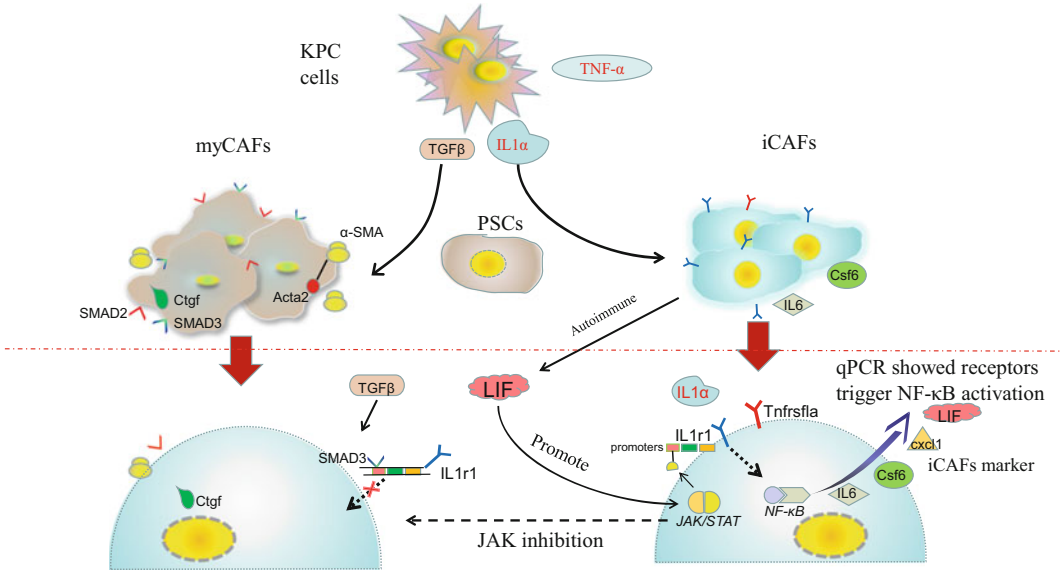


Fig. 12.2 TGF-β and IL-1 secreted by PDAC induce differentiation of PSCs into iCAFs and myCAFs subtypes, JAK/STAT and SMAD3 are the vital signal pathway that facilitate the activation

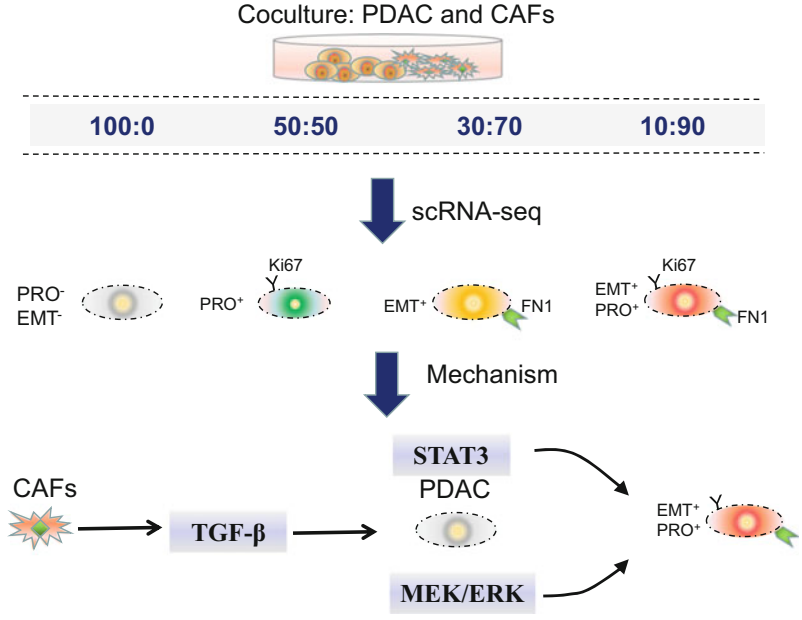


Fig. 12.3 CAFs drive PRO, EMT, and DP phenotype in PDAC. TGF-β, STAT3 and MEK/ERK signal pathway promote the development of DP population

12.5 Dissection of Circulating Tumor Cells Using Single-Cell Analysis

With a poor prognosis and highly invasive tumors, pancreatic cancer is extremely difficult to diagnose early. Most of the patients have local or distant metastasis tumor when they are initially diagnosed with PDAC. Despite this, there is a series of treatment methods, including chemotherapy, radiotherapy, and tumor resection; however, the 5-year survival rate is less than 9% [49, 50]. Therefore, there is urgency in the search for a sensitive tumor biomarker for the management of PDAC. Circulating tumor cells (CTCs) flow into the peripheral blood after they stem from solid tumors [51]. They share specific molecular markers and cell heterogeneity with the primary tumors. For instance, CTC subtypes expressing EMT or PRO traits have an obvious correlation with tumor invasiveness and patient survival [48]. However, research of hepatocellular cancer showed that CTCs with EMT had an inconspicuous relationship with the prognosis of patients, despite opinions that epithelial and mesenchymal features might exhibit characteristics such as those of cancer stem cells (CSCs) [9, 52]. On the other hand, not all CTCs can produce a metastatic tumor, even if CTCs have been detected in most of the malignant tumors, and the total number of CTCs is not associated with the recurrence rate or metastatic risk [8]. We hypothesize that these rare CTCs possess high potential of invasion and capacity for tumor regeneration if the CSCs escape from primary tumor. Conducting a further investigation to uncover the heterogeneity and biological features of CTCs would be very meaningful for the prediction and therapy of PDAC.

It is beneficial to take CTCs as pathologic biopsies in PDAC in response to the complex operation and risks in fine needle aspiration (FNA). However, it is truly a challenging thing to separate and extract CTCs for sequencing from blood because of the minute amounts [51]. The application of single-cell sequencing has promoted not only CTC detection as a "liquid biopsy" but also the analysis of tumor heterogeneity and its molecular mechanism in PDAC. Min Yu et al. first applied scRNA-seq in the analysis of mouse PDAC CTCs. They identified nine genes related to tumorigenesis that were obviously highly expressed in CTC and overlapped with transcripts sequenced from mouse and human PDAC solid tumors [53]. Wnt2 was demonstrated to be especially upregulated in primary tumors, ascites cells and CTCs. Notably, the enrichment of Wnt2 in CTCs was found to be present in most PDAC patients. Another study analyzed individual transcripts sequenced from CTCs and mouse PDAC cell lines by scRNA-seq. They defined several specific molecules expressed by CTCs that are connected with the CSCs, EMT, and ECM. Among these, the SPARC protein increased significantly in both mouse and human PDAC CTCs, which contributed to development of the ECM, and has been identified in expressed clusters as playing an important role in cell invasion. Additionally, they also discovered a CTC subpopulation that has proliferative traits and named this population PRO [45]. Similarly, mRNA expression of human PDAC CTCs was detected by single-cell sequencing technology. The results indicated that CTC-enriched clusters including CD24, CD44, and ALDH1A1 resemble CSCs and that SPARC is correlated with EMT [10]. However, the investigation also illustrated that scRNA-seq was deficient in coverage of the target genes, especially in detecting heterozygous CTCs [54]. This deficiency produced a low detection rate of differentially expressed genes and target variations in CTCs. For instance, the positive rate of KRAS variation in PDAC CTCs was measured at 27.7% by scRNA-seq, while it was detected at 92% in bulk CTCs. Thus, the technical bias of scRNA-seq is still a major problem that must be improved to promote the application of scRNA-seq to CTC exploration [55].

12.6 Conclusions

Single-cell transcriptomics analysis has brought new insight into cell heterogeneity and specific cell-type biomarkers [56]. The pancreas is a

complex organ comprising the endocrine pancreas, which includes α, β, ε, δ, and PP cells, and the exocrine portion, which includes acinar and ductal cells. Each cell type has unique respective functions. The scRNA-seq technology has facilitated the transcriptomics and cell heterogeneity studies in pancreatic islets and PDAC. Cell-specific markers and their potential functions were identified in islet cells [13, 19]. Meanwhile, novel molecules expressed in β cells were found to contribute to the development of T2D [21]. Some genes and pathways may play critical roles in the differentiation, immunoreaction, and aging of the pancreas [22]. The transcriptomics atlas provides a foundation for deep exploration of the biology and pathology of the pancreas [12, 17]. Investigations in PDAC revealed a more accurate understanding of CSCs, CTCs, and cell heterogeneity using the scRNA-seq method. The consensus now is that intratumoral cell-to-cell distinction has a crucial role in tumor recurrence and drug resistance. CSCs are the pivotal population related to clinical prognosis and can protect themselves against toxins and the immune system [57]. scRNA-seq permits analysis of differentially expressed genes of individual cells such as CSCs and CTCs, which provides an approach to dissecting the cell features and gene functions of these cells. CSCs circulating in peripheral blood might be the crucial population of CTCs that result in distant metastasis. In addition, cancer stromal cells were found to have effects on tumor progression and invasiveness. Both stromal cells and tumor cells have subtypes that are associated with different cancer processes. Degeneration of the microenvironment may precede tumorigenesis, which could bring new insight into the prevention and early diagnosis of cancers. The scRNA-seq technologies will be a potent tool in promoting these explorations.

Conflicts of Interest Statement Qiankun Luo, Qiang Fu, Xu Zhang, Hongwei Zhang, and Tao Qin have no conflicts of interest.

References

1. Talchai C et al (2012) Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. Cell 150(6):1223–1234
2. Tang Y et al (2014) Genotype-based treatment of type 2 diabetes with an α2A-adrenergic receptor antagonist. Sci Transl Med 6(257):257ra139
3. Taneera J et al (2012) A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab 16(1):122–134
4. Storz P (2017) Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol 14(5):296–304
5. Kopp JL et al (2012) Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. Cancer Cell 22(6):737–750
6. Ferreira RMM et al (2017) Duct- and acinar-derived pancreatic ductal adenocarcinomas show distinct tumor progression and marker expression. Cell Rep 21(4):966–978
7. Biffi G et al (2019) IL1-induced JAK/STAT signaling is antagonized by TGFbeta to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. Cancer Discov 9(2):282–301
8. Court CM et al (2015) Improving pancreatic cancer diagnosis using circulating tumor cells: prospects for staging and single-cell analysis. Expert Rev Mol Diagn 15(11):1491–1504
9. Agnoletto C et al (2019) Heterogeneity in circulating tumor cells: the relevance of the stem-cell subset. Cancers 11(4):483
10. Lapin M et al (2017) Single-cell mRNA profiling reveals transcriptional heterogeneity among pancreatic circulating tumour cells. BMC Cancer 17(1):390
11. Lytle NK et al (2019) A multiscale map of the stem cell state in pancreatic adenocarcinoma. Cell 177 (3):572–586.e22
12. Muraro MJ et al (2016) A single-cell transcriptome atlas of the human pancreas. Cell Syst 3(4):385–394. e3
13. Li L-C et al (2018) Single-cell transcriptomic analyses reveal distinct dorsal/ventral pancreatic programs. EMBO Rep 19(10):e46148
14. Baron M et al (2016) A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst 3(4):346–360. e4
15. Zhou Y et al (2014) TCF7L2 is a master regulator of insulin production and processing. Hum Mol Genet 23 (24):6419–6431
16. Taneera J et al (2015) Identification of novel genes for glucose metabolism based upon expression pattern in human islets and effect on insulin secretion and glycemia. Hum Mol Genet 24(7):1945–1955

17. Xin Y et al (2016) RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab 24 (4):608–615
18. Wang YJ et al (2016) Single cell transcriptomics of the human endocrine pancreas. Diabetes 65 (10):3028–3038
19. Li J et al (2016) Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep 17(2):178–187
20. Xin Y et al (2016) Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proc Natl Acad Sci U S A 113(12):3293–3298
21. Segerstolpe A et al (2016) Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab 24(4):593–607
22. Enge M et al (2017) Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell 171(2):321–330. e14
23. Arda HE et al (2016) Age-dependent pancreatic gene regulation reveals mechanisms governing human beta cell function. Cell Metab 23(5):909–920
24. Xin Y et al (2016) Single-cell RNAseq reveals that pancreatic β-cells from very old male mice have a young gene signature. Endocrinology 157 (9):3431–3438
25. Kenyon CJ (2010) The genetics of ageing. Nature 464:504
26. Lu GS et al (2018) APE1 stimulates EGFR-TKI resistance by activating Akt signaling through a redox-dependent mechanism in lung adenocarcinoma. Cell Death Dis 9(11):1111
27. Chen T et al (2017) The expression of APE1 in triple-negative breast cancer and its effect on drug sensitivity of olaparib. Tumor Biol 39(10):1010428317713390
28. Lou D et al (2014) Aberrant expression of redox protein Ape1 in colon cancer stem cells. Oncol Lett 7 (4):1078–1082
29. Shah F et al (2017) APE1/Ref-1 knockdown in pancreatic ductal adenocarcinoma—characterizing gene expression changes and identifying novel pathways using single-cell RNA sequencing. Mol Oncol 11 (12):1711–1732
30. Whitaker AM, Flynn TS, Freudenthal BD (2018) Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. Nat Commun 9(1):399–399
31. Moncada R et al (2019) Building a tumor atlas: integrating single-cell RNA-Seq data with spatial transcriptomics in pancreatic ductal adenocarcinoma bioRxiv 254375. https://doi.org/10.1101/254375
32. Fazal FM et al (2019) Atlas of subcellular RNA localization revealed by APEX-Seq. Cell 178(2):473–490. e26
33. Dougan SK (2017) The pancreatic cancer microenvironment. Cancer J 23(6):321–325
34. Ohlund D, Elyada E, Tuveson D (2014) Fibroblast heterogeneity in the cancer wound. J Exp Med 211 (8):1503–1523
35. Kleeff J et al (2016) Pancreatic cancer. Nat Rev Dis Primers 2:16022
36. Ozdemir BC et al (2014) Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell 25(6):719–734
37. Rhim AD et al (2014) Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer Cell 25(6):735–747
38. Smith NR et al (2013) Tumor stromal architecture can define the intrinsic tumor response to VEGF-targeted therapy. Clin Cancer Res 19(24):6943–6956
39. Zhang J et al (2013) Fibroblast-specific protein 1/S100A4–positive cells prevent carcinoma through collagen production and encapsulation of carcinogens. Cancer Res 73(9):2770
40. Garin-Chesa P, Old LJ, Rettig WJ (1990) Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers. Proc Natl Acad Sci U S A 87(18):7235–7239
41. Ohlund D et al (2017) Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med 214(3):579–596
42. Feig C et al (2012) The pancreas cancer microenvironment. Clin Cancer Res 18(16):4266–4276
43. Elyada E et al (2019) Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov 9(8):1102–1123
44. Bernard V et al (2019) Single-cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. Clin Cancer Res 25 (7):2194–2205
45. Ting DT et al (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. Cell Rep 8 (6):1905–1918
46. Gore J, Korc M (2014) Pancreatic cancer stroma: friend or foe? Cancer Cell 25(6):711–712
47. Laklai H et al (2016) Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. Nat Med 22 (5):497–505
48. Ligorio M et al (2019) Stromal microenvironment shapes the intratumoral architecture of pancreatic cancer. Cell 178(1):160–175.e27
49. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. CA Cancer J Clin 69(1):7–34
50. Aier I et al (2019) A systematic assessment of statistics, risk factors, and underlying features involved in pancreatic cancer. Cancer Epidemiol 58:104–110
51. Lianidou ES, Strati A, Markou A (2014) Circulating tumor cells as promising novel biomarkers in solid cancers. Crit Rev Clin Lab Sci 51(3):160–171
52. Chen Y et al (2019) Circulating tumor cells undergoing EMT are poorly correlated with clinical

stages or predictive of recurrence in hepatocellular carcinoma. Sci Rep 9(1):7084
53. Yu M et al (2012) RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature 487(7408):510–513
54. Court CM et al (2016) Reality of single circulating tumor cell sequencing for molecular diagnostics in pancreatic cancer. J Mol Diagn 18(5):688–696
55. Luo Q, Zhang H (2018) Emergence of bias during the synthesis and amplification of cDNA for scRNA-seq. Adv Exp Med Biol 1068:149–158
56. Niu F et al (2016) Potentials of single-cell biology in identification and validation of disease biomarkers. J Cell Mol Med 20(9):1789–1795
57. Clarke MF (2019) Clinical and therapeutic implications of cancer stem cells. N Engl J Med 380 (23):2237–2245

Qiankun Luo is a Doctoral Candidate at Zhengzhou University People's Hospital. Acute pancreatitis and pancreatic cancer are the primary focus of his research. He is also interested in genomics and tumor microenvironment. He has published six scientific papers in well-known international and domestic journals.

Qiang Fu is a surgeon of the Department of Hepatobiliary Pancreatic Surgery at Zhengzhou University People's Hospital. His main research interests are molecular mechanism of acute pancreatitis. He has participated in one project funded by the National Natural Science Foundation of China and published several articles in international journals.

Xu Zhang, Doctoral Candidate, Surgery major, Zheng Zhou University (China), devoted to the clinical and basic research of hepatobiliary and pancreatic surgical diseases, participated in two national research projects and published several articles in international journals.

Hongwei Zhang is a ward director of the Department of Hepatobiliary Pancreatic Surgery, a master tutor of Zhengzhou University People's Hospital. His main research is focused on the accurate resection of liver and pancreatic cancer, and molecular biomarkers of acute pancreatitis. He presided one project funded by the National Natural Science Foundation of China.

Tao Qin is a doctoral tutor of Zhengzhou University People's Hospital, master tutor of Henan University People's Hospital. His main research is focused on pathogenesis, molecular therapies, and specific biomarkers of acute pancreatitis and pancreatic cancer. He presided two projects funded by the National Natural Science Foundation of China and published over 30 scientific papers on the international and domestic authoritative journal.

Single-Cell Sequencing in Genitourinary Malignancies

13

Neal Murphy, Pratik Shah, Andrew Shih, Houman Khalili, Anthony Liew, Xinhua Zhu, and Annette Lee

Abstract

Single-cell sequencing (SCS) is a powerful new tool that applies Next Generation Sequencing at the cellular level. SCS has revolutionized our understanding of tumor heterogeneity and the tumor microenvironment, immune infiltration, cancer stem cells (CSCs), circulating tumor cells (CTCs), and clonal evolution. The following chapter highlights the current literature on SCS in genitourinary (GU) malignancies and discusses future applications of SCS technology. The renal cell carcinoma (RCC) section highlights the use of SCS in characterizing the initial cells driving tumorigenesis, the intercellular mutational landscape of RCC, intratumoral heterogeneity (ITH) between primary and metastatic lesions, and genes driving RCC cancer stem cells (CSCs). The bladder cancer section will also illustrate molecular drivers of bladder cancer stem cells (BCSCs), SCS use in reconstructing tumor developmental history and underlying subclones, and understanding the effect of cisplatin on intratumoral heterogeneity in vitro and potential mechanisms behind platinum resistance. The final section featuring prostate cancer will discuss how SCS can be used to identify the cellular origins of benign prostatic hyperplasia and prostate cancer, the plasticity and heterogeneity of prostate cancer cells with regard to androgen dependence, and the use of SCS in CTCs to understand chemotherapy resistance and gene expression changes after androgen deprivation therapy (ADT). The studies listed in this chapter illustrate many translational applications of SCS in GU malignancies, including diagnostic, prognostic, and treatment-related approaches. The ability of SCS to resolve intratumor heterogeneity and better define the genomic landscape of tumors and CTCs will be fundamental in the new era of precision-based care.

Keywords

Single-cell sequencing · Next generation sequencing · Renal cell carcinoma · Bladder cancer · Prostate cancer · Tumor heterogeneity · Tumor microenvironment · Cancer stem cells · Circulating tumor cells · Clonal evolution · Chemotherapy resistance

N. Murphy (✉)
Department of Medicine, Long Island Jewish Medical Center, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, New Hyde Park, NY, USA
e-mail: nmurphy2@northwell.edu

P. Shah
Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA

A. Shih · H. Khalili · A. Liew · A. Lee
Feinstein Institute for Medical Research, Manhasset, NY, USA

X. Zhu
Northwell Health Cancer Institute, Lake Success, NY, USA

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_13

13.1 Introduction

Genitourinary (GU) malignancies, including prostate, bladder, and renal cell carcinoma (RCC), are associated with significant morbidity and mortality in the United States, and individually have an annual incidence that ranks in the top ten of all malignancies [1]. Next Generation Sequencing (NGS) is changing the way we approach these cancers. By allowing rapid and accurate sequencing of individual tumors, NGS has helped drive a precision oncology approach, in which the best targeted and immunological therapies can be chosen leading to improved outcomes and better overall survival [2]. Moving progress further along, single-cell sequencing (SCS) is a powerful new tool that applies NGS at the cellular level. SCS has revolutionized our understanding of tumor heterogeneity and the tumor microenvironment, immune infiltration, cancer stem cells (CSCs) and rare tumor cells, clonal evolution and circulating tumor cells (CTCs). The following chapter highlights the current literature on SCS in GU malignancies and discusses future directions.

13.2 Renal Cell Carcinoma

The first paper for the section of this chapter on renal cell carcinoma by Young et al. 2018 illustrates how SCS can be used to identify the cellular origin of a specific cancer. The authors studied 72,501 single-cell transcriptomes of human renal tumors, integrated with tumor whole-genome DNA sequences. Single cells were taken from Wilms tumors ($n = 3$), clear cell RCC (ccRCC, $n = 3$), papillary RCC (pRCC, $n = 1$) and compared to healthy fetal ($n = 2$), pediatric ($n = 3$), adolescent ($n = 2$), and adult ($n = 5$) kidneys, and adult ureters ($n = 4$). Normal tissue biopsies were taken from macroscopically normal areas of kidneys being resected for either transplantation or for cancer. The 72,501 fetal, normal, and cancer cells and their respective transcriptomes were split into distinct clusters based on a community detection algorithm. Normal and fetal cells were then assigned to a reference map by cross-referencing their assigned cluster-defining transcripts with canonical markers known from established studies. Of the 42,809 non-malignant cells, 37,951 mature kidney cells represented epithelial cells from throughout the nephron, mostly consisting of proximal tubular cells, but fibroblasts, myofibroblasts, and vascular endothelial cells were also included. The remaining 4858 fetal cells were grouped into developing nephron cells, vascular endothelial cells, myofibroblasts, fibroblasts, and ganglion cells [3].

After establishing the landscape of healthy kidney cells, the authors characterized the identity of 17,821 immune cells and 6333 nonimmune cells from the tumors listed previously. By combining a genotyping and similarity analysis, the authors found that Wilms cells represent abnormal fetal cells, and the cells matched with specific developing nephron populations. To validate the single-cell identification of the Wilms tumors, the authors compared the bulk transcriptomes of an independent group of Wilms tumors ($n = 124$) to a group of fetal, pediatric, and adult normal tissue, and other childhood tumors ($n = 135$). The specific developing populations in the single-cell analysis were evident and limited to the Wilms tumors and normal tissues from the bulk analysis [3].

When comparing the ccRCC and pRCC tumors to normal mature cells, the authors found that the tumor cells retained transcriptional features similar to the PT1 cluster, a subtype of a proximal convoluted tubular cells. The authors observed that within their data, the PT1 cluster was defined by expression of SLC17A3 and VCAM1, and an absence of SLC17A13. They hypothesized that PT1 cells are the nearest normal cell correlate of ccRCC cells, and that given both ccRCC and pRCC have the presence of a PT1 signature, these cancers may have a similar origin with different fates. When analyzing the tumor microenvironment, the authors focused on VEGF signaling and uncovered a complex circuit involving tumor infiltrating macrophages, and two populations of ascending vasa recta cells [3]. Overall, by comparing normal tissue to

tumor cells this study showed how we can translate our understanding of tumor pathogenesis from the micro-anatomical to a more defined cellular and molecular level, highlighting which cells may be the initial driving force for tumorigenesis.

The next paper by Xu et al. 2012 illustrates how SCS can be applied to understand the intercellular mutational landscape of RCC. The authors studied clonal diversity in a single clear cell RCC tumor from a 59-year-old Chinese male with Stage IV disease according to the 2002 AJCC TNM classification [4]. Whole exome sequencing was first performed on a sample from the patient's tumor and local adjacent tissue. Exome sequencing revealed that the patient's tumor was unlikely to be related to the presence of VHL or PBRM-1 mutation. Instead, there were mutant alleles with higher frequency in two genes, specifically AHNAK and SRGAP3. Subsequently, single-cell exome sequencing was next performed on 20 cells from the tumor and 5 cells from adjacent normal tissue using a modified bioinformatics pipeline described in Hou et al. 2012 [5]. Two hundred and sixty somatic mutation sites in the coding region were first identified between cancer cells and the normal population. However, PCA analysis using the somatic mutation sites clustered three of the cancer cells with the normal adjacent tissue. PCA analysis did not reveal obvious cell subpopulations among the cancer cells. With removal of the three cancer cells from the initial group of 20 cells, 229 somatic mutations sites were identified between the two cell groups, 120 of them in coding regions [4].

To further define the intratumoral landscape, the frequency of mutant alleles among the cells was calculated. A small number of mutant genes, termed "mountain" genes, were present in a large fraction of cells (28 genes), while a large number of genes, termed "hill" genes, were present in only a few cells (66 genes). For the mountain genes, in addition to correlating with the whole exome sequencing of the sample taken from the tumor, AHNAK was noted to be of interest due to the gene's relation to HIF1A and the chromatin-remodeling process. For the hill genes, USP6, a ubiquitin-mediated proteolysis pathway (UMPP) gene, was found to be of interest secondary to role in tumorigenesis by promoting matrix metalloproteinases through NF-KB activation. Other genes of interest in biologically relevant pathways included: TUBB, CCKBR, and SULT1A1. The authors concluded that mutations in the hill genes appeared to be involved in cellular modification roles, and may be relevant in promoting progression once the cells have undergone mutations that initiate cancer [4].

The next paper by Kim et al. 2016, examined intratumoral heterogeneity (ITH) between a patient's ccRCC primary and metastatic lung lesion, and used their results to test a combinatorial regimen targeting two mutually exclusive pathways in the patient's metastatic cells. The basis for their approach considered the assumption that agents targeting a single pathway will terminate a subpopulation of tumor cells without affecting the remaining cells. Therefore, analyzing the tumor transcriptome at high resolution could possibly reveal drug-resistant subpopulations and guide additional therapeutic treatments that overcome resistance [6].

The patient included in the paper was initially diagnosed with T1NxM0 disease, had metastasectomy of a solitary pulmonary lesion within 1 year of initial diagnosis, and subsequent disease progression on sequential therapies, including pazopanib, everolimus, and high-dose interleukin-2 resulting in multi-organ metastasis. The authors used paired primary RCC (prRCC) and metastatic RCC (mRCC) patient-derived xenografts (PDX) in an attempt to better understand ITH and recreate the molecular landscape seen in the primary and metastatic tumors. Histopathologic and genomic analysis revealed consistent features between prRCC, mRCC, and their respective xenografts. Whole exome sequencing (WES) of prRCC and mRCC xenograft samples showed that 23.5% of somatic single nucleotide variants (SSNVs) were shared, including a mutation in the VHL gene, which may have been a founder event for this patient's tumor. Copy number variations detected from array comparative genomic hybridization (aCGH), revealed similar somatic copy number alterations (SCNAs) between the two groups, however 5q

amplifications were only seen in prRCC. Taken together, WES and aCGH major clones harboring driver mutations were seen in both groups and minor subclones were enriched in mRCC, indicating nonlinear, branching clonal evolution to support ITH for this patient instead of parallel evolution [6].

The authors next used scRNA-seq to profile single cells from both the mRCC ($n = 34$), PDX mRCC ($n = 36$), and PDX prRCC ($n = 46$). PCA analysis generated three clusters: normal kidney cortex, parental mRCC and the PDX mRCC, and the PDX prRCC cells. Unsupervised clustering demonstrated correlation of the average bulk cell populations to that of the averaged expression of single cells. PCA analysis also revealed a statistically significant molecular signature in the parental and PDX mRCC cells, specifically an epithelial-mesenchymal transition signature and Gene Ontology terms supporting "regulation of cell proliferation." The authors then used predefined gene sets involved with known drug targets to estimate drug sensitivity of the PDX cells. PDX-prRCC exhibited higher gene expression in C-met and PI3K/AKT pathways, while PDX-mRCC cells exhibited higher expression in EGFR, Src, and BRAF/MEK pathways. With these findings, Kim et al. 2016 hypothesized that the PDX metastatic cells would respond to afatinib and dasatinib, two inhibitors targeting EGFR and Src, respectively. This was further supported by the fact that only 14.7% of parental mRCC and 13.9% of PDX mRCC had neither activation of both pathways in subpopulation analysis. Subsequent in vitro and in vivo testing of PDX mRCC cells proved that combining afatinib and dasatinib was superior to single agent treatment [6]. Overall, Kim et al. 2016 proved that to overcome resistance created by ITH, SCS can be used in the translational design of personalized therapeutic strategies.

The final research to be presented in the RCC section comes from Li et al. 2017 and their work using SCS to identify genes driving RCC cancer stem cells (CSCs). Given that CD133 is a common CSC marker [7], the authors sorted 10 CD133+ RCC cells, 10 CD133- RCC cells, and 10 normal renal cells for single-cell WES from a 57-year-old male patient with T3aN1M0 RCC. Both the RCC and adjacent normal tissue were used for bulk tissue WES. One hundred and sixty SNVs were seen in the cancer bulk tissue WES and 297 SNVs in the 20 RCC cells. Commonly mutated genes in RCC (VHL, BAP1) were found to have variations in both bulk tissue and single cells, strengthening the reliability of the WES analysis. PCA analysis of all somatic mutations separated normal renal, CD133- and CD133+ cells, proving that the isolated CD133- and CD133+ cells were tumor cells. After constructing a neighbor-joining tree for the 30 cells, Li et al. 2017 found that the evolutionary distance was larger between CD133+ to normal cells compared to CD133- cells to normal cells, indicating that CD133+ cells likely originated from cancer cells and not normal cells [8].

The authors then found three missense mutations that were unique to the CD133+ cells in the following genes: KCP, LOC44040, and LOC440563. In addition to these genes, the authors found that 29 mutated genes were detected in 3 or more CD133+ RCC cells, KCP being one of them. Of these 29 genes, 18 were listed in the TCGA database with a combined frequency in the TCGA patients of less than 2%. Next, the authors used CRISPR-Cas9 to assess the tumor propagating potential of the 18 genes (including KCP) plus LOC44040 and LOC440563 in RCC 786-O cells. Mutations in the KPC gene had the highest cancer spherogenesis, and LOC44040 in combination with KCP was the most effective in enhancing spherogenesis. In addition, 786-O and 769-P RCC cells with triple mutations (KCP/LOC44040/LOC440563) had a significant increase in spherogenesis compared to wild-type cells. Engraftment of these triple mutated cell lines in mouse xenograft models had more mice developing tumors and higher enrichment of CSCs compared to wild-type cells. The last part of their analysis assessed the mutation rates of these three genes in 57 RCC patients using Sanger sequencing. They found that patients with triple mutations had shorter disease-free time after primary resection [8]. In conclusion, Li et al. 2017 was able to better characterize

RCC CSCs, showing that these CSCs likely originate from RCC cells, and KCP, LOC440563, and LOC44040 likely drive CSC survival and progression.

13.3 Bladder Cancer

The first paper in the bladder cancer section illustrates how SCS can be applied to further support the development of a diagnostic test in bladder cancer. Chen et al. 2018 published their results on a microfluidic approach on harvesting intact urinary-exfoliated cells (UETCs) for the detection of bladder cancer. To accurately identify these UETCs, they developed a quantitative immunoassay using the oncoproteins CK20 and CD44v6 antigen. The authors subsequently enrolled 79 patients with bladder cancer and 43 age-matched controls, and were able to detect a significantly greater number of UETCs in the cancer group compared to the control group [53.3 (10.7–1001.9) vs. 0.0 (0–3.0) UETCs/10 mL; $P < 0.0001$]. A stratified tenfold cross validation was used to assess bladder cancer detection, demonstrating 89.8% sensitivity (95% CI, 71.5–86.4%) and 71.5% specificity (95% CI, 59.7–83.3%) [9].

Of relevance to this review article, the authors performed single-cell whole-genome sequencing on captured UETCs, to confirm that their immunoassay was collecting tumor cells. Chen et al. 2018 collected from 3 patients with bladder cancer, a total of 15 single cells comprising of 12 immunofluorescence identified UETCs and 3 normal urothelial cells that were referenced control cells. The authors used a low-pass sequencing strategy to assess the copy number alterations (CNAs) of the single cells. They found that 11/12 of the UETCs possessed an unstable genome with CNAs while the 3 control cells were found to be diploid, thus confirming that the immunoassay is able to identify individual cancer cells from primary bladder tumors [9].

Similar to the work by Li et al. 2017 in RCC, Yang et al. 2017 examined the origination and genetic drivers of human bladder cancer stem cells (BCSCs) [10]. The authors conducted sequencing on 59 cells including bladder epithelial stem cells (BESCs, pan-CK^+CD44^+), bladder epithelial non-stem cells (BENSCs, pan-CK^+CD44^-), BCSCs ($CD31^-CD45^-CD44^+$), and bladder cancer non-stem cells (BCNSCs, $CD31^-CD45^-CD44^-$) [11] from three bladder cancer specimens. Analysis of these four cell types showed that the authors were able to establish tumors in NOD/SCID mice using BCSCs and BCNSCs, but not the other two cell types. Furthermore, BCSCs and BESCs in comparison to BCNCs and BENSCs, had upregulated expression of stem cell-related genes, and exhibited better spherogenesis and self-renewal properties. In serial transplantation assays, BCSCs were better at initiating tumor formation when compared to BCNSCs, further validating the identity of the four cells types being examined [10].

Yang et al. 2017 next isolated 59 cells from three bladder cancer specimens and subsequently performed exome sequencing and phylogenetic analysis using the modified neighbor-joining method to assess cell clonality. They found that BCSCs likely originated from mutations in BESCs or BCNSCs. Four hundred and six non-synonymous mutant genes were identified, with BCSCs showing a higher frequency of non-synonymous mutant genes when compared to BCNSCs. 21 out of 46 genes were distributed among five functional pathways: cell differentiation and self-renewal (FAT4 and GPRC5A) cell-cycle regulation (TP53, ATM, and CREBBP, STAG2), transcription regulation (TP53, BRF1, PAWR, ERCC2, MKL1, SIN3A, and ETS1), chromatin remodeling (ARID1A, CREBBP and MLL2). To test the function of BCSCs mutations, 15 genes with a mutational rate >50% were introduced individually using CRISPR/Cas9 into BCNSCs taken from primary bladder tumors. MLL2 was the only gene that slightly enhanced spheroid formation. However, further combinatorial and stepwise introduction of mutations proved that a trio of mutations consisting of $MLL2^+ARID1A^+GPRC5A^+$ brought the sphere forming ability of BCNSCs on par with BCSCs. In addition, BCNSCs with these three mutations were better at initiating bladder cancer and serial tumor formation compared to wild-type BCNSCs

in limiting dilution xenograft and serial transplantation assays. Overall, the authors were able to provide information on the origin of BCSCs, key gene mutations in these cells, and which mutations promote stemness [10].

Li et al. 2012 performed single-cell exome sequencing of a patient's muscle-invasive transitional cell carcinoma (MI-TCC) to reconstruct the developmental history of the tumor and underlying subclones. By understanding the timing of key mutations in the patient's tumor, the authors sought to discover genes driving carcinogenesis and subclone development. Forty-four single cells from the tumor and 11 from normal adjacent tissue were sequenced, in addition to whole exome sequencing of the bulk tissue used to obtain the individual cells. Four hundred and forty-three somatic mutations were identified from single-cell analysis. The authors randomly selected 17 genes to verify with PCR-Sanger capillary sequencing with 100% of predicted genes being confirmed. None of these mutations appeared in RB1 or TP53, and the majority of mutations were C:G > T:A. Exome sequencing of DNA from the bulk tissue could only identify 134 of the 443 somatic mutations (30.25%) found in SCS. The authors noted significant copy number variations and loss of heterozygosity in chromosomes 9 and 11 that were consistent with MI-TCC [12].

The authors next applied population genetics to analyze bladder cancer development by deriving the somatic mutant allele frequency spectrum (SMAFS) between normal and bladder tumor cells. The majority of tumor cells had a peak in SMAFS around 50%, indicating that TCC likely originated from a single cell. PCA analysis using identified mutations was able to separate the normal cells from tumor cells. The tumor cells had significant diversity across the principal vectors supporting a heterogenous makeup of the tumor. However, the authors were able to cluster the tumor cells into three subclones (A, B, and C) using mutational heat map changes (146 non-synonymous mutations in 113 genes). Almost all three subclones had mutations in 22 genes, which were thought to be driver genes that initiated tumorigenesis from a common ancestral cell. Clones B and C were estimated to have emerged late in the tumor history, but surprisingly made up a larger portion of the tumor than expected (35%). The authors hypothesized that these two subclones had conferred additional growth advantages and were undergoing positive selection in comparison to Clone A. The authors used conventional exome sequencing in a cohort of 99 TCC patients and found that 4 of their 22 genes had non-silent mutations in at least three of these patients: CFTR, NIPBL, ASTN1, and DHX57. The authors also found three recurrent mutations in this 99 patient cohort that were unique to clone B and C, potentially serving as driving genes for these two subclones: ATM in subclone B and COL6A3, KIAA1958 in subclone C. Taken together, Li et al. were able to provide significant insight on the monoclonal origin of their patient's tumor supporting a clonal evolution model, and the subsequent subclonal development and genetic changes [12].

Zhang et al. 2016 used single-cell RNA-seq on 67 tumor cells and 7 normal cells to characterize the heterogenous gene expression profile of a patient with localized squamous cell carcinoma. PCA analysis indicated no normal cell populations amongst the tumor cells. However, no specific subclonal population was identified. NOISeq identified differentially expressed genes in significant pathways between normal versus tumor cells (p53 signaling, cell-cycle pathways). Between individual tumor cells, the authors calculated individual gene coefficient of variation (CV) as a way to define intratumor heterogeneity. They analyzed the 100 most variable and 100 most stably expressed gene sets. The variably expressed gene group had enrichment in many cell-cycle related genes, specifically six genes from the MAPK pathway (RPS6KA1, RAC2, CACNG4, CACNA1E, CACNA1H, and MAPKAPK5), while the stably expressed genes had enrichment in the expected housekeeping genes. The authors also looked at co-expression profiles between 5530 highly variable genes at a system level (RPM > 10, SD > 100) using Weighted Gene Co-Expression Network Analysis (WGCNA). Hub-gene-network-analysis revealed several significant "hub genes:" GCC2, OR9Q1,

LINC00189, NKD1, POU2F3. Of these genes, POU2F3 was found to play a significant role in squamous epithelial stratification and is reported to have tumor suppressor function in cervical cancer. Overall, the authors further illustrated how SCS can define intratumor heterogeneity and identify significant genetic changes that could lead to further functional validation tests and therapeutic studies [13].

The final paper in the bladder cancer section published by Tanaka et al. 2018 examined platinum-based therapy's effect on intratumoral heterogeneity in vitro using scRNA-seq, in order to elucidate a potential mechanism behind platinum resistance. The authors used urothelial cell line 5637 cells, passaged 1–2 times per week in medium containing platinum-based cisplatin (CDDP) over 6 months, increasing the CDDP gradually to 3 μmol/L. Parental cells were continually cultured and passaged without CDDP. SCS analysis occurred 3 months after CDDP was discontinued. When comparing gene expression between the two cell lines, 12 genes were found to be consistently downregulated in platinum resistant cells: COX7B, MT1E, LGALS1, KRT17, EIF3E, TMA7, ARL6IP1, HES1, UQCR10, MORF4L1, CDKN3, and PSMD10. Transfected siRNAs against these 12 genes showed that 4 genes (COX7B, MT1E, LGALS1, and KRT17) caused the highest CDDP resistance. However, a Kaplan–Meier analysis of TCGA bladder cancer cohort revealed that out of these four genes only COX7B was associated with worse mortality and predicted poor prognosis. Therefore, COX7B could potentially be used as a marker for platinum resistance, and overexpression of COX7B re-sensitized cell lines to CDDP. During their in vitro analysis, the authors were able to identify a subclone of platinum-naive cells with low-COX7B that behaved as if they already had acquired platinum resistance. Using fluorescent-activated cell sorting (FACS) the authors found that CD63 could sort this subclone from bulk cancer cells. In conclusion, the work by Tanaka et al. 2018 highlights how scRNA-seq can be used to examine gene expression that promotes chemotherapy resistance, which may be used to guide future treatments in bladder cancer [14].

13.4 Prostate Cancer

Identifying the cellular origins of benign prostatic hyperplasia and prostate cancer first requires an understanding of the identity and function of each cell type within the prostate. Defining the cellular origins of prostate organ cells is not only important in understanding the diseases of prostate cancer and BPH, but also important in developing novel therapies against these conditions. To properly define human prostate cellular anatomy and create a baseline for understanding the cellular origins of disease, Henry et al. 2018 performed single-cell RNA sequencing (scRNA-seq) on approximately 98,000 cells from five young adult human prostates [15]. Initially, scRNA-seq transcriptome analysis was used to broadly define cell types either in the epithelial, stromal, or neuroendocrine lineage. The authors found a molecular identity for five epithelial and two stromal cell types. Following transcriptome analysis, the authors used the scRNA-seq dataset to identify cell surface markers more specific to each individual prostate cell subtype to enhance the enrichment yield by FACS. The transcriptomic signature was used to describe the function and spatial location of the cell types.

Furthermore, their data led to the discovery of two previously unknown epithelial cell types marked by a high expression of SCGB1A1 and KRT13. Prostate SCGB1A1$^+$ cells are similar in morphology and transcriptomic profile to Clara cells (or club cells) of the lungs. Prostate club cells may function to enrich immunomodulatory programs and further testing of their function in the prostate or prostatic urethra needs to be done. Prostate KRT13$^+$ cells are similar in morphology and transcriptomic profile to hillock basal cells of the lung. Prostate hillock cells are found rarely in adult prostates but in high density in fetal prostate. KRT13$^+$ cells are also enriched in localized prostate tumors, and in stem-like cells that display androgen resistance and a capacity for branching morphogenesis. Two of the top genes associated

with the KRT13^{+} subtype (AKR1C1 and AKR1C2) are part of the androgen metabolism pathway. Findings from this study show that hillock cell type may be responsible for the KRT13 expression that was hypothesized to originate from basal and luminal cell types. Overall, Henry et al. 2018 provide a better understanding of how to best characterize prostatic cell subtypes to allow for further research to elucidate the cellular mechanisms of prostatic disease [15].

One of the major challenges in the management of prostate cancer addressed by Horning et al. 2018 is the plasticity and heterogeneity of prostate cancer cells with regard to androgen dependence. The stepwise model and punctuated models are two theories that potentially explain how prostate cancer cells develop resistance to androgen deprivation therapy (ADT). The stepwise model suggests that a single cell or clone develops a mutation that favors ADT resistance, and that clone subsequently takes over under ADT. The punctuated model suggests that the progression of cancer cells and clones is nonlinear. Instead, there are likely multiple subclones with their own molecular alterations, each with a different degree of androgen sensitivity. Eventually, through ADT selection and clonal expansion, an androgen insensitive subclone grows and becomes the predominant cell type. Horning et al. 2018 attempted to stratify these cell subpopulations using scRNA transcriptome profiling of 144 single LNCaP prostate cancer cells first treated with ADT, and then either treated or untreated with androgens after cell-cycle synchronization. This method identified 397 differentially expressed genes in eight potential subpopulations of LNCaP cells, revealing a previously unknown level of cellular heterogeneity among the LNCaP cells [16].

Of note, one of the subpopulations exhibited stem-like features (slow doubling rate, increased spherogenesis) and enhanced growth of this subpopulation showed increased expression in ten cell-cycle genes: CCN, HMMR, CDC20, CCNB2, DLGAP5, CENPF, PLK1, CENPE, MKI67, PTTG1. The authors used in silico RNA-seq data from the TCGA to show that these genes are linked to prostate recurrence. Overall, the results from Horning et al. 2018 are more consistent with a punctuated model, in which pre-existing subpopulations have stem cell like features that promote resistance to ADT. Furthermore, the ten gene panel discovered by the authors may potentially be used for subpopulation stratification, and may identify tumors at high risk for recurrence if validated in a large-scale study [16].

In order to apply precision medicine for cancer patients with metastatic disease, a tissue sample is required that is sometimes difficult to obtain. Circulating tumor cells in the blood (CTCs) circumvents this issue, and the remainder of the prostate cancer section will focus on these cells. CTCs mediate the metastatic spread of many solid tumors and when found are generally a poor prognostic indicator. The difficulty of isolating and analyzing individual CTCs has limited progress in understanding the molecular expression of these cells. Cann et al. 2012 published their results on the use of MagSweeper to harvest individual CTCs without contaminating leukocytes. The authors performed single-cell transcriptome analysis on CTCs isolated by MagSweeper to confirm their identity. All but one of the cells were confirmed to be CTCs given their expression of the androgen receptor and downstream target genes (KLK3, TMPRSS2). Pathway analysis further confirmed activation of the AR pathways, cell-cycle regulation, and mitotic spindle genes. As expected, mitotic spindle genes had increased expression in patients on taxane therapy. Several transcripts associated with aggressiveness in localized prostate cancer were seen in the CTCs (PLK-1, TOPA). Furthermore, the authors noted that several transcripts that were upregulated in CTCs may serve as potential targets (BIRC5, SPINK1). In conclusion, Cann et al. 2012 highlight the effectiveness of MagSweeper to capture individual CTCs that offers further potential to understand the genetic makeup of these cells and potential druggable targets [17].

Similar to Cann et al. 2012, Lohr et al. 2014 used MagSweeper and developed a set of experimental and analytical protocols for the sequencing of whole exomes of prostate CTCs and

confidential calling of SSNVs. They compared a patient's CTC sequencing to their primary tumor, sequencing nine spatially distinct foci from the primary tumor. Ten SSNVs were found in all primary foci and the CTCs (including TP53), suggesting a single ancestor initiated the patient's cancer with divergent evolution. Fifty-six mutations were present in both the primary tumors (any foci) and CTCs. Together, both of these findings by Lohr et al. 2014 show how CTCs can be used to understand tumor evolution [18].

CTCs provide a noninvasive way to assess chemotherapy resistance during therapy and to monitor the genetic changes that are driven by treatment. Dago et al. 2014 used the High Definition-CTC method for the identification and isolation of CTCs in a patient with metastatic prostate cancer being treated with chemotherapy and abiraterone at four different time points [19]. To correlate protein expression data with genome-wide CNV alterations in 41 CTCs taken at each time point, the authors used the protocols established by Navin et al. 2011 [20] and Baslan et al. 2012 [21]. Bulk metastatic biopsy prior to therapy provided the root CNV profile that subsequent CTC profiles were compared against. The first two draws were taken after being treated with ADT (leuprolide acetate), in which a subpopulation (clone A) was found to be a descendant of the initial bulk metastatic tumor with the development of high-copy AR amplification. The cells in this clone likely evolved to overexpress the androgen receptor protein due to pressure from ADT. Clone A was surprisingly found to have few changes between the first two treatments of ADT [19].

However, the patient was switched to abiraterone acetate and at draw 3 the androgen-dependent AR positive cells were mostly absent. These cells were replaced by AR negative pseudodiploid cells. During draw 4, AR+ cells had returned as demonstrated by the dominant subpopulation clone C. Clone C had first become evident during draw 3, but was not the dominant clone at that time. The authors concluded that clone C was selected as a drug-resistant subclone from one of the initially depleted metastatic sites. Of note, a complex rearrangement of 8q causing amplification in MYC for the cells in the later draws may be related to the re-emergence of AR protein expression. c-MYC expression is related to androgen independent growth [22]. Thus, targeting c-MYC and AR at the time point of draw 4 may have delayed progression or prevented the development of resistance to abiraterone acetate. Overall, Dago et al. 2014 demonstrate how "fluid biopsies" looking at CTCs may be able to guide treatments and support a minimally invasive, precision based approach to treat metastatic disease [19]. The PROPHECY study by Armstrong et al. 2019 further supports the "fluid biopsy" approach demonstrated by Dago et al. 2014, in which the authors used a qPCR CTC assay to detect the androgen receptor splice variant (AR-V7). The authors found that AR-V7 detection in CTCs is associated with shorter PFS and OS in men with metastatic castrate resistant prostate cancer prior to receiving abiraterone or enzalutamide [23].

The last paper in this chapter by Miyamoto et al. 2015 also examines CTC heterogeneity and changes in gene expression after ADT [24]. The authors performed single-cell RNA-seq on 77 CTCs isolated by microfluidic enrichment from 13 patients. In addition to the CTCs, the authors obtained bulk transcriptomes from the primary tumors of 12 patients, 30 single cells taken from different prostate cancer lines and 5 patient-derived leukocyte controls. Unsupervised hierarchical clustering analysis was able to separate the CTCs, primary tumors, and cell lines. Gene markers of prostate lineage were examined (epithelial, mesenchymal, and stem cells). The CTCs had upregulation in epithelial markers, but not mesenchymal markers when compared to primary tumors and prostate cell lines. Sixty percent of CTCs had increased expression in three stem cell markers: ALDH7A1, CD44, and KLF4. The authors found 711 genes that were upregulated in CTCs compared to primary tumors. One gene specifically, HSP90AA1, is known to regulate the AR receptor [25]. The authors used the Pathway Interaction Database to identify pathways updated in CTCs and found 21 pathways, mostly

related to cell adhesion, growth factor, and hormone signaling. Analysis of mRNA splice variants revealed heterogeneous and complex patterns of AR splice-variant expression among the CTCs [24].

The authors next performed retrospective differential analysis in CTCs to identify potential molecular mechanisms of resistance to enzalutamide. The authors compared 41 CTCs from 8 patients without enzalutamide treatment (Group A) to 36 CTCs from 5 patients with evidence of progression of disease on enzalutamide (Group B). Disease progression was determined by either a rising PSA or radiographically. A gene set enrichment analysis (GSEA) showed significant enrichment in the Wnt signaling group for the group B CTCs. This signaling pathway controls multiple downstream regulators of cell proliferation, survival, and motility: specifically, RAC1, RHOA, and CDC42. Surprisingly, AR abnormalities were not significantly increased amongst group B when compared to group A. The authors next ectopically expressed WNT ligands (WNT4, WNT5A, WNT7B, WNT11) in LNCaP androgen-sensitive cell lines. Survival of these AR positive cells was increased in the presence of enzalutamide by increased expression of all the ligands tested, and most significantly for WNT5A. In addition, the authors found that WNT5A expression increased in untreated LNCaP cells with the addition of enzalutamide to cell medium. Overall, similar to the previous papers publishing results using CTCs, Miyamoto et al. 2015 provided additional support to the utility of using CTCs to understand the molecular mechanisms of therapy acquired resistance and disease progression.

13.5 Conclusion

The studies listed in this chapter illustrate many translational applications of SCS in GU malignancies, including diagnostic, prognostic, and treatment-related approaches. Collectively, these studies provide evidence for clonal evolution from a single ancestral cell, and proof that cancer stem cells play a vital role in tumorigenesis. The ability of SCS to resolve intratumor heterogeneity and better define the genomic landscape of tumors and CTCs will be fundamental in the new era of precision-based care. The authors' research highlighted in this chapter represents the first critical step toward better defining GU cancers at the cellular level. Future analysis with a larger number of cells and improved SCS technology will be the next step to successfully continue moving the field forward.

References

1. Common cancer sites—cancer stat facts. In: SEER [Internet]. [cited 16 Sep 2019]. https://seer.cancer.gov/statfacts/html/common.html
2. Morash M, Mitchell H, Beltran H, Elemento O, Pathak J (2018) The role of next-generation sequencing in precision medicine: a review of outcomes in oncology. J Pers Med 8(3):pii: E30. https://doi.org/10.3390/jpm8030030
3. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR et al (2018) Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361:594–599
4. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L et al (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148:886–895
5. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X et al (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 148:873–885
6. Kim K-T, Lee HW, Lee H-O, Song HJ, Jeong DE, Shin S et al (2016) Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. Genome Biol 17:80
7. Florek M, Haase M, Marzesco A-M, Freund D, Ehninger G, Huttner WB et al (2005) Prominin-1/CD133, a neural and hematopoietic stem cell marker, is expressed in adult human differentiated cells and certain types of kidney cancer. Cell Tissue Res 319:15–26
8. Li C, Wu S, Yang Z, Zhang X, Zheng Q, Lin L et al (2017) Single-cell exome sequencing identifies mutations in KCP, LOC440040, and LOC440563 as drivers in renal cell carcinoma stem cells. Cell Res 27:590–593
9. Chen A, Fu G, Xu Z, Sun Y, Chen X, Cheng KS et al (2018) Detection of urothelial bladder carcinoma via microfluidic immunoassay and single-cell DNA copy-number alteration analysis of captured urinary-exfoliated tumor cells. Cancer Res 78:4073–4085

10. Yang Z, Li C, Fan Z, Liu H, Zhang X, Cai Z et al (2017) Single-cell sequencing reveals variants in ARID1A, GPRC5A and MLL2 driving self-renewal of human bladder cancer stem cells. Eur Urol 71:8–12
11. Chan KS, Espinosa I, Chao M, Wong D, Ailles L, Diehn M et al (2009) Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. Proc Natl Acad Sci U S A 106:14016–14021
12. Li Y, Xu X, Song L, Hou Y, Li Z, Tsang S et al (2012) Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. Gigascience 1:12
13. Zhang X, Zhang M, Hou Y, Xu L, Li W, Zou Z et al (2016) Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder. Oncotarget 7:66069–66076
14. Tanaka N, Katayama S, Reddy A, Nishimura K, Niwa N, Hongo H et al (2018) Single-cell RNA-seq analysis reveals the platinum resistance gene COX7B and the surrogate marker CD63. Cancer Med 7:6193–6204
15. Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J et al (2018) A cellular anatomy of the normal adult human prostate and prostatic urethra. Cell Rep 25:3530–3542.e5
16. Horning AM, Wang Y, Lin C-K, Louie AD, Jadhav RR, Hung C-N et al (2018) Single-cell RNA-seq reveals a subpopulation of prostate cancer cells with enhanced cell-cycle-related transcription and attenuated androgen response. Cancer Res 78:853–864
17. Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat M et al (2012) mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. PLoS One 7:e49144
18. Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, Cruz-Gordillo P et al (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. Nat Biotechnol 32:479–484
19. Dago AE, Stepansky A, Carlsson A, Luttgen M, Kendall J, Baslan T et al (2014) Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. PLoS One 9: e101777
20. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J et al (2011) Tumour evolution inferred by single-cell sequencing. Nature 472:90–94
21. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A et al (2012) Genome-wide copy number analysis of single cells. Nat Protoc 7:1024–1041
22. Li B, Simon MC (2013) Molecular pathways: targeting MYC-induced metabolic reprogramming and oncogenic stress in cancer. Clin Cancer Res 19:5835–5841
23. Armstrong AJ, Halabi S, Luo J, Nanus DM, Giannakakou P, Szmulewitz RZ et al (2019) Prospective multicenter validation of androgen receptor splice variant 7 and hormone therapy resistance in high-risk castration-resistant prostate cancer: the PROPHECY study. J Clin Oncol 37:1120–1129
24. Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT et al (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science 349:1351–1356
25. Trepel J, Mollapour M, Giaccone G, Neckers L (2010) Targeting the dynamic HSP90 complex in cancer. Nat Rev Cancer 10:537–549

Neal Murphy is an Assistant Professor at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell where he is currently practicing as a hospitalist. He graduated from Cornell University and received his MD from the Lewis Katz School of Medicine at Temple University. He subsequently completed his internal medicine training at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell. His current research is focused on identifying predictive biomarkers in genitourinary cancers. Dr Murphy will begin hematology/oncology fellowship July 2020 as part of the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell program.

Pratik Shah graduated from the dual-degree BA/MD program at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell. He will begin internal medicine residency at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell program in July 2020. He is interested in pursuing a career in hematology/oncoloy and pursuing further research in solid tumor maligancies.

Andrew Shih is an Assistant Investigator at The Feinstein Institute for Medical Research. He completed his undergraduate study at Duke University obtaining a BSE in Biomedical and Electrical Engineering. He subsequently went on to complete a PhD at the University of Pennsylvania in Bioengineering, and a postdoctoral fellowship at the Feinstein Institute in bioinformatics and human genetics.His career has focused on using computational techniques to analyze large and complex datasets to reveal biological insights. Dr Shih's early work focused on looking at cellular consequences of perturbations in the ErbB family using protein simulations, but he transitioned into bioinformatics as sequencing technology matured. He has substantial experience in analyzing and interpreting many of the data types in the -omics world (e.g., DNA-seq, RNA-seq, ATAC-seq, single-cell RNA-seq, etc.) across multiple tissue types and biological conditions. He also has specific expertise with endometriosis, in collaboration with Dr Peter Gregersen and Dr Christine Metz, in which they have reported potential biomarkers for detection of endometriosis noninvasively by using menstrual effluent. In addition, he actively participates in the Systems Biology Group of the Accelerating Medicine Partnership (AMP) with Dr. Gregersen, where

bioinformatics methods, designs, and results of RNA-Seq, scRNA-Seq, and ATAC-Seq experiments applied to rheumatoid arthritis are discussed.

Houman Khalili is Senior Research Assistant at the Feinstein Institutes for Medical Research, specializing in high-throughput "-omics" technologies.

Anthony Liew is Senior Research Assistant at the Feinstein Institutes for Medical Research, specializing in high-throughput "-omics" technologies.

Xinhua Zhu is a hematologist/oncologist at the Northwell Health Cancer Institute and an Assistant Professor at the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell. He initially obtained his MD at the Medical College of SooChow University, China, MS and PhD at the Shanghai Second Medical University. He went on to complete postdoctoral fellowships at NYU and Memorial Sloan-Kettering Cancer Center where he then proceeded to complete internal medicine residency at Mount Sinai School of Medicine and hematology/oncology fellowship at NYU. As a PhD student, Dr Zhu discovered the effect of arsenic trioxide on growth inhibition for malignant lymphocytes with in vitro cell lines in comparison to primary cells from patients. His work has also demonstrated the novel mechanism of treatment with histone deacetylase inhibitors in acute promyelocytic leukemia through the use of animal models. As a postdoctoral Fellow at NYU School of Medicine, he carried out transgenic, knockout and protein work for Tamm-Horsfall Protein (THP) gene function studies, which were associated with the finding of the role of THP in pathogenesis of urinary tract infection and carcinogenesis of bladder cancer. As a postdoctoral Fellow at Memorial Sloan-Kettering Cancer Center, he discovered that the non-catalytic cyclin A-CDK2 could enhance the proteolysis of p27 which is responsible for the regulation of cell cycle from G1 to S phase, and that hDM2 regulates p21 proteolysis in the S-Phase. As a clinical Fellow in Hematology and Oncology at NYU he found that the chronic inflammation created by anti-androgen therapy may be associated with the development of castration-resistant prostate cancer. He also discovered potential targets of WNT-signaling pathway to inhibit the activity of prostate cancer stem cells for potentially treating patients with castration-resistant prostate cancer. Dr Zhu has completed two R01 projects, having collaborated with Dr Xue-Ru Wu and Dr Andrew Koff with related publications in prestigious journals. Clinically he completed his training in both China and United States, and is a board certified physician in Internal Medicine, Hematology and Medical Oncology. Dr Zhu's training as a physician scientist allows him to bridge the gap between basic scientific studies and clinical application. As a postdoctoral fellow at NYU and Memorial Sloan-Kettering Cancer Center, he has had many experiences in the guidance of PhD candidates and junior postdoctoral fellows. He continues to work on several translational studies as a principle investigator, including a "Phase I study of the safety and tolerability of protein tyrosine phosphatase (PTP1B) inhibitor for metastatic breast cancer patients," "Identifying biomarkers and potential targets through investigation of first line cisplatin sensitive and resistant patients with bladder cancer," and "Identifying molecular targets and prognostic biomarkers for metastatic progression in stage I and stage II clear cell renal cell carcinomas."

Annette Lee is an Associate Professor at the Robert S. Boas Center for Genomics & Human Genetics and Dean of the Elmezzi Graduate School of Molecular Medicine at the Feinstein Institutes for Medical Research. Dr Annette Lee graduated from Northeastern University and received her PhD from The Rockefeller University. She joined The Center for Genomics and Human Genetics as an Assistant Professor in 2001 and worked on the genetics of autoimmune diseases such as rheumatoid arthritis, lupus and inflammatory bowel disease. More recently as an associate investigator and director of the Laboratory of Translational Genetics, she has established her own area of research investigating the genetics of cancer—including chronic lymphocytic leukemia, breast, and ovarian cancers.

Dr. Lee has coauthored almost 100 peer-reviewed papers and directed the sample genotyping for several large autoimmune genome-wide association studies to identify risk genes for rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease (Crohn's disease and ulcerative colitis), scleroderma, alopecia areata, IgA deficiency, myasthenia gravis, and myositis. All of these studies have either resulted in publications in high profile journals or are in the process of being analyzed for manuscript submission. As a follow up to genome-wide studies, she has also supervised the selection and development of targeted genetic studies which range from a few variants to dense mapping of over 12,000 SNPs for several autoimmune diseases, Alzheimer's, Parkinson's, diabetic nephropathy, and asthma. Several key genetic variants associated with human diseases have been identified as a direct result of these studies. More recently she has begun to study biomarkers of breast, ovarian cancers, and genitourinary cancers.

PI3K Isoform-Selective Inhibitors in Cancer

14

Leslie Duncan, Chloe Shay, and Yong Teng

Abstract

PI3K inhibitors are a common area of research in finding a successful treatment of cancer. The PI3K pathway is important for cell growth, apoptosis, cell metabolism, cell survival, and a multitude of other functions. There are multiple isoforms of PI3K that can be broken down into three categories: class I, II, and III. Each isoform has at least one subunit that helps with the functionality of the isoform. Mutations found in the PI3K isoforms are commonly seen in many different types of cancer and the use of inhibitors is being tested to stop the cell survival of cancer cells. Individual PI3K inhibitors have shown some inhibition of the pathway; however, there is room for improvement. To better treat cancer, PI3K inhibitors are being combined with other pathway inhibitors. These combination therapies have shown better results with cancer treatments. Both the monotherapy and dual therapy treatments are still currently being studied and data collected to better understand cancer and other treatment options.

Keywords

PI3K · Isoforms · Cancer · Mutations · Inhibitor · Anticancer

Abbreviations

Akt	Protein kinase B
ALK	Anaplastic lymphoma kinase
Bad	BCL2 associated against cell death
EGFR	Epidermal growth factor receptor
HER-2	Human epidermal growth factor receptor 2
mTOR	Mechanistic target of rapamycin
mTORC1	Mechanistic target of rapamycin complex 1
mTORC2	Mechanistic target of rapamycin complex 2
PI3K	Phosphoinositide 3-kinase
PIP2	Phosphatidylinositol 4,5-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate

L. Duncan
Department of Oral Biology and Diagnostic Sciences, Dental College of Georgia, Augusta University, Augusta, GA, USA

Department of Biology, College of Science and Mathematics, Augusta University, Augusta, GA, USA

C. Shay
Department of Pediatrics, Emory Children's Center, Emory University, Atlanta, GA, USA

Y. Teng (✉)
Department of Oral Biology and Diagnostic Sciences, Dental College of Georgia, Augusta University, Augusta, GA, USA

Georgia Cancer Center, Medical College of Georgia, Augusta University, Augusta, GA, USA
e-mail: yteng@augusta.edu

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_14

PtdIns	Phosphatidylinositol
PtdIns (3) P	Phosphatidylinositol 3-phosphate
PtdIns (4) P	Phosphatidylinositol 4-phosphate
PTEN	Phosphatase tensin homologue
Rheb	Ras homolog enriched in brain
Tor	Target of rapamycin
Wnt	Wingless type MMTV integration site family

14.1 Introduction

Phosphatidylinositol 3-kinase, otherwise known as PI3K, is a lipid based enzyme, responsible for phosphorylating molecules when activated by a growth factor [1–6]. The activation of the PI3K in turn initiates a signaling pathway in different cells. The PI3K pathway is important in cell growth, cell survival, apoptosis, metabolism, and a variety of other cellular functions [2, 7–10]. The pathway starts with a growth factor binding to a tyrosine kinase receptor [11]. This binding stimulates many signaling pathways, including the PI3K pathway. The binding of the growth factor to the receptor activates PI3K which catalyzes the 3′ hydroxyl group on a phosphatidylinositol, or PtdIns, specifically phosphatidylinositol (3,4,5)-triphosphate, or PIP3 [1, 11]. The now phosphorylated PIP3 binds to protein kinase B, otherwise known as Akt, initiating a conformational change. Akt is then activated by being phosphorylated by 3′ phosphoinositide dependent protein kinase [11]. The now activated Akt further phosphorylates different substrates, one being BAD otherwise known as, BCL2 associated against cell death. Bad normally promotes apoptosis; however, when it is inhibited, apoptosis is suppressed [11–18]. Akt also inhibits cell cycle arrest by activating a G protein that is a RAS homolog enriched in the brain, known as Rheb [11, 13, 17, 19–21]. The activation of Rheb then leads to the activation of a cell growth regulator target of rapamycin, known as Tor [11, 13, 17, 21]. This activation stimulates cell growth. The PI3K signaling pathway ultimately promotes cell survival and cell growth [2, 7, 8, 11, 13, 17, 21].

The PI3K signaling pathway is a focus of many research topics and is important in cancer development. Mutations in the pathway increase the activity of the protein kinase Akt increasing cell proliferation and survival [22]. The increase concentration of Akt inhibits Bad and Rheb decreasing apoptosis and increasing cell growth [2, 7, 8]. This increased Akt activity is found frequently in many types of cancer cells including breast cancer, prostate cancer, head and neck cancer, leukemia, and many others. Research suggests the inhibition of the PI3K signaling pathway may be a successful treatment for cancer by suppressing the cancer cell proliferation. Isoforms of PI3K have the same functionality with a difference in physical makeup. PI3K isoforms are named based on their substrate preference and structure. There are three classes of isoforms: class I, II, and III. In mammals the first class is further subdivided into class IA and IB [2, 7, 8, 23]. Class IA consists of three types of catalytic subunits: p110α, p110β, and p110δ and a regulatory subunit p85 as seen in Table 14.1 [24]. The isoforms are made up of a catalytic and regulatory subunit making a heterodimer. Class IB only has one catalytic subunit, p110γ, and two regulatory subunit p101 and p87 [2]. Class IA and class IB both possess many different domains each with a responsibility, including the phosphorylation of phosphatidylinositol 4,5-biphosphate or PIP2 [2]. Class II isoforms are still being researched as little is known about them. It is known that there are three catalytic subunits: PI3K-C2α, PI3K-C2β, and PI3K-C2δ and no regulatory subunits as seen in Table 14.1 [2]. The lack of the regulatory subunit makes the class II isoforms monomeric. It was thought that class II isoforms could phosphorylate PtdIns and phosphatidylinositol 4- phosphate (PtdIns (4) P); however, recent studies show that class II isoforms may prefer PtdIns producing phosphatidylinositol 3- phosphate (PtdIns (3) P) [2]. Class III includes one catalytic subunit, VPS34, and one regulatory subunit, VPS15. Like class I, class III isoforms are heterodimers with a catalytic and

Table 14.1 PI3K isoform subunits

	Catalytic subunit	Regulatory subunit
PI3K class IA isoforms	P110α	P85
	P110β	
	P110δ	
PI3K class IB isoforms	P110γ	P101
		P84
PI3K class II isoforms	PI3KC2α	
	PI3KC2β	
	PI3KC2δ	
PI3K class III isoforms	VPS34	VPS15

regulatory subunit. Class III isoforms phosphorylate PtdIns to generate PtdIns (3) P [2].

One of the most frequent occurrences in cancer is increased activation of the PI3K pathway [2, 7, 8]. This can occur by activation of tyrosine kinase receptors, loss of phosphatase tensin homologue (PTEN), and changes in isoforms [25]. PTEN is a protein made up of phosphatase and tensin that encodes the PTEN gene [2]. PTEN negatively regulates the PI3K pathway by dephosphorylation of PIP2 and PIP3 [25]. Class IA mutations are commonly found in many types of cancer. Mutations are found in three genes affecting the corresponding catalytic subunits as seen in Table 14.2. Gene PIK3CA mutations occur in all domains of p110α specifically the helical and kinase domains [7, 22]. Helical mutations limit the inhibition of p110α by p85 or they ease the interaction between p110α and receptors. Kinase mutations increase the interaction between p110α with membranes. PIK3CB gene mutations were seen in a breast cancer case where the helical domain increased the PI3K activation, therefore increasing p110β association with membranes [8]. PIK3CD gene mutations, however, have not been linked to cancers. Mutations found in the genes coding for regulatory subunits have also been seen in cancer. PIK3R1 gene mutations are frequently substitutions, insertions, or deletions in a domain of p85α [8]. These mutations prevent the binding of the regulatory subunit to the catalytic subunit ultimately preventing inhibition of p110. p85α can negatively regulate the PI3K signaling. Mutations in other genes such as PIK3R2, coding for p85β, have also been found but at a lower rate. Class IB mutations in PIK3CG are generally found to be over expressed but not mutated in cancers, similar to the PIK3CB and PIK3CD genes [8]. Little information is known regarding classes II and III. It is known that the genes PIK3C2A and PIK3C2B have been expressed in a few types of cancer; however, the outcome of that expression remains unknown [8]. Class III has minimal evidence that VSP34 plays a role in cancer and research is still ongoing [8]. Inhibitors have been a prime target of therapeutic treatment and work by binding to the target active site and preventing the substrates binding. Several inhibitors were tried as mono therapies including epidermal growth factor receptors (EGFR), BRAF, anaplastic lymphoma kinase (ALK), and PI3K [2]. As mono therapies the inhibitors have shown little progress and efficacy probably due to the lack of specificity. Instead, PI3K-based combination therapies were applied for better specificity and results. These combination therapies are in different stages of trial; however, they are showing better results.

14.2 Application of PI3K Isoform Inhibitors in Cancer Treatment

14.2.1 The Impact of PI3K Inhibitors on Breast Cancer

Inhibitors affect different processes in cells depending on their target of action. They also affect different types of cancer in different ways. Breast cancer inhibitors such as BYL719, Anastrozole, Buparlisib, and Pictilisib, as seen in Table 14.3, are studied along with other

Table 14.2 PI3K isoforms and corresponding genes

	Gene	Catalytic subunit	Gene	Regulatory subunit
PI3K class IA isoforms	PIK3CA	P110α	PIK3R1	P85α, p55α, p50α
	PIK3CB	P110β	PIK3R2	P85β
	PIK3CD	P110δ	PIK3R3	P85γ
PI3K class IB isoforms	PIK3CG	P110γ	PIK3R5	P101
			PIK3R6	P87, p84, p87(PIKAP)
PI3K class II isoforms	PIK3C2A	PI3K-C2α		
	PIK3C2B	PI3K-C2β		
	PIK3C2G	PI3K-C2γ		
PI3K class III isoforms	PIK3C3	VPS34	PIK3R4	VPS15

Table 14.3 Inhibitors tested in treatment for types of cancer

Cancer type	Inhibitor	Target of inhibitor
Breast cancer	BYL719	P110 inhibitor
	Anastrole	Aromatase inhibitor
	Buparlisib (BKM120)	PI3K inhibitor
	Pictilisib	PI3K inhibitor
Prostate cancer	GSK 2636771 (β)	P110 inhibitor
	AZD 8186 (β)	P110 inhibitor
	Everolimus	mTORC1 inhibitor
	Uprosertib (GSK2141795)	AKT inhibitor
Head and neck cancer	Buparlisib (BKM120)	PI3k inhibitor
	Copanlisib (BAY80–6946)	PI3K inhibitor
	MK2206	AKT inhibitor
	Everolimus (RAD001)	mTOR inhibitor
	Temsirolimus (CCI-779)	mTOR inhibitor
Leukemia	Idelalisib (CAL101)	PI3Kδ inhibitor
	Duvelisib (ABBV-954, INK-1197, IPI-145)	PI3Kδ/PI3Kγ inhibitor
	Dactolisib (BEZ235)	PI3K/mTOR inhibitor
	Buparlisib (BKM120)	Pan-PI3K inhibitor

inhibitors to understand their effect on breast cancer [26, 27]. Clinical studies have shown that pan- class I PI3K inhibitors have been affective in patients with cancer; however, new PI3K p110α specific inhibitors are showing even more effective in PIK3CA mutations [18, 27]. One of these specific inhibitors is BYL719 (also known as Alpelisib) and is a p110α inhibitor currently being studied and researched for its possible role in cancer treatment, specifically breast cancer [27]. Clinical trials showed a good response in most patients, but not all patients. This lack of response in some patients may be due to some resistance to PI3Kα inhibitors. When BYL719 was tested in multiple cell lines, it was determined that the marker Akt was being inhibited; however, the mechanism target of rapamycin complex 1 (mTORC1) component of the pathway was not being inhibited [27]. This lead to the combination of an mTORC1 inhibitor with BYL719 resulting in better outcomes. Another inhibitor currently being studied for its role in breast cancer is the aromatase inhibitor, Anastrozole. Anastrozole inhibits aromatase, which is an enzyme that aids in converting different hormones into estrogen [27]. This inhibitor, when added with other inhibitors, was seen to suppress cancer cell growth. Pictilisib is a PI3K inhibitor that was found to be toxic when used as a monotherapy treatment; however, is currently being studied when combined with other inhibitors [18, 27]. Pictilisib and Anastrozole

were combined in a study to test the effects of the combination therapy, and significant inhibition of cell growth was seen. Pictilisib has also been combined with Fulvestrant with little to no benefit on inhibition due to the toxicity of Pictilisib [18, 27]. Another PI3K inhibitor studied is Buparlisib. Two studies were directed at determining the most effective dosage without toxic effects on patients. When Buparlisib was paired with a human epidermal growth factor receptor 2 (HER-2) drug, the result was a patient population with slight evidence of antitumor activity, with more data needing to be collected [27].

14.2.2 The Impact of PI3K Inhibitors on Prostate Cancer

Inhibitors such as GSK2636771, AZD8186, Everolimus, and Uprosertib are studied for their effect on patients with prostate cancer [28]. GSK2636771 is a p110β inhibitor studied in PTEN-deficient tumors. This selective inhibitor is considered to help avoid toxicities found in other treatments because GSK2636771 targets p110β only rather than the other isoforms. This specific p110 inhibition is also being studied in combination with androgen receptor antagonists [28, 29]. Similar to GSK2636771, AZD8186 is a selective inhibitor of p110β as well as p110δ [28, 30]. AZD8186 is showing sufficient progress in cancer research as a monotherapy and in combination with other drugs. Isoform-specific PI3K inhibitors do however, show limited capability due to the toxicities and upregulation of pathways [30]. Everolimus is a mechanism target of rapamycin (mTOR) inhibitor also explored for its effect on prostate cancer [28]. When Everolimus was combined with another drug there were few patients with positive results compared to the higher percentage of patients with positive affects with just the other drug. This study indicated that Everolimus is not a successful inhibitor in prostate cancer [18]. Uprosertib, or GSK2141795, is an Akt inhibitor. In a study investigating Uprosertib, a majority of patients had stable disease and measurable responses after being treated [28, 29].

14.2.3 The Impact of PI3K Inhibitors on Leukemia

Research on leukemia is currently studying multiple inhibitors including Idelalisib, Duvelisib, Dactolisib, and Buparlisib [31]. Dactolisib is currently being studied for its influence on leukemia and other cancers. Dactolisib inhibits mTOR activity as well as PI3K activity, including the different isoforms [31, 32]. Akt phosphorylation activation must be initiated after Dactolisib therapy or the mechanism target of rapamycin complex 2 (mTORC2) inhibition will be very unstable [32]. Unregulated activation of the PI3K pathway has been seen in many patients with leukemia. As such, Buparlisib is a class I PI3K inhibitor, is being tested for its influence on regulating and inhibiting this pathway [31, 33]. It was reported that Buparlisib produced phosphorylation of PI3K target Akt suggesting a neutralized feedback [33]. A study with leukemia cells treated with Duvelisib, a PI3Kδ and PI3Kγ inhibitor, displayed a diminished adhesion of lymphocytes to endothelial cells [31]. This discovery suggests Duvelisib may be used for impairment of signaling between tumor cells [31]. A PI3Kδ inhibitor, known as Idelalisib, has been studied and found to be active against leukemia. The inhibition of the PI3K pathway decreases the tumor size and strengthens antileukemia activity [31, 34–38]. Idealalisib also has been found to minimize growth and increases apoptosis of leukemia cells [31, 34–38].

14.2.4 The Impact of PI3K Inhibitors on Head and Neck Cancer

Similarly to the other cancers discussed, head and neck cancer is also being studied to test if the use of inhibitors is a helpful treatment. Some inhibitors for head and neck cancer that are being examined are Buparlisib, Copanlisib, MK2206, Everolimus, and Temsirolimis [39–41]. Buparlisib is a PI3K inhibitor, inhibiting all of the class I isoforms. In a study conducted on Buparlisib, it was seen that there was a similar

impact on both the mutated and wild-type forms of the PIK3CA gene [39]. There is a concern with the inhibitor Buparlisib in the clinical application due to its higher toxicity levels than other PI3K inhibitors. Copanlisib is another PI3K inhibitor specific to isoforms p110α and p110δ. Like Buparlisib, Copanlisib shows similar influence on both mutant and wild-type forms of the PIK3CA gene and stipulates antitumor activity [39, 40]. Other inhibitors are used to inhibit other portions of the pathway. Akt activation is commonly connected to resistance of cancer treatment such as chemotherapy. For this reason it has been important to find Akt inhibitors such as MK2206. MK2206 has shown to be a very forceful Akt inhibitor and increases antitumor activity. In a trial in patients, MK2206 was seen to slow or stop progression of the head and neck cancer [39]. Final results for that trial are still being collected; however, they show promise. Everolimus is an mTOR inhibitor being tested in head and neck cancer. Alone, Everolimus was determined to be unsuccessful at inhibiting the pathway and current studies are testing it in combination with other inhibitors [39]. Temsirolimus is an inhibitor that has mTOR inhibition along with the ability to become hydrolyzed to form sirolimus, another mTOR inhibitor, after application [39].

14.3 PI3K Inhibitor-Based Combination Therapies for Cancer

PI3K inhibitors along with other treatments such as epidermal growth factor receptor (EGFR) inhibitors and Wnt inhibitors have shown little efficacy when given to patients as monotherapies, so combinations of the different inhibitors are being studied as combination therapies.

14.3.1 EGFR Inhibitors in Combination with PI3K Inhibitors for Cancer

Epidermal growth factor receptor inhibitors were used as a monotherapy to treat cancer patients. Patients showed antitumor activity; however, with long-term treatment a resistance was acquired and the cancer relapsed. The resistance to the EGFR inhibitors are likely due to the reactivation of the PI3K pathway [42–48]. It is known that the inhibition of a pathway can lead to the activation of other pathways in a feedback loop style. For this reason, a combination therapy is being tested combining EGFR inhibitors with PI3K inhibitors. The PI3K pathway involves cell proliferation and because of this, regulation of the pathway by combined therapy is efficient in apoptosis sensitive cell lines [46]. The combination of PI3K with EGFR have promising therapeutic outcomes; however, recent studies have shown anti-climatic data with the dual therapy. In head and neck cancer, an EGFR targeting antibody, known as Cetuximab, was combined with PI3K inhibitors such as PX-866 [43, 46]. This combination did not show any improvement in patient survival, leading scientists to believe there needs to be a deeper understanding to find a successful treatment for cancer. Studies are being continued in hopes to find the correct combination by combining different PI3K and EGFR inhibitors together on differing cell lines [42, 46]. In one experiment, different inhibitors were tested and it was determined that PI3Kα inhibitors in combination with EGFR inhibitors, show better outcomes in cancer patients [46].

14.3.2 Wnt Pathway Inhibitors in Combination with PI3K Inhibitors for Cancer

Abnormal wingless type MMTV integration site family, Wnt, signaling is seen in many types of cancer and is being studied as a form of treatment [49–51]. Using a dual therapy approach, the Wnt pathway inhibitors are combined with PI3K inhibitors for reasons similar to the EGFR pathway. The Wnt pathway inhibitors are studied in combination with PI3K inhibitors due to PI3K inhibitor resistance when used as a monotherapy. It was seen in some cancers, such as breast cancer, that the Wnt pathway was activated when treated with a PI3K inhibitor [49–51]. For this reason,

Table 14.4 The effects of the addition of Buparlisib

	Components of PI3K pathway that become inhibited	Components of Wnt pathway that become activated
Addition of Buparlisib	PIK3CA	Wnt ligands (WNT 5, WNT6, WNT7, etc.)
	PIK3CB	Frizzled receptors (FZD)
	PIK3CD	Beta catenin (CTNNB1)
	PIK3CG	LRP 1/5/6

dual combination therapy began being studied using a Wnt inhibitor, known as WNT974, in combination with PI3K inhibitor, Buparlisib [39]. WNT974 inhibits a protein known as porcupine which is important in the Wnt pathway [50, 51]. Cells were first treated with Buparlisib, which inhibited PI3K pathway components and activated Wnt components as seen in Table 14.4 [51]. Treatment of WNT974 was then administered to the cell lines and inhibition of Akt as well as increased expression of porcupine was observed [50, 51]. This combination of inhibitors taken together was seen to show significant effects in breast cancer, and further testing is being done to determine if it is plausible to use in humans [51].

14.4 Conclusions and Perspectives

PI3K inhibition is studied as a broad treatment for cancer showing varied results. The use of PI3K inhibitors is still currently being studied with some showing promising outcomes and others showing little to no outcomes. This data is used to generate a new treatment for cancer with the aim of better initiating apoptosis in cancer cells rather than all cells. Current treatments are very harsh on the human body including the good cells. The study of PI3K inhibitors is trying to better target cancer cells. PI3K inhibitors are showing some positive results; however, they are very limited. For this reason, PI3K inhibitors are also studied in combination with other therapies. The combination therapy treatment studies are seen to have better effects on treatment of cancer while still initiating apoptosis predominantly in cancer cells unlike current treatments.

Acknowledgments This work was supported in part by Augusta University CURS Summer Scholars.
Competing interests: The authors declare that they have no competing interests.

References

1. Vadas O, Burke JE, Zhang X, Berndt A, Williams RL (2011) Structural basis for activation and inhibition of class I phosphoinositide 3-kinases. Sci Signal 4(195): re2
2. Wang X, Ding J, Meng L (2015) PI3K isoform-selective inhibitors: next-generation targeted cancer therapies. Acta Pharmacol Sin 36(10):1170–1176
3. Wong KK, Engelman JA, Cantley LC (2010) Targeting the PI3K signaling pathway in cancer. Curr Opin Genet Dev 20(1):87–90
4. Wu P, Liu T, Hu Y (2009) PI3K inhibitors for cancer therapy: what has been achieved so far? Curr Med Chem 16(8):916–930
5. Yuan TL, Cantley LC (2008) PI3K pathway alterations in cancer: variations on a theme. Oncogene 27(41):5497–5510
6. Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer 9(1):28–39
7. Jean S, Kiger A (2014) Classes of phosphoinositide 3-kinases at a glance. J Cell Sci 127(5):923–928
8. Thorpe L, Yuzugullu H, Zhao J (2015) PI3K in cancer: divergent roles of isoforms, modes of activation, and therapeutic targeting. Nat Rev Cancer 15(1):7–24
9. Vanhaesebroeck B, Stephens L, Hawkins P (2012) PI3K signalling: the path to discovery and understanding. Nat Rev Mol Cell Biol 13(3):195–203
10. Vivanco I, Sawyers CL (2002) The phosphatidylinositol 3-kinase-AKT pathway in human cancer. Nat Rev Cancer 2(7):489–501
11. Cantley LC (2002) The phosphoinositide 3-kinase pathway. Am Assoc Adv Sci 296(5573):1655–1657
12. Engelman JA (2009) Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer 9(8):550–562
13. Engelman JA, Luo J, Cantley LC (2006) The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. Nat Rev Genet 7(8):606–619

14. Kok K, Geering B, Vanhaesebroeck B (2009) Regulation of phosphoinositide 3-kinase expression in health and disease. Trends Biochem Sci 34(3):115–127
15. Liu P, Cheng H, Roberts TM, Zhao JJ (2009) Targeting the phosphoinositide 3-kinase pathway in cancer. Nat Rev Drug Discov 8(8):627–644
16. LoRusso PM (2016) Inhibition of the PI3K/AKT/mTOR pathway in solid tumors. J Clin Oncol 34 (31):3803–3815
17. Rodon J, Dienstmann R, Serra V, Tavernero J (2013) Development of PI3K inhibitors: lessons learned from early clinical trials. Nat Rev Clin Oncol 10(3):143–153
18. Elkabets M, Vora S, Juric D, Morse N, Mino-Kenudson M, Muranen T et al (2014) mTORC1 inhibition is required for sensitivity to PI3K p110α inhibitors in PIK3CA-mutant breast cancer. Sci Transl Med 5(196):196ra99
19. Sadeghi N, Gerber DE (2012) Targeting the PI3k pathway for cancer therapy. Future Med Chem 4 (9):1153–1169
20. Osaki M, Oshimura M, Ito H (2004) PI3K-Akt pathway: its functions and alterations in human cancer. Apoptosis 9(6):667–676
21. Schink KO, Raiborg C, Stenmark H (2013) Phosphatidylinositol 3-phosphate, a lipid that regulates membrane dynamics, protein sorting and cell signalling. BioEssays 35(10):900–912
22. Samuels Y, Waldman T (2010) Oncogenic mutations of PIK3CA in human cancers. Curr Top Microbiol Immunol 347:21–41
23. Hawkins PT, Anderson KE, Davidson K, Stephens LR (2006) Signalling through class I PI3Ks in mammalian cells. Biochem Soc Trans 34:647–662
24. Vanhaesebroeck B, Guillermet-Guibert J, Graupera M, Bilanges B (2010) The emerging mechanisms of isoform-specific PI3K signalling. Nat Rev Mol Cell Biol 11(5):329–341
25. Parsons R (2004) Human cancer, PTEN and the PI-3 kinase pathway. Semin Cell Dev Biol 15(2):171–176
26. Bendell JC, Rodon J, Burris HA, de Jonge M, Verweij J, Birle D et al (2012) Phase I, dose-escalation study of BKM120, an oral pan-class I PI3K inhibitor, in patients with advanced solid tumors. J Clin Oncol 30(3):282–290
27. Qin H, Liu L, Sun S, Zhang D, Sheng J, Li B et al (2018) The impact of PI3K inhibitors on breast cancer cell and its tumor microenvironment. PeerJ 6:e5092
28. Crumbaker M, Khoja L, Joshua A (2017) AR signaling and the PI3K pathway in prostate cancer. Cancers 9(4):34
29. Edlind MP, Hsieh AC (2014) PI3K-AKT-mTOR signaling in prostate cancer progression and androgen deprivation therapy resistance. Asian J Androl 16 (3):378–386
30. Hancox U, Cosulich S, Hanson L, Trigwell C, Lenaghan C, Ellston R et al (2015) Inhibition of PI3Kbeta signaling with AZD8186 inhibits growth of PTEN-deficient breast and prostate tumors alone and in combination with docetaxel. Mol Cancer Ther 14 (1):48–58
31. Sanchez V, Nichols C, Kim H, Gang E, Kim Y (2019) Targeting PI3K signaling in acute lymphoblastic leukemia. Int J Mol Sci 20(2):412
32. Schult C, Dahlhaus M, Glass A, Fischer K, Lange S, Freund M et al (2019) The dual kinase inhibitor NVP-BEZ235 in combination with cytotoxic drugs exerts anti-proliferative activity towards acute lymphoblastic leukemia cells. Int J Cancer Res Treat 32 (2):463–474
33. Ragon B, Kantarjian H, Jabbour E, Ravandi F, Cortes J, Borthakur G et al (2017) Buparlisib, a PI3K inhibitor, demonstrates acceptable tolerability and preliminary activity in a phase I trial of patients with advanced leukemias. Am J Hematol 92(1):7–11
34. Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P et al (2014) Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. N Engl J Med 370(11):997–1007
35. Gopal AK, Kahl BS, de Vos S, Wagner-Johnston ND, Schuster SJ, Jurczak WJ et al (2014) PI3Kδ inhibition by idelalisib in patients with relapsed indolent lymphoma. N Engl J Med 370(11):1008–1018
36. Herman SE, Gordon AL, Wagner AJ, Heerema NA, Zhao W, Flynn JM et al (2010) Phosphatidylinositol 3-kinase-delta inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. Am Soc Hematol 116:2078–2088
37. Hoellenriegel J, Meadows SA, Sivina M, Wierda WG, Kantarjian H, Keating MJ et al (2011) The phosphoinositide 3′-kinase delta inhibitor, CAL-101, inhibits B-cell receptor signaling and chemokine networks in chronic lymphocytic leukemia. Am Soc Hematol 118 (13):3603–3612
38. Lannutti BJ, Meadows SA, Herman SE, Kashishian A, Steiner B, Johnson AJ et al (2011) CAL-101, a p110delta selective phosphatidylinositol-3-kinase inhibitor for the treatment of B-cell malignancies, inhibits PI3K signaling and cellular viability. Am Soc Hematol 117(2):591–594
39. Cai Y, Dodhia S, Su G (2017) Dysregulations in the PI3K pathway and targeted therapies for head and neck squamous cell carcinoma. Oncotarget 8 (13):22203–22217
40. Garcia-Carracedo D, Villaronga M, Alvarez-Teijeiro S, Hermida-Prado F, Santamaria I, Allonca E et al (2016) Impact of PI3K/AKT/mTOR pathway activation on the prognosis of patients with head and neck squamous cell carcinomas. Oncotarget 7(20):29780–29793
41. Vander B, Mohan S, Eytan DF, Chen Z, Van W (2015) The PI3K/Akt/mTOR axis in head and neck cancer: functions, aberrations, cross-talk, and therapies. Oral Dis 21(7):815–825
42. Benavente S, Huang S, Armstrong EA, Chi A, Hsu KT, Wheeler DL et al (2009) Establishment and characterization of a model of acquired resistance to

epidermal growth factor receptor targeting agents in human cancer cells. Clin Cancer Res 13(10):935–945
43. Boeckx C, Baay M, Wouters A, Specenier P, Vermorken JB, Peeters M et al (2013) Anti-epidermal growth factor receptor therapy in head and neck squamous cell carcinoma: focus on potential molecular mechanisms of drug resistance. Oncologist 18 (7):850–864
44. Ercan D, Xu C, Yanagita M, Monast CS, Practilas CA, Montero J et al (2013) Reactivation of ERK signaling causes resistance to EGFR kinase inhibitors. Cancer Discov 2(10):934–947
45. Niederst MJ, Engelman JA (2013) Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Am Assoc Adv Sci 6(294): re6
46. Sato H, Yamamoto H, Sakaguchi M, Shien K, Tomida S, Shien T et al (2018) Combined inhibition of MEK and PI3K pathways overcomes acquired resistance to EGFR-TKIs in non-small cell lung cancer. Cancer Sci 109(10):3183–3196
47. Shien K, Toyooka S, Yamamoto H, Soh J, Jida M, Thu KL et al (2013) Acquired resistance to EGFR inhibitors is associated with a manifestation of stem cell-like properties in cancer cells. Cancer Res 73 (10):3051–3061
48. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W et al (2013) Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res 19(8):2240–2247
49. Anastas JN, Moon RT (2013) WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer 13(1):11–26
50. Liu J, Pan S, Hsieh MH, Ng N, Sun F, Wang T et al (2013) Targeting Wnt-driven cancer through the inhibition of porcupine by LGK974. Proc Natl Acad Sci U S A 110(50):20224–20229
51. Solzak JP, Atale RV, Hancock BA, Sinn AL, Pollok KE, Jones DR et al (2017) Dual PI3K and Wnt pathway inhibition is a synergistic combination against triple negative breast cancer. NPJ Breast Cancer 3:17

Leslie Duncan is an undergraduate pre-medical student at Augusta University completing a Bachelor of Science degree in Cell and Molecular Biology with a minor in Chemistry. She is an Honors student conducting basic research under the supervision of Dr Teng and the awardee of CURS Summer Scholars 2019 and 2020.

Chloe Shay is the Senior Associate Director of Programs at Emory University. She received the Doctor of Business Administration degree from Georgia State University and Executive Leadership for Healthcare Professionals Certificate from Cornell University. In the last 15 years, she has authored more than 50 peer-reviewed articles and book chapters in the fields of behavioral research and medical sciences, and overseen numerous clinical trials funded by the federal government, institutions, and international pharmaceutical companies.

Yong Teng is an influential young scientist at Dental College of Georgia in Augusta University, with a joint professor appointment in Medical College of Georgia and College of Allied Health Sciences. He is also a member of Molecular Oncology and Biomarkers Program at Georgia Cancer Center. The main research activity in Dr Teng's lab is not limited to understand and reverse mechanisms of cancer metastasis and metabolism. His research team also focuses on using novel 3D culture systems and tumor-targeting nanoparticles to develop highly effective anticancer modalities in order to improve the quality of cancer patients. He has been the principal investigator on multiple studies funded by DOD and NIH. His ongoing projects seek to shift current research and clinical practice paradigm, which will significantly impact the future development of advanced materials and novel strategies for cancer treatment.

Till 2019, Dr Teng has authored more than 80 articles and chapters in peer-reviewed journals and books, in most of which he was the first or/and corresponding author. He serves on multiple national and international grant review panels and scientific advisory boards, and successfully organized many prestigious conferences. In addition, he holds professional memberships with several scientific and professional societies (AACR, AHA, ADEA, etc.). Dr Teng is currently the editorial board member of Journal of Experimental and Clinical Cancer Research and other reputed journals. He also serves as an Associate Editor of many journals, including Frontiers in Oncology, Cancer Management and Research, Frontiers in Molecular Biosciences, and OncoTargets and Therapy, and edits the Special Issue and Collection for the Journal Cells, Frontiers in Oncology, Frontiers in Molecular Biosciences, Amino Acids and others.

Single Cell Sequencing in Cancer Diagnostics

15

Frederik Otzen Bagger and Victoria Probst

Abstract

Personalized medicine has been driven by improvements in genomic sequencing and analysis. For several diseases, in particular cancers, it has for nearly a decade been standard clinical practice to analyze the genome and expression of the genes of patients. The results are reflected directly in the treatment plan for the patient, in targeted medical inventions. This specialized mode of diagnostics has been restricted to account for averaged trends in the tumor. The approach sharply contrasts our knowledge on heterogeneity within tumors. Several studies further describe how treatment against one tumor subclone in some cases merely serves to provide space and support for uncontrolled growth of more aggressive subclones. In this chapter, we describe current possibilities for implementation of single cell sequencing of malignomas in clinic, as well as discuss hands-on practical advice for single cell routine diagnostics that allows for full delineation of tumor clonality.

Keywords

Single cell sequencing · Cancer · Diagnostics · Precision medicine · Clinical research

F. O. Bagger (✉) · V. Probst
Department of Genomic Medicine, Rigshospitalet, Centre of Diagnostics, Section 4113, Copenhagen, Denmark
e-mail: frederik.otzen.bagger@regionh.dk

15.1 Introduction

To this date, assessment of cancer tumor sub-clonality in single patients has not been technically feasible. However, recent development of technologies for Single Cell Sequencing (SCS) has progressed research on tissue heterogeneity to a new paradigm, and today single cell technologies are applied on a large scale, famously in the Human Cell Atlas (HCA) project, aiming at building a reference map of all human cells [1]. Similar initiatives applying single cell sequencing include National Institutes of Health (NIH) Human Biomolecular Atlas Program (HuBMAP), and The LifeTime Initiative [2, 3]. Also, SCS is being used to monitor disease progression and response to therapy in cancer research [4–8]. In the following, we briefly introduce the field of SCS in clinic and the implications for cancer, followed by a review of clinically relevant single cell technologies, with key guides for clinical implementation of single cell sequencing of malignomas.

15.2 Importance of Single Cell Technologies

Currently, Clinical next generation sequencing (NGS) techniques are built upon bulk DNA- and RNA sequencing, where tumors are treated as a homogenous tissue specimen. SCS has the potential to unravel sub-clonality of tissue (e.g., tumor

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_15

samples) by looking at samples at the level of an individual cell. Bulk-sequencing has limited capabilities to uncover the cellular heterogeneity of tumors, because it provides an averaged signal from a complex cell population [9]. Deep-sequencing, and sequencing of different spatial regions of tumors, have proven to disclose clonal heterogeneity to some extent, however, the data always reflects an averaged signal [10]. This potentially makes identifying distinct cell populations difficult, and the ability to uncover populations important for disease progression might be lost. Another issue with bulk-sequencing of heterogeneous samples is the incapability of distinguishing whether transcriptomic changes are caused by alteration in gene regulation or by a shift in the ratio of different cell populations. These caveats associated with traditional bulk NGS potentially result in patients being treated with therapies targeting only the most abundant cancer cell population, whereafter less abundant populations might sustain and spread. Applying technologies for unraveling tumor heterogeneity are hypothesized to make targeted therapies more efficient and potentially capable of minimizing risk of disease relapse. Single cell genomics is believed to have the ability to account for traditional NGS-associated biases by uncovering the heterogeneous nature of tumors (Fig. 15.1).

15.3 Mutagenesis and Cancer

Cancerous genetic variants are either acquired or inherited. Inherited, or germline, variants may predispose individuals to develop certain types of cancer. Cancer is a condition where cells divide uncontrollably, and development requires acquired mutations that circumvent advanced networks of defense mechanisms. Mutagenesis may occur from intrinsic risk factors, being random errors in DNA replication, or non-intrinsic endogenous risk factors such as genetic susceptibility or biological ageing. Mutagenesis might also occur from non-intrinsic exogenous risk factors such as radiation, chemical substances, or lifestyle choices [11].

Mutations that enhance cell proliferation or affect the stability of the entire genome, are important for cancer development. Some initial mutations increase the risk of a cell to acquire additional mutation by either inducing genomic instability or by giving the cell a growth advantage [12]. Mutations to genes that have the potential to increase the overall mutational rate are divided into two broad categories—tumor suppressor genes (TSGs) and proto-oncogenes [13]. The mechanism of oncogene activation is gene amplification and may be caused by point mutations leading to hyperactive gene products or overall increased transcription of the gene. This can occur via gene duplications which is classified as a copy number variation (CNVs), or it can occur via translocations that relocate the gene to a chromosomal region with higher expression [14–16].

The normal role of TSGs is to restrain inappropriate cell growth, and they contribute to tumorigenicity by loss-of-function mutations leaving both maternal and paternal copy of the gene dysfunctional [13]. Cancerous mutation affecting TSGs are often point mutations or small deletions resulting in nonfunctional gene products, or chromosomal breaks that destroy the gene [17]. Alternatively, TSG inactivation can be acquired through epigenetic changes, such as DNA methylation or modification of histone tails [6]. Hereditary cancer is often associated with mutations affecting one copy (either maternal or paternal) of a TSG. One mutant TSG copy increases an individual's susceptibility to loss of heterozygosity (LOH). LOH in tumors refers to loss of the functional copy of a TSG in a heterozygous person (born with only one functional copy), making tumor cells homozygous for loss of the gene. Individuals born heterozygous for TSGs mutations are thus at higher risk of developing cancer [18].

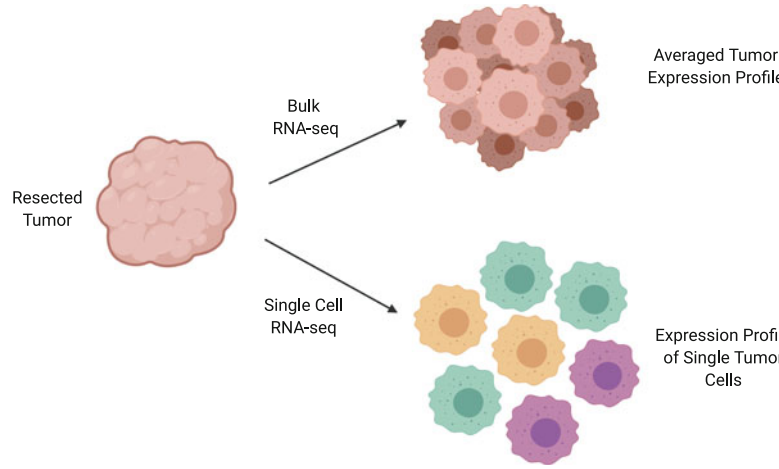


Fig. 15.1 Bulk vs. single cell RNA sequencing. Bulk RNA sequencing provides an averaged transcriptional profile across all cell populations residing within the tumor sample. Single cell sequencing gives an individual transcriptional profile from single cells of the tumor specimen, allowing assessment of intra-tumor heterogeneity. (Figure Illustrated using *Biorender*)

15.4 Precision Diagnostics

Precision diagnostics refers to tailoring a medical treatment plan to suit the characteristics of each patient's disease. In precision diagnostics each patient is classified into a subgroup, depending on e.g., genetics, gene expression patterns, disease susceptibility, treatment options, or treatment response. Targeted treatment refers to drugs targeting specific genes, proteins or parts of the tissue environment, and is highly applied in personalized cancer therapy. One of the most prominent challenges in cancer diagnostics is matching each patient with the most suitable treatment, among the available options. Tailoring medical treatment to each patient is predicted to become more common in the future, with the aim of minimizing disease progression, period, and risk of relapse [19]. Current workflows for precision diagnostics of cancer patients are built on knowledge from bulk RNA and DNA sequencing of the tumor specimen. A whole-blood (WH) sample is also retrieved from the patient for germline analysis by whole-genome sequencing (WGS). WH sequencing commonly acts as a normal background for the patient, allowing to identify tumor specific (somatic) changes resulting from the cancer. Both DNA- and RNA sequencing can directly determine treatment options for the specific patient, and whether the patient is eligible for targeted treatment. DNA sequencing of tumor tissue determines the genetics of a cancer by identifying mutation from a panel of variants known to have clinical implications. RNA sequencing assigns a cancer subtype by comparing the expression profile with that of previous cancer patients. This is achieved either by mathematical transformation of full gene expression pattern in the tumor, or as determined by expression levels of small sets of predefined *marker genes*.

15.4.1 Therapeutic Resistance

A challenge in clinical application of targeted therapy is when patients acquire therapeutic resistance following a period of treatment. Resistance to therapy and disease relapse has been correlated with advanced disease progression following targeted treatment [20, 21]. Cancer clonal evolution has been suggested to take place within tissue ecosystems where cancer cells interact with stromal and immune cells in their neighboring microenvironment [22]. These interactions are crucial for providing cancerous cells with resources, however also limits their developmental potential, due to repression from the immune system, and space limitations. Limited resources and space provide a natural selective pressure within the tumor, leading certain subpopulation to proliferate to a greater extent than others (Fig. 15.2).

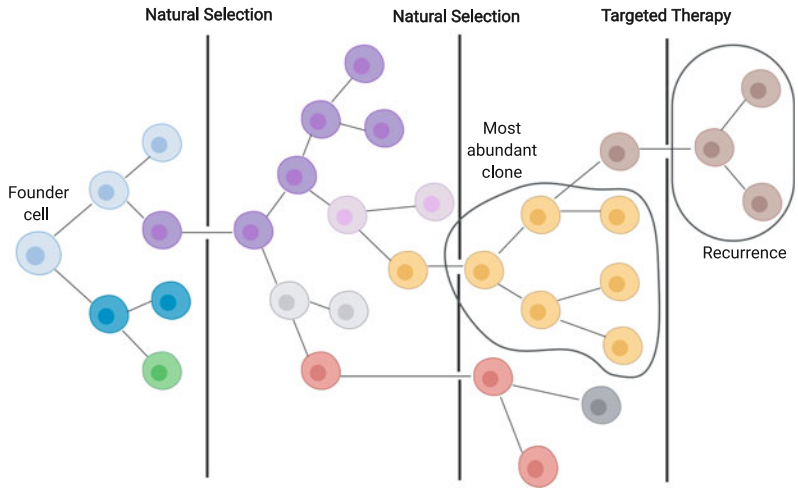


Fig. 15.2 Tumor ecosystem. Illustration showing branching of subclones from a founder cell. The tumor microenvironment allows for selective pressure of specific subclones to expand and others to become extinct. Introduction of an artificial selective pressure (e.g., Targeted therapy) may eliminate therapy sensitive cells, leaving space and resourcing for resistant cells to cause cancer recurrence. (Figure inspired from Greaves and Maley [22]. Illustrated using *Biorender*)

Treating patients with drugs or radiotherapy introduces an artificial selective pressure, changing the dynamics within the tumor. It has been suggested that the introduction of an artificial pressure may aid therapeutic resistance by eliminating therapy sensitive cells, while potentially leaving resistant cancerous cells with excess space and resources [22]. These remaining cells can cause cancer recurrence (Fig. 15.2). Following targeted treatment, the genetic profile of the tumor is likely to be changed, and new strategies for patient treatment must be planned. Tumor cell drug resistance is enabled in several ways including drug inactivation, change of drug target or drug target levels, increased drug efflux, decreased uptake, increased DNA damage repair, or apoptosis suppression [23].

15.5 Single Cell Sequencing: A Clinical Aspect

NGS implementation in diagnostics of cancer patients is an important technique for providing prognosis of disease state, and to propose suitable patient treatment. DNA sequencing provides information about CNV, Single nucleotide variants (SNV), and structural variants (SV). RNA-seq provides information on which gene products are expressed and in what quantity. Techniques for SCS has within recent years been developed to enable studying human health and development as well as providing potential for diagnostics, monitoring, and treatment of disease. Several different approaches for single cell RNA and DNA sequencing have been developed within the last decade [24, 25].

Techniques for SCS are roughly divided in two categories—plate-based (e.g., SMART-seq) and droplet-based techniques (e.g., 10x Genomics Chromium™ Single Cell 3′ RNA-seq) (Fig. 15.3). Both strategies start with dissociated single cells in suspension and aims at adding a unique identifier (barcode) to each cell, allowing for multiplexed sequencing on Illumina sequencers. For plate-based techniques each single cell is deposited into a chamber in a multi-well PCR plate or a tube, and subsequently indexes are added to each, as a part of final library preparation. Droplet-based techniques are based on microfluidics, where single cells in suspension and primer-covered beads are capsuled in small emulsion droplets, leading to random indexing. In each droplet, a single cell is lysed and combined

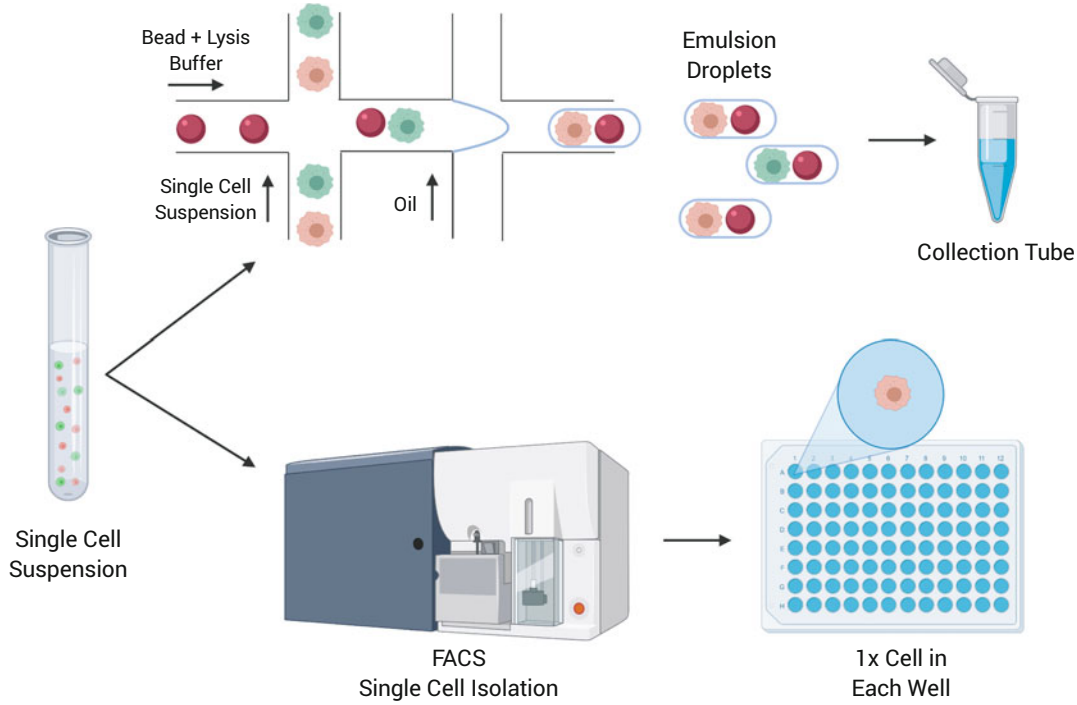


Fig. 15.3 Single cell sequencing technologies. Single cell sequencing technologies are roughly divided into two categories—Plate-based approaches (Bottom) and Droplet-based approaches (Top). Plate-based approaches isolate single cells into lysis buffer containing multi-well PCR plates or tubes. Isolation may be performed using fluorescent activated cell sorting (FACS). Barcodes are added during final library preparation steps. Droplet-based approaches combine primer-covered beads and single cells in emulsion droplets. Cells are lysed within the droplet, and reverse transcription carried out. Single cell cDNA is subsequently pooled and processed in bulk. (Figure Illustrated using *Biorender*)

with a bead covalently bound to primers containing both unique molecular identifiers (UMIs) and a bead-specific barcode. In general, droplet-based approaches allows for preparation of a great number of single cell libraries simultaneously. However, it only allows for detection of highly expressed genes. For most diagnostic purposes this level of resolution is inadequate, mainly because a high number of important marker genes will never be detected, only because they have low expression. Plate-based methods require more technical know-how and does not allow for processing as many cells in parallel, as droplet-based methods. Plate-based approaches does, however, allow for detection of more genes per cell, important in a clinical aspect [26]. Selecting the most appropriate single cell RNA- or DNA-sequencing method is furthermore a question of price versus gained information. Droplet-based methods are in general cheaper per cell, but the information gained from each cell is shallow, and currently limiting for both research and clinic.

15.6 Single Cell Genome Sequencing

With single cell genome sequencing, genomic heterogeneity of cell populations can be explored, such as mutations or chromosomal abnormalities. Plate-based single cell DNA-sequencing (scDNA-seq) have few commonly used—Multiple Annealing and Looping-Based Amplification Cycles (MALBAC), Multiple Displacement Amplification (MDA), and PicoPLEX®. All techniques are available as commercial kits. Common for all current scDNA-seq methods is

the requirement of an amplification step, which has the potential to introduce an amplification bias, well known from low-input bulk DNA sequencing techniques [27]. Importantly, quantification of repetitive regions which are notoriously hard to sequence with standard short read sequencing, is sensitive to even amplifications, where not all methods perform equally. Repetitive regions include gene duplication CNVs, important for cancers.

15.6.1 MDA

MDA is based on annealing random hexamer primers to denatured DNA, followed by primer extension and strand displacement enabled by *Bacillus Stearothermophilus* (Bst) phi29 DNA polymerase at isothermal conditions [28]. MDA has been suggested to surpass PicoPLEX® and MALBAC for detection of Small nucleotide polymorphisms (SNPs) and Structural variants (SVs) [24]. These properties are enabled by the low error rate of Bst phi29 DNA polymerase and the generation of larger amplicons than PCR-based methods. However, MDA exhibits amplification bias by over-representing some areas while under-representing others. When sequencing complex diploid genomes, such as human, MDA may give rise to uneven genome coverage yielding false quantification of CNVs.

15.6.2 MALBAC

MALBAC is a technique combining MDA with PCR. Random primers are annealed to single-stranded templates similar to MDA and extended using Bst phi29 DNA polymerase at isothermal conditions. Using a PCR machine, several cycles of annealing, extension, melting, and looping are performed, ensuring a more equal distribution of transcripts across the genome compared to MDA. This property makes MALBAC more suitable for detection of CNVs [24, 29]. However, MALBAC suffers a greater proportion of false positives by amplification of polymerase errors occurring in early PCR cycles, which makes MALBAC less suitable than MDA for detection of SNVs [24].

15.6.3 PicoPLEX®

PicoPLEX® is a purely PCR-based technique. Denatured template DNA is pre-amplified using random primer extension with primers containing sequences that hybridize to single-stranded DNA (ssDNA), and with a PCR handle at its 3′ end. Pre-amplification steps accumulate a hairpin library which is amplified by PCR with primers toward the handle sequences. PicoPLEX® technique has been found to be most suitable for CNV detection. New generation PicoPLEX® Gold promises improved detection of both SNVs and SVs to levels exceeding MDA [24, 30].

15.7 Single Cell Transcriptome Sequencing

Single cell mRNA sequencing allows for the study of intercellular transcriptional variability, enables identification of marker genes and related pathways in specific subpopulation of cells. All current fully functioning techniques require isolation and subsequent lysis of single cells, conversion of RNA to cDNA, and amplification of cDNA. Due to low concentration of starting material, necessitating multiple rounds of amplification, the process of single cell RNA sequencing (scRNA-seq) suffers from several technical biases. Choosing your method for scRNA-seq depends on which transcript properties one wishes to uncover.

ScRNA-seq techniques enabling full-length transcripts are preferred when wishing to uncover structural variation, mutations within transcripts (still problematic from mRNA) or for the detection of pseudogenes and splice variants. However, a disadvantage to this feature is the preclusion of early barcoding steps and incorporation of Unique Molecular Identifiers (UMIs). UMIs are 4–10 random nucleotides introduced in the primer oligo for 5′ and 3′ cDNA synthesis. The principle for adding UMIs

is to establish a unique identity of each RNA molecule. During PCR amplification, each cDNA containing the same UMI will be considered derived from the same transcript molecule. This, in theory, eliminates PCR amplification bias commonly seen when applying a high number of PCR cycles, leading to falsely counting identical reads as different transcripts [31, 32]. Counting UMIs instead of actual reads make protocols suitable for transcript CNV detection. A way of adjusting for amplification bias in full-length RNA-seq protocols is by adding External RNA Control Consortium (ERCC) spike-ins to the experiment [33]. ERCCs are 92 synthetic transcripts that function to standardize sequencing experiments by adding an equal amount to each single cell reaction prior to processing steps. ERCCs are of bacterial origin and designed to show minimal sequence homology with endogenous eukaryotic transcripts, but features a poly (A)-tail, different GC-content, and varying lengths. Applying ERCCs in a sequencing experiment can account for biases such as primer capture efficiency and are used for normalization of gene expression across cells during data analysis.

Droplet-based methods have a limited transcript information compared to full-length protocols, by only capturing 3′ or 5′ transcript-ends for sequencing. Techniques that are UMI-based are primarily suitable for high-throughput experiments of gene expression heterogeneity, since these have an advantage with small technical biases. UMI-based techniques are not suitable in experiments designed to uncover specific transcript properties, or for diagnostic purposes where specific variants are of great importance and severe dropouts are not tolerated [25, 34].

15.7.1 First Generation Full-Length ScRNA-seq

Fuchou Tang's method was the first plate-based technique for full-length scRNA-seq, and built upon the unique feature of mRNA, the poly-A tail at the end of each mRNA which is to be translated into protein [35]. In brief, a poly (T) primer, which will bind the poly-A tail, coupled to a specific anchor sequence allows to make a cDNA copy, by reverse transcription, of full length of the mRNA. The single-stranded cDNA is subsequently added a 3′ end poly(A)-sequence to which a new poly(T) primer coupled to a different anchor sequence is annealed, facilitating the formation of double stranded cDNA molecules. These cDNA molecules are then amplified using polymerase chain reaction (PCR), and amplicons of each cell in the plate is then a sample that can be sequenced with standard sequencing machines.

15.7.2 SMART-seq1/2/3

Switching Mechanism at the 5′ End of RNA Template (SMART-seq) is a plate-based technique also built around the polyadenylation (poly(A)) feature of mRNA transcripts. SMART-seq is a protocol for full-length transcriptome sequencing relying on Reverse Transcription (RT) followed by template switching (TS) [36]. The poly(A)-tails of mRNA transcripts are primed using an oligo-d(T) primer coupled to a PCR handle. The primed mRNA is reverse transcribed by Moloney Murine Leukemia Virus (M-MLV) RT, which has terminal transferase activity, and adds non-templated nucleotides to the 3′ end of cDNA ends. These non-templated nucleotides are preferentially cytosines, which allow annealing of a template switching oligo (TSO) containing riboguanosines at its 3′ end. The second generation, SMART-seq2, applies a TSO carrying a locked nucleic acid (LNA) in its last 3′ end position. The LNA locks the nucleotide in endo-formation, improving base-stacking and annealing ability yielding raised melting temperature between the cDNA strand and the TSO [37]. This feature provides SMART-seq2 with a stronger ability for transcript capture, which mounts in improved gene detection [38]. In general, SMART-seq2 is believed to have good sensitivity and accuracy in regards to gene detection, and to give even read coverage [25, 38]. Several different SMART-seq kits are commercially available from different vendors (e.g., Clontech and New England

Biolabs) differing in chemistry, price-point, and hands-on processing time. SMART-seq3 is the newest generation full-length scRNA-seq protocol and differs from traditional SMART-seq by the implementation of UMIs in the 5′ end of full-length RNA transcripts [39]. Inclusion of UMI counting gives the protocol higher power in regard to transcript copy number, detection, and PCR bias. The protocol features a next-generation MMLV RT, switching of salt component from KCl to NaCl, and implements PEG to improve cDNA yield during reverse transcription. The protocol also adds GTPs during RT to support TS. SMART-seq3 has been suggested to improve the sensitivity of original SMART-seq protocols to sensitivity levels approaching single-molecule RNA fluorescence in situ hybridization (smRNA FISH) [39].

15.7.3 MARS-seq

Mars-seq is a plate-based method implementing Unique Molecular Identifiers (UMIs) and barcoding of each single cell during transcript priming. Single cells are, as in the case of SMART-seq, sorted into PCR plates or tubes, in which libraries are prepared. MARS-seq has been found to have a high dropout rate and a low detection of genes per cell compared to remaining SCS protocols [32]. Second generation MARS-Seq 2.0 was released as an improvement to the original MARS-Seq protocol. By lowering RT working volumes and quantity of reaction reagents, price per single cell was reduced six-fold, and background noise reduced from 15 to 2% [40].

15.7.4 Seq-well

Seq-well is a portable UMI-based single cell RNA-seq method [41]. Seq-well differs from both droplet-based and plate-based RNA-seq approaches by using sealed arrays of sub-nanoliter wells. Primer-covered mRNA capture beads are isolated in each well, and subsequently a single cell suspension is poured over the array, before sealing the plate using a semi-permeable membrane. The size of the well should limit the majority of wells to contain only a single primer-covered bead and a single cell. Newest generation seq-well S^3 have been found to have more sensitive transcript capture and higher gene detection than drop-seq and 10x Chromium, however it cannot outperform plate-based SMART-seq2 [42]. Due to its simplicity, low cost, and high-throughput seq-well has been applied in clinical studies to identify HIV-infected immune cells and research on cells infected with Malaria in developing countries [43]. Seq-well is not sold as a commercially available kit, however due to its simplistic experimental design, doesn't require a plenitude of specialized equipment.

15.7.5 Drop-seq

Drop-seq is a droplet-based method where a flow of primer-covered beads suspended in lysis buffer and a flow of single cell suspension are combined using a microfluidic chip generating emulsion droplets. In each droplet a single cell is lysed and combined with a bead covalently bound to oligo-d(T) primers containing both UMIs and a bead-specific barcode. Cells are lysed within each droplet, and mRNA attach to the oligo-d(T) covered beads. Subsequently droplets are dissolved to allow for parallel cDNA amplification, and final library preparation of all cells within a single tube [25, 44]. The technique has a low cost per cell, however the cell input requirement is fairly high (~150.00 cells per run) [25, 45]. Furthermore, drop-seq has been found to have low detection of genes per single cell compared to remaining SCS protocols [32]. Drop-seq is not available as a commercial kit, which can make experiment setup time consuming and technically challenging.

15.7.6 10x Genomics Chromium Single Cell 3′ RNA-seq (10x Chromium)

The 10x Chromium system is a droplet-based semiautomated platform, commercially available with reagent kits to be used in small specialized liquid handler machine. It works by combining single cells in suspension and primer-covered beads within a single emulsion droplet enabled by the use of a microfluidic chip. Each primer contains both a UMI and a unique barcode. Within the droplet, mRNA transcripts are converted to cDNA. Following RT, emulsion droplets are dissolved, and final library preparation is performed in bulk. 10x Chromium system can prepare up to 10,000 single cell libraries in a single run and requires a read depth of approximately 25,000 reads per single cell. Due to its ability to cover a large number of cells, 10x Chromium has been found better at detecting rare cell types in tissue compared to SMART-seq2 [26]. 10x Chromium have a slightly better transcript capture and requires a lower minimum cell input compared to other droplet-based approaches [45]. 10x Chromium data has been found to have severe gene dropout problems, especially of lowly expressed genes [26]. Plate-based method SMART-seq2 has a higher gene detection, especially in regard to low abundance transcripts and spliced transcripts [26]. A great advantage of 10x Chromium is that it's almost fully automated, minimizing both hands on time and risk of inducing confounding factors (e.g., pipetting errors). Even though the price of 10x Chromium is slightly higher than similar droplet-based approaches (e.g., drop-seq), the per cell library price is lower than plate-based approaches (e.g., SMART-seq2) [45, 46].

15.8 Single Cell Parallel Genome and Transcriptome Sequencing

Single cell G&T-seq, is a method developed for studying both the genome and transcriptome of the same single cell [47] (Fig. 15.4). This method enables investigation of genetic variations and its correlation with gene expression. The core of the method is a step separating mRNA and DNA from the same single cell into different plates or tubes. Following separation, DNA and mRNA is eligible for a range of library preparation protocols, individually processed and sequenced. The protocol was originally described applying a modified SMART-seq2 protocol for transcriptome amplification, and PicoPLEX® or MDA for genome amplification [36, 47]. The separation of DNA from mRNA is enabled by primer coupled magnetic beads (Oligo-d(T)30VN beads). The primer is an oligo-nucleotide containing a PCR sequence, a stretch of 30 thymidine residues (oligo-d(T)30), and an anchor sequence (VN) (V = A,G, or C; N = A,G,C or T) coupled to biotin in the 5′ end (Fig. 15.4). The 5′ Biotin modification enables conjugation to streptavidin coated magnetic Dynabeads®. Oligo-d(T)30VN beads capture mRNA transcripts in the cell lysate, and transcripts are moved to one part of the well using a magnet. This allows for transferring DNA in the supernatant to a new plate, which can be stored at −80 °C for later processing. mRNA is immediately processed, performing on-bead RT. The protocol applies a strong lysis buffer containing guanidinium salt and sodium dodecyl sulfate (SDS) to enable complete lysis of both the cell membrane and nucleus membrane. Before processing, DNA must be liberated from this lysis buffer using solid-phase reversible immobilization (SPRI) beads to avoid inhibition of downstream amplification protocols.

G&T-seq is relevant for cancer diagnostic purposes by the ability of sequencing both RNA and DNA from the same single cell. Single cell transcriptome data can be used to cluster cells into subpopulations, which may show differences in proliferation, marker gene expression, or metastatic potential. DNA-seq from these single cells may be pooled according to assigned subpopulations for disclosing relevant diagnostic mutagenic variants between and within subpopulations.

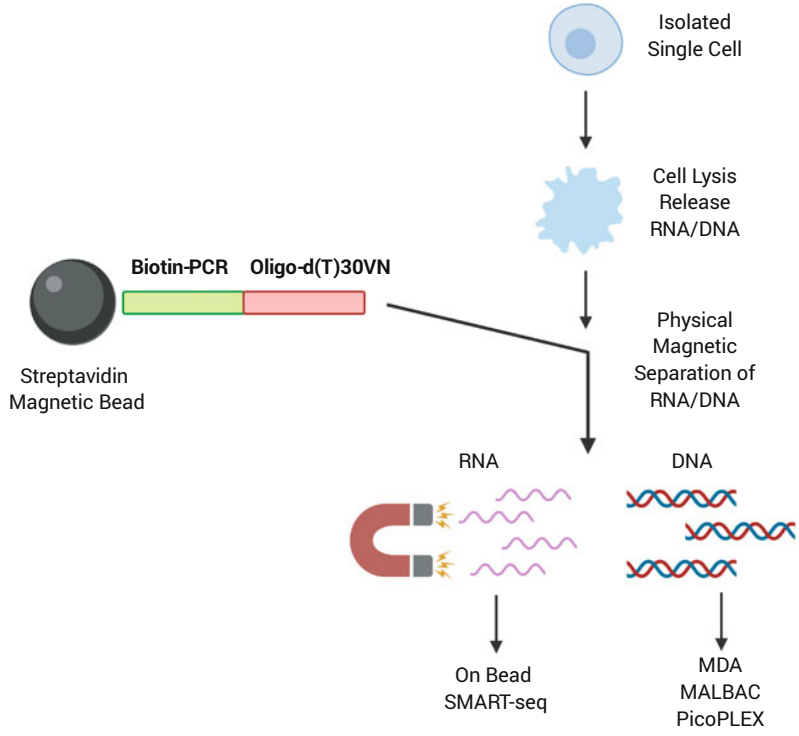


Fig. 15.4 Genome and transcriptome sequencing. Overview of G&T-seq technique. DNA and mRNA from the same single cell are physically separated using streptavidin magnetic beads coupled to a biotin oligo-d(T)30VN primer. Subsequently both DNA and mRNA can be amplified and sequenced separately. DNA is stored at −80 °C for later processing, and may be amplified by either MDA, MALBAC, or PicoPLEX®. MRNA is amplified by on-bead SMART-seq technology immediately following separation. VN = anchor sequence (V = A,G, or C; N = A,G,C, or T). oligo-d(T)30 = a stretch of 30 deoxythymidine. (Figure inspired from Macaulay et al. [48]. Illustrated using *Biorender*)

15.9 Single Cell RNA-seq Data Processing

ScRNA-seq initial data processing comprise steps of alignment, gene annotation, count matrix processing, and normalization. Following initial processing biological analysis of single cells often include identification of differentially expressed (DE) genes, clustering of single cells into distinct populations, classification of cells, and trajectory reconstruction. Normalization has been suggested to be the step influencing biological DE analysis the most. Second most important factor for DE analysis performance was suggested to be library preparation protocol [49].

Most appropriate choice of alignment tool depends on the sequencing protocol. 3′ UMI-based protocols are more sensitive to the choice of aligner and annotation tool. In a comparison of tools, Spliced Transcripts Alignment to a Reference (STAR) tool in combination with GENCODE assigned most reads to 3′ UMI-based protocols. Kallisto was found to assign fewest reads, has the lowest mean gene expression and highest gene dropout rates for 3′ UMI-based protocols [49]. Aligning 3′ UMI reads using Burrows-Wheeler Aligner (BWA) were found to give a higher false mapping fraction, and expression variance than UMI reads mapped using STAR. BWA also has reduced power to detect differentially expressed features compared to STAR, or kallisto alignment for UMI reads. In

general, UMI-based protocols are recommended aligned using STAR with GENCODE annotation, whereas for full-length protocol SMART-seq2, kallisto with RefSeq annotation performed slightly better than STAR and have advantage of lower computation time [49]. ScRNA-seq data has lower signal-to-noise ratio, and higher drop-out of both cells and transcripts compared to bulk data. Careful filtering prior to normalization and biological analysis can minimize this issue [50]. A number of checks can be performed, bearing in mind that sequencing is a sampling of fragments of cDNAs, where quantity is only meaningful relative to the total number of reads. The total number of reads depend only on how much sequencing time and reagents is invested, but one to ten million reads per cell have been reported to capture most mRNA information, compared to cost of more cells, but it will vary, depending on the platform [51]. Common quality control (QC) metrics for filtering scRNA-seq data are library size, feature count (features are genes, fragments, or transcripts, depending on alignment strategy), and proportion of ERCC spike-ins or mitochondrial features [52]. Library size constitutes the total sum of counts across all features for each single cell. Cells with small libraries are considered of low quality. This could be the result of RNA degradation by either RNase contamination or a product of inefficient transcript capture before first strand synthesis and cDNA amplification. Library size, however, also depends on sequencing depth. The number of expressed features is defined as the number of features with nonzero counts for each specific cell. Cells with very few expressed genes are considered to be of poor quality as the wide range of expressed transcripts have not been efficiently captured.

If spike-in controls were added to the sample, the proportion of reads mapped to ERCC spike-ins can be calculated relative to the library size of each single cell. ERCC spike-ins are added in equal amount to each cell lysis prior to processing. An enrichment in spike-in counts represents either loss of endogenous RNA, empty wells, or simply too high concentration of spike-ins added each sample. The proportion of mitochondrial features (mtRNA) can be assessed as a QC metric the same may ERCCs spike-ins are used. The reasoning is that due to the size of mitochondria, and physical protection of mtRNAs, they will not as readily leak from perforated cells as individual transcripts. Mitochondrial transcripts are thus more likely to be protected from degradation, captured and sequenced, and will take up a larger proportion of the reads in damaged or stressed cells.

In general, there should be rough agreement between QC metrics. If there is no clear agreement, it might correspond to technical bias in the data—pipetting uncertainty, sample preparation steps, or PCR bias. However, a lack of agreement between metrics may also represent genuine biological differences—RNA content, size differences, or different mtRNA content.

15.10 Before Starting a Plate-Based Single Cell Experiment

Handling single cell RNA and DNA is much more technically demanding than handling a pool of cells. Single cell experiments can be challenging to perform because it requires both specialized equipment and technical know-how on sample and reagent handling. Diploid human cells contain between 0.5–50 pg of RNA and 6.6 pg of DNA, which make single cell experiments extremely sensitive to contamination, sample degradation, and elevated background noise in the final sequencing data. All pre-amplification steps must be performed within a positive air flow hood, located in a clean pre-PCR room. Instruments must be cleaned using UV hood, ethanol, and DNA/RNase away. Experimenter must wear gloves and a clean lab coat when processing single cells.

Establishment of protocols for generating successful plate-based single cell libraries often requires some adaptation to the specific cell type. When working with primary cells, it is recommended to first test and practice protocols using a comparable cell line. Often protocols for

plate-based SCS do not provide guidance on either preparation of single cell suspensions nor sorting of your cells. However, both are pivotal starting points for generating successful single cell libraries.

15.10.1 Single Cell Suspensions

Preparation of single cell suspension is an often disregarded but critical starting point of SCS protocols. Cell line single cell suspensions compatible with SCS are performed similar to subculturing protocols. However, it is important to thoroughly wash cells from tryptic medium. Preparing single cell suspensions from primary cells are more challenging. First, if performing scRNA-seq, cells must be processed max. 4 h post surgery and kept on ice, to minimize risk of RNA degradation [53]. Second, it's pivotal to design your enzymatic dissociation assay to fit the tissue of origin and avoid damage to sorting surface markers. No standard protocols exist to fit all tissues, and this step will require optimization for each tissue. Furthermore, working with epithelial cells, it is recommended to use correct plastic equipment (e.g., LoBind Eppendorf tubes), to avoid losing sticky cells. Finally, single cell suspension, either primary or cell lines, must be resuspended for isolation in a suitable sorting buffer. Sorting buffer must be free of agents, salts or enzymes, that can interfere with downstream amplification protocols. Cell viability between 60–90% is generally to be expected for primary cells when dissociating using a fast disaggregation protocol. Cell viabilities below this threshold should cause reason for concern, as cells might be damaged, apoptotic or necrotic, induced either prior to or during processing. This will cause a shift in transcriptomic activity not representing tissue of origin, which might reflect as untruthful conclusions in your final sequencing data. Enrichment for viable cells can be performed by applying several rounds of low force centrifugation, removing dead cells and cell debris. Working with bloody tissue, it's recommended to remove erythrocytes and enrich for cells of interest by applying red blood cell (RBC) lysis buffer (Do not apply if working with mammalian alpha-intercalated kidney cells). Results not shown are from work at Department of Genomic Medicine, Danish State Hospital, Rigshospitalet, Denmark.

15.10.2 Single Cell Isolation

The handling and isolation of single cells is a technically challenging task. A variety of methods are available for the isolation and sorting of single cells into lysis buffer containing PCR plates or tubes [54]. For plate-based SCS purposes, the most common method for single cell isolation is fluorescence activated cell sorting (FACS). FACS allows sorting of a heterogeneous single cell suspension into populations based on the cells physical properties enabled by light scattering. Fluorescent labeling allows further distinguishment between single cells based on surface markers and viability. When designing a fluorescent assay, it's important to minimize spectral overlap between fluorochromes. Cell lysis should take place as the single cell hits the well. It is thus of great importance to carefully center the stream before each sort, making sure the cell hits the lysis buffer and not the sides of the well. Immediately after sorting, samples should be thoroughly vortexed to make sure single cells are completely lysed and cell content placed within the buffer. Finally, samples should be spinned down, and placed in dry ice for a few minutes until completely frozen before storing at −80 °C for future use.

15.10.3 PCR Cycling

When working with single cells, the number of PCR cycles must be adjusted according to the RNA content of the cell. In general, the number of pre-amplification cycles should be kept as low as possible to avoid introducing PCR bias into the experiment. However, if cDNA concentration is too low, it is difficult to assess the quality of the cDNA fragments. In general, between 17 and 20 cycles is suggested for cell lines, depending

on the RNA content of the strain. For single primary cells the approximate RNA content should determine the number of PCR cycles for proper amplification. The optimal number of cycles should be determined according to trial-and-error for each tissue and cell type when conducting a new single cell experiment. Bear in mind that commercial kits are often designed and tested on cell lines, thus suggesting lower cycles than needed for live tissue.

15.11 Research Studies of Carcinomas by Single Cell Sequencing

The first single cell study by *Navin et al*. [7] on breast cancer heterogeneity applied single nuclei-sequencing of 100 cells from both a polygenomic and a monogenomic triple-negative (HER2-/ER-/PR-) breast cancer patient [7]. The study proved it was possible to flow-sort and amplify genomic DNA of a single cell nuclei for NGS quantification of gene copy numbers. Analysis of the polygenomic patient tumor showed presence of four distinct populations. The first population consisted of CNV-neutral diploid cells and "pseudo diploid" cells, that is cells with normal copy numbers but abnormal chromosomal composition. The three remaining populations had specific CNVs. Each subpopulation showed relation to one another by shared genomic alterations, while also having distinct characteristics. These findings correlate with beliefs of cancer progression by clonal expansions. Investigation of CNV breakpoints within the monogenomic tumor (52 cells) and its liver metastasis (48 cells) showed high genomic similarities between cells of the metastasis and one population within the original tumor. This attributed to a belief that cancer metastasis occur from a single clonal expansion within the original tumor. The authors introduce the term "punctuated clonal evolution," describing how distinct mutations give rise to complicated genomic alterations within very few cell generations. A more extensive study on CNV breakpoints supporting punctuated clonal evolution was published a few years later [5]. In this study 1000 single nuclei were sequenced from a total of 12 HER2-/ER-/PR- breast tumors (Average 83 cells per patient). This study also provided insight into tumor heterogeneity by finding that patient tumors had 1–3 major subpopulations of aneuploid cells, supporting subclonal architecture of breast tumors.

A study of single cell transcriptomes by *Chung et al*. [55] further attributed to exploration of inter- and intra-tumor heterogeneity in breast cancer [55]. In this study 515 single cell transcriptomes were amplified by full-length SMART-seq1 technology from 11 patients with different breast cancer subtypes. Both tumor and immune cells were isolated together by "marker-free" microfluidics. The study showed that each cancer subtype did have a core gene expression profile correlating to its assigned breast cancer subtyped by bulk RNA- seq. However, SCS revealed intra-tumor heterogeneity. First, it was found that only one of three patients characterized as HER2+ had actual activation of HER2 downstream pathway. One HER2+ patient had both HER2+ and ER+ expression, with predominating ER pathway initiation, and categorized as LumB subtype. This HER2+/ER+ LumB subtype patient might initially correspond well to hormonal therapy targeting ER, however may retain HER2+ cells. Extinction of the ER+ subpopulation potentially leave space and resources for the HER2+ subgroup to remain and spread, leading to prolonged disease for the patient. Another HER2+ patient had been treated using HER2 targeted therapy based upon expression analysis at the sampling time point. However, following targeted therapy, the genetic profile of the tumor had changed, showing low level HER2+ pathway activation and higher basal gene expression resembling HER2-/ER-/PR- tumors.

A study by *Karaayvaz et al.* was published in 2018, applying SMART-seq2 technology on >1500 breast tumor cells from six untreated HER2-/ER-/PR- patients with the aim of unraveling sub-clonality [56]. Cells were sorted by FACS using both a "marker-free" approach and by negative selection of CD45+ cells. Cell types were subsequently identified by cell type specific marker genes in the expression data.

Malignant cells were selected as cycling cells and showed that most G1/S and G2/M cells (98.5%) were of epithelial origin, consistent with the idea that malignant cells exhibit higher proliferation than non-malignant cells. Furthermore, clustering of all cells showed that while immune, stromal, and endothelial cell types clustered together, both inter- and intra-tumoral, epithelial cells formed several distinct clusters. Epithelial cell clusters from tumors were also compared to normal mammary epithelia, and showed that normal epithelial cells had high concordance, while tumor epithelial cells had weaker concordance both inter- and intra-tumorally. In general, tumor epithelial cells clustered according to tumor, even when some clusters consisted of cells from different patients, suggesting the existence of subpopulations defined by common states between patients. The study also investigated the correlation between RNA expression and CNVs from bulk whole-exome sequencing of four out of six patients. The analysis found high correlation between CNV in the genome and copy number in expression data, suggesting that genomic aneuploid alterations determine transcriptional alterations, giving rise to intra-tumor heterogeneity.

Single cell transcriptomic analysis of head and neck squamous cell carcinoma (HNSCC) also highlighted carcinogenic intra-tumor heterogeneity [57]. Single cell RNA was processed using a modified SMART-seq2 protocol, featuring a RNA cleanup step, on 2215 malignant and 3363 non-malignant cells from 18 different tumors. Malignant epithelial cells from different patients were found to have concordance in expression signatures, suggesting common patterns of intra-tumor expression. Seven different expression programs were identified within malignant populations, associated with either cell cycling, hypoxia, stress, epithelial-mesenchymal transition, or epithelial differentiation. The authors also suggested a refined HNSCC tumor classification system. Four different bulk-sequencing classification subtypes of HNSCC tumors exist—basal, classical, atypical, and mesenchymal. However, malignant cells from 10 patient tumors mapped to just three subtypes (basal = 7, classical =2, atypical = 1). No malignant cells mapped to the mesenchymal subtype, even though this represents the second most common subtype among HNSCC tumors. Malignant cells from mesenchymal subtype assigned tumors instead mapped to the basal subtype. Expanding the analysis to include stromal and immune cells, suggested that tumors mapped to mesenchymal subtype by bulk-sequencing merely had a higher degree of stromal and immune cell infiltration. The authors thus suggested a refined classification system for HNSCC tumors, eliminating mesenchymal subtype.

In a study from 2017, single cell RNA signatures of colorectal cancer cells showed that tumors previously assigned to a single cancer subtype by bulk-sequencing could be divided into subgroups with different survival probabilities [58]. *Li et al.* sequenced transcriptomes of 375 tumor cells from 11 different colorectal cancer patients and 215 cells from matched normal tissue using Fluidigm C1 SMART-seq1 protocol. The authors identified three clusters of tumor epithelial cells assigned as Stem/transit-amplifying-like, enterocyte 2B-like, and goblet-like. Ninety-three percent of tumor epithelial cells were identified as Stem/transit-amplifying-like, in contrast to only 30% of normal epithelia. Comparing differentially expressed (DE) genes between tumor and normal tissue from single cell data with DE genes identified by bulk-sequencing, it was found that the majority of DE genes from single cell analysis were lost in bulk analysis. The list of tumor-normal tissue DE genes identified by bulk-sequencing were also found contaminated by cell type markers relating to differences in composition of cell types within the tissue. The authors suggest that single cell transcriptomics and subsequent cell type clustering provides a better approach for identification of DE genes between tumor and normal tissue. Furthermore, it was found that single cell transcriptome sequencing could refine traditional colorectal tumor classification by introducing new tumor

subclasses defining differences in patient survival outcomes.

15.12 Conclusions

Future precision diagnostics of cancer patients are predicted to include both RNA and DNA sequencing of single cells for identification of tumor subpopulations (Fig. 15.5). Several different research grade SCS methods have been developed within the last decade and choosing between them depends on the biological question at hand. In general plate-based full-length RNA-seq approaches are evaluation to be better suited for diagnostics purposes, due to higher gene detection and genome coverage. Current droplet-based UMI approaches are too insensitive and inconsistent for implementation in robust diagnostic workflow. Research on intra-tumor

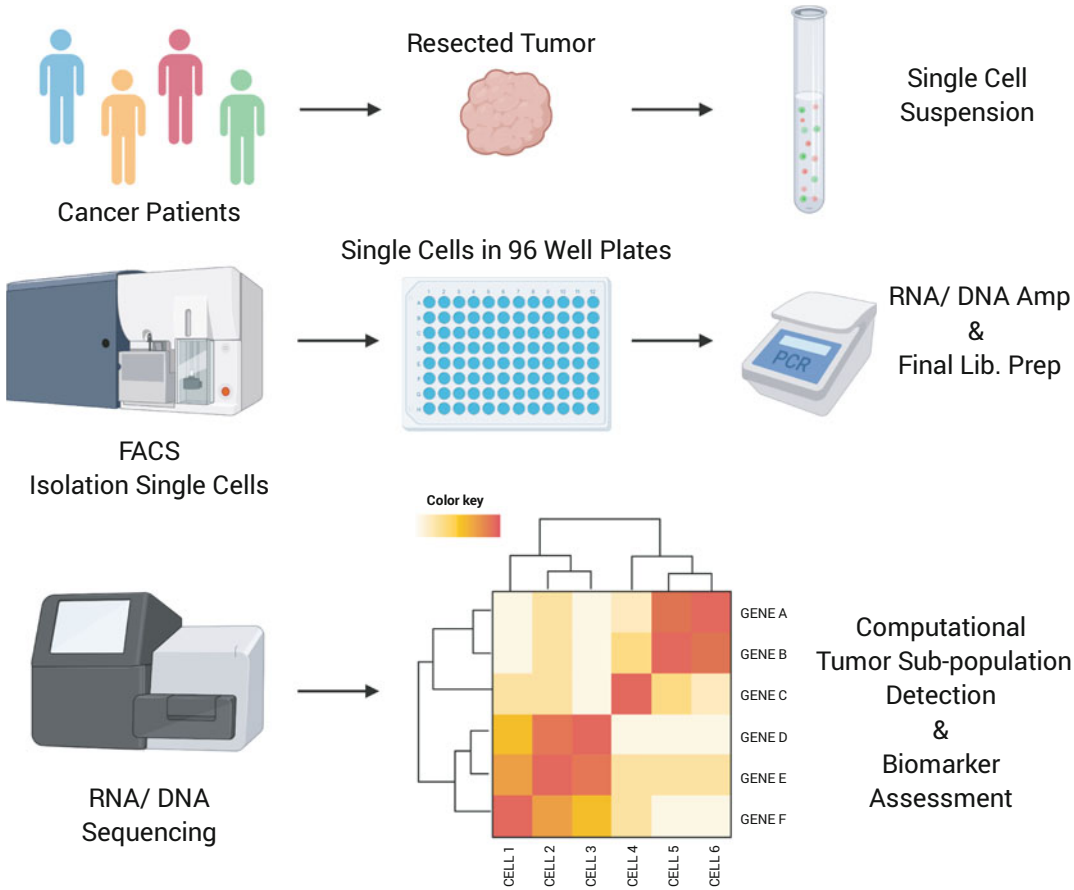


Fig. 15.5 Improved cancer precision diagnostics workflow. Diagnostics of cancer patients are predicted to include single cell sequencing in the near future. Plate-based full-length RNA-seq approaches are best suited for diagnostic cancer subtype classification, due to higher gene detection and genome coverage than droplet-based RNA-seq approaches. Cancerous tumors are collected max. 4 h post surgery, and tumors dissociated to single cell suspensions. Single cells are isolated in multi-well PCR plates using fluorescent activated cell sorting (FACS). RNA and DNA is amplified using PCR, and subsequently subjected to library preparation allowing for multiplexed sequencing on Illumina machines. Following sequencing, computational analysis allows detection of tumor subpopulations, and identification of possible biomarkers for targeted therapy. (Figure Illustrated using *Biorender*)

heterogeneity applying both RNA and DNA single cell sequencing has highlighted the importance of applying such technologies, also in the clinic. Identification of subclones within cancer patient tumors are hypothesized to improve precision diagnostics, minimize disease period and risk of cancer relapse, by allowing for treatment of all cancer subclones, and not only the major clone. A general pitfall with sequencing single cells from a tumor specimen, is evaluating the amount of cells needed for authentic mapping of the tumor landscape. Sequencing too few cells or only one area of a tumor might still lead to lost detection of subclones. Thorough testing and evaluation of SCS on clinical samples must be performed to determine whether clinical implementation of these technologies provide novel information important for efficient patient treatment. Immediate challenges with implementation of single cell sequencing in clinic is the lack of standardized computational pipelines for downstream analysis of subclones and clusters of cells. Previous studies have suggested that similarities between tumor subclones are higher than differences between individuals. This opens for automated approaches and clinical use of also global mRNA expression patterns with little batch, or individual, effect to account for, which has been a major challenge in diagnostics of bulk samples. When larger number of patients and tumors have been profiled at single cell level it is possible that future subtyping, identification of origin of metastasis and comparison with previous treatment responses will be computationally feasible, without current research driven analysis by bioinformatics experts in batch-driven cohorts.

References

1. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E et al (2017) The human cell atlas. elife 6:pii: e27041. https://doi.org/10.7554/eLife.27041
2. HuBMAP (2019) The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature 574(7777):187–192. https://doi.org/10.1038/s41586-019-1629-x
3. LifeTime (2019) The LifeTime initiative. https://lifetime-fetflagship.eu/. Accessed 15 Oct 2019
4. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S et al (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell 174(5):1293–1308. e1236. https://doi.org/10.1016/j.cell.2018.05.060
5. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y et al (2016) Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nat Genet 48(10):1119–1130. https://doi.org/10.1038/ng.3641
6. Kazanets A, Shorstova T, Hilmi K, Marques M, Witcher M (2016) Epigenetic silencing of tumor suppressor genes: paradigms, puzzles, and potential. Biochim Biophys Acta 1865(2):275–288. https://doi.org/10.1016/j.bbcan.2016.04.001
7. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J et al (2011) Tumour evolution inferred by single-cell sequencing. Nature 472(7341):90–94. https://doi.org/10.1038/nature09807
8. Sun Z, Wang CY, Lawson DA, Kwek S, Velozo HG, Owyong M et al (2018) Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells. Oncotarget 9 (13):10945–10961. https://doi.org/10.18632/oncotarget.23760
9. Trapnell C (2015) Defining cell types and states with single-cell genomics. Genome Res 25(10):1491–1498. https://doi.org/10.1101/gr.190595.115
10. Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I et al (2008) Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci U S A 105 (35):13081–13086. https://doi.org/10.1073/pnas.0801523105
11. Wu S, Zhu W, Thompson P, Hannun YA (2018) Evaluating intrinsic and non-intrinsic cancer risk factors. Nat Commun 9(1):3490. https://doi.org/10.1038/s41467-018-05467-z
12. Aguilera A, Gomez-Gonzalez B (2008) Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet 9(3):204–217. https://doi.org/10.1038/nrg2268
13. Lee EYHP, Muller WJ (2010) Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol 2 (10):a003236. https://doi.org/10.1101/cshperspect.a003236
14. Botezatu A, Iancu IV, Popa O, Plesa A, Manda D, Huica I, Vladoiu S, Anton G, Badiu C (2016) Mechanisms of oncogene activation
15. Anderson MW, Reynolds SH, You M, Maronpot RM (1992) Role of proto-oncogene activation in carcinogenesis. Environ Health Perspect 98:13–24. https://doi.org/10.1289/ehp.929813
16. Croce CM (2008) Oncogenes and cancer. N Engl J Med 358(5):502–511. https://doi.org/10.1056/NEJMra072367
17. Wang LH, Wu CF, Rajasekaran N, Shin YK (2018) Loss of tumor suppressor gene function in human cancer: an overview. Cell Physiol Biochem 51 (6):2647–2693. https://doi.org/10.1159/000495956

18. Larsen MJ, Thomassen M, Gerdes AM, Kruse TA (2014) Hereditary breast cancer: clinical, pathological and molecular characteristics. Breast Cancer (Auckl) 8:145–155. https://doi.org/10.4137/BCBCR.S18715
19. Love-Koh J, Peel A, Rejon-Parrilla JC, Ennis K, Lovett R, Manca A et al (2018) The future of precision medicine: potential impacts for health technology assessment. PharmacoEconomics 36(12):1439–1451. https://doi.org/10.1007/s40273-018-0686-6
20. Hagio K, Baba M, Ishida N, Oshino T, Kasahara R, Nara M, Yamashita H (2018) Tamoxifen withdrawal in women with progressive metastatic breast cancer: a case series of six patients. Int Cancer Conf J 7 (4):142–147. https://doi.org/10.1007/s13691-018-0339-9
21. Mittendorf EA, Wu Y, Scaltriti M, Meric-Bernstam F, Hunt KK, Dawood S et al (2009) Loss of HER2 amplification following trastuzumab-based neoadjuvant systemic therapy and survival outcomes. Clin Cancer Res 15(23):7381–7388. https://doi.org/10.1158/1078-0432.Ccr-09-1735
22. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481(7381):306–313. https://doi.org/10.1038/nature10762
23. Kartal-Yandim M, Adan-Gokbulut A, Baran Y (2016) Molecular mechanisms of drug resistance and its reversal in cancer. Crit Rev Biotechnol 36 (4):716–726. https://doi.org/10.3109/07388551.2015.1015957
24. Wang K, de Bourcy CFA, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR (2014) A quantitative comparison of single-cell whole genome amplification methods. PLoS One 9(8):e105585. https://doi.org/10.1371/journal.pone.0105585
25. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M et al (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65(4):631–643.e634. https://doi.org/10.1016/j.molcel.2017.01.023
26. Wang X, He Y, Zhang Q, Ren X, Zhang Z (2019) Direct comparative analysis of 10X Genomics Chromium and Smart-seq2. bioRxiv. https://doi.org/10.1101/615013
27. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods 6(4):291–295. https://doi.org/10.1038/nmeth.1311
28. Dean FB (2001) Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 11(6):1095–1099. https://doi.org/10.1101/gr.180501
29. Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338 (6114):1622–1626. https://doi.org/10.1126/science.1229164
30. Takarabio (2019) High-resolution CNV detection using PicoPLEX Gold DNA-seq. https://www.takarabio.com/learning-centers/next-generation-sequencing/technical-notes/single-cell-rna-and-dna-seq/high-resolution-cnv-detection-using-picoplex-gold-dna-seq. Accessed 1 Mar 2019
31. Sena JA, Galotto G, Devitt NP, Connick MC, Jacobi JL, Umale PE et al (2018) Unique molecular identifiers reveal a novel sequencing artefact with implications for RNA-seq based gene expression analysis. Sci Rep 8(1):13121. https://doi.org/10.1038/s41598-018-31064-7
32. Ziegenhain C, Vieth B, Parekh S, Hellmann I, Enard W (2018) Quantitative single-cell transcriptomics. Brief Funct Genomics 17(4):220–232. https://doi.org/10.1093/bfgp/ely009
33. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M et al (2011) Synthetic spike-in standards for RNA-seq experiments. Genome Res 21 (9):1543–1551. https://doi.org/10.1101/gr.121095.111
34. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC et al (2017) Power analysis of single-cell RNA-sequencing experiments. Nat Methods 14(4):381–387. https://doi.org/10.1038/nmeth.4220
35. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N et al (2009) mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods 6(5):377–382. https://doi.org/10.1038/nmeth.1315
36. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9 (1):171–181. https://doi.org/10.1038/nprot.2014.006
37. Grunweller A, Hartmann RK (2007) Locked nucleic acid oligonucleotides: the next generation of antisense agents? BioDrugs 21(4):235–243. https://doi.org/10.2165/00063030-200721040-00004
38. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10(11):1096–1098. https://doi.org/10.1038/nmeth.2639
39. Hagemann-Jensen M, Ziegenhain C, Chen P et al. (2020) Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol. https://doi.org/10.1038/s41587-020-0497-0
40. Keren-Shaul H, Kenigsberg E, Jaitin DA, David E, Paul F, Tanay A, Amit I (2019) MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. Nat Protoc 14(6):1841–1862. https://doi.org/10.1038/s41596-019-0164-4
41. Gierahn TM, Wadsworth MH II, Hughes TK, Bryson BD, Butler A, Satija R et al (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods 14(4):395–398. https://doi.org/10.1038/nmeth.4179
42. Hughes TK, Wadsworth MH, Gierahn TM, Do T, Weiss D, Andrade PR et al (2019) Highly efficient,

massively-parallel single-cell RNA-seq reveals cellular states and molecular features of human skin pathology. bioRxiv. https://doi.org/10.1101/689273
43. Landhuis E (2018) Single-cell approaches to immune profiling. Nature 557(7706):595–597. https://doi.org/10.1038/d41586-018-05214-w
44. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161(5):1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
45. AlJanahi AA, Danielsen M, Dunbar CE (2018) An introduction to the analysis of single-cell RNA-sequencing data. Mol Ther Methods Clin Dev 10:189–196. https://doi.org/10.1016/j.omtm.2018.07.003
46. 10x Genomics (2019) Single-cell RNA-seq: an introductory overview and tools for getting started. https://community.10xgenomics.com/t5/10x-Blog/Single-Cell-RNA-Seq-An-Introductory-Overview-and-Tools-for/ba-p/547. Accessed 26 Feb 2019
47. Macaulay IC, Teng MJ, Haerty W, Kumar P, Ponting CP, Voet T (2016) Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. Nat Protoc 11(11):2081–2103. https://doi.org/10.1038/nprot.2016.138
48. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ et al (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 12(6):519–522. https://doi.org/10.1038/nmeth.3370
49. Vieth B, Parekh S, Ziegenhain C, Enard W, Hellmann I (2019) A systematic evaluation of single cell RNA-seq analysis pipelines. Nat Commun 10(1):4667. https://doi.org/10.1038/s41467-019-12266-7
50. Soneson C, Robinson MD (2018) Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods 15(4):255–261. https://doi.org/10.1038/nmeth.4612
51. Svensson V, da Veiga Beltrame E, Pachter L (2019) Quantifying the tradeoff between sequencing depth and cell number in single-cell RNA-seq. bioRxiv. https://doi.org/10.1101/762773
52. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA (2016) Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17:29. https://doi.org/10.1186/s13059-016-0888-1
53. O'Flanagan CH, Campbell KR, Zhang AW, Kabeer F, Lim JLP, Biele J et al (2019) Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. Genome Biol 20(1):210. https://doi.org/10.1186/s13059-019-1830-0
54. Gross A, Schoendube J, Zimmermann S, Steeb M, Zengerle R, Koltay P (2015) Technologies for single-cell isolation. Int J Mol Sci 16(8):16897–16919. https://doi.org/10.3390/ijms160816897
55. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT et al (2017) Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun 8(1):15081. https://doi.org/10.1038/ncomms15081
56. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC et al (2018) Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nat Commun 9(1):3588. https://doi.org/10.1038/s41467-018-06052-0
57. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S et al (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 171(7):1611–1624.e1624. https://doi.org/10.1016/j.cell.2017.10.044
58. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL et al (2017) Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet 49(5):708–718. https://doi.org/10.1038/ng.3818

Frederik Otzen Bagger, Senior researcher and PI at Department of Genomic Medicine, Copenhagen State Hospital, Visiting researcher at University of Basel, Universitäts-Kinderspital beider Basel (UKBB), Switzerland. PhD from University of Copenhagen, Post-Doctoral from EMBL-EBI and University of Cambridge, with broad experience within 'omics techniques, data analysis, and deep learning, with focus on single cell resolution methods for clinic and research. >35 published papers in the past 5 years with over 2600 citations. Founder and consultant in Fobinf ApS, Scientific advisory board of Immunitrack ApS. Inventor of the SinaPlot, founder of BloodSpot.eu and HemaExplorer.

Victoria Probst is involved in clinical and diagnostic research at department of Genomic Medicine (GM). Miss Probst was trained in single cell technologies at Wellcome Trust Genome Campus, Cambridge. Her role has been to develop and perform single cell RNA-sequencing (scRNA-seq) protocols at GM. Developing protocols for dissociation of fresh tumor tissue, sorting of single tumor cells using FACS, and sequenced transcriptomes of primary single cells from breast cancer (BC) patients using these advanced technologies. Victoria has set up nontrivial protocol for performing scRNA-seq by G&T-seq at GM, in collaboration with leading professor within the field Prof. Thierry Voet, department of human genetics, university of Leuven, Belgium. Finally miss Probst has implemented bioinformatic pipelines for trimming, alignment, and clustering of scRNA-seq data for visualization of tumor populations, and assessment of their transcriptional variability. Miss Probst aim to use the established protocols of single cell sequencing on BC patients to investigate its clinical significance for use in routine diagnostics.

Single Cell RNA Sequencing in Human Disease: Renal, Pancreatic, and Viral Diseases

16

Sayra Garcia, Evan Der, and Chaim Putterman

Abstract

In this chapter, we discussed some of the specific uses of scRNA-seq in exploring viral infections and diseases of the kidney and pancreas. This review, however, is by no means exhaustive, and indeed this technology has advanced the study of pulmonary and cardiac diseases, transplant immunology, cancer, and many others as well. Nevertheless, the above reviewed studies do illustrate the utility and resolution of scRNA-seq in understanding exact cellular compositions, discovering heterogeneity within cellular expression patterns, and uncovering clues that may eventually lead to the development of more targeted and personalized therapies. Additionally, the increasing availability of whole tissue cellular atlases in both health and disease as a result of scRNA-seq studies provides an important resource to better understand complicated molecular signaling patterns and events that are similar and different between human diseases.

Keywords

scRNA-seq · Virus · Infection · Kidneys · Pancreas

16.1 Introduction

Single-cell RNA sequencing (scRNA-seq), a technology with increasing impact, has advanced our understanding of the development and function of organ systems and the pathophysiology of diseases that affect them. scRNA-seq is built on the foundation provided by conventional bulk RNA sequencing (RNA-seq) and is fundamentally similar, differing only in the power of its resolution. Bulk RNA-seq, a powerful whole transcriptomic tool, has provided major insights into human diseases by elucidating differences in gene expression patterns at the level of tissue or cell populations. Bulk RNA-seq, however, suffers from a lack of sensitivity to gene expression from rare cell types or cell states which can be lost in the average of the dominant cells, potentially masking important contributors to disease pathogenesis. As scRNA-seq becomes more widely applied, more novel types and subtypes of cells are being discovered, further underscoring the need to increase the resolution at which diseases are studied.

S. Garcia · E. Der
Sue Golding Graduate Division of the Albert Einstein College of Medicine, Bronx, NY, USA

C. Putterman (✉)
Division of Rheumatology, Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA

Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel

Research Institute, Galilee Medical Center, Nahariya, Israel
e-mail: chaim.putterman@einstein.yu.edu

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_16

scRNA-seq has developed rapidly in recent years; from the original tube-based technology, several platforms have emerged. Common to all platforms is the need to generate a single-cell suspension which is typically accomplished by gentle enzymatic digestion. Early scRNA-seq used flow cytometry (FACS) to separate cells of interest into standard 96- or 384-well plates. Later, advances in microfluidics allowed more sophisticated capture using microfluidic devices, such as integrated fluidic chips. Most recently, platforms using microfluidic oil droplets to capture cells including 10x Genomics Chromium, DropSeq, and inDrop have become the gold standard, due to the massive potential throughput offered by these techniques [1, 2]. Concurrently, advances in scRNA-seq chemistry has allowed for individual transcripts to be barcoded with a unique molecular identifier (UMI), providing an absolute count of transcripts per cell instead of a relative abundance. Droplet-based scRNA-seq runs typically generate thousands of single-cell transcriptomic profiles, which can be mined for gene expression patterns to reveal signatures that may have otherwise been overlooked.

This chapter demonstrates the power of scRNA-seq to advance the understanding of human diseases by highlighting some recent applications of this technology. While the focus is on the kidney, pancreas, and viral infections, many other diseases and organs have been studied using scRNA-seq. For instance, scRNA-seq approaches have provided important new insights into mechanisms of disease affecting the lung and liver, and those involved in neoplasia. Additionally, studies of healthy organs and normal organogenesis have revealed previously unknown cellular heterogeneity, states, and interactions.

16.2 Kidney

The kidney is a homeostatic organ with several important physiologic functions including blood filtration, and consists of several distinct cell types typically classified by their function and location [3]. scRNA-seq of kidney tissue initially was focused on exploring the heterogeneity of tumor cells in renal cell carcinoma, but more recently has expanded to other human renal diseases, including lupus nephritis (LN) and chronic kidney disease (CKD). A major undertaking recently has been the generation of healthy renal tissue cell atlases, which can be used as a reference point in the comparison of renal cell profiles from disease states. Park et al. applied droplet-based scRNA-seq to healthy adult mouse renal cells discovering novel transition states between cell types and revealing a novel population of collecting duct cells [4]. The authors additionally identified a population of cells expressing both intercalated cell and principal cell markers, indicating that these cells exist in a transitional continuum. When transcriptional profiles of renal cells from a mouse model of fibrosis analogous to CKD was compared to controls, Park et al. found that the balance in renal tissue was shifted toward principal cells. Since intercalated cells are key in proton secretion through ATP-dependent proton pumps, their decreased number could explain the acid buildup contributing to the metabolic acidosis seen in mouse models and patients with CKD [4]. This finding was extended to the human disease in which CKD was associated with an increase in principal cells as compared to intercalated cells. Park et al. further used scRNA-seq to categorize the specific cellular populations expressing susceptibility genes identified by GWAS studies for various renal diseases, including tubular acidosis and nephrotic syndrome. From this analysis it was hypothesized that there is a cell-type-specific contribution of these genes in each disease.

Sivakamasundari et al. similarly created a single-cell atlas of the human kidney [5]. This analysis revealed novel cell populations, and markers for previously known cell types similar to those identified by Park et al. in mice. Sivakamasundari et al. tracked the expression of known LN-associated genes, such as *NOTCH4*, which in healthy conditions was highly expressed in endothelial cells of the ascending and descending vasa recta. This observation localizes the cellular compartment responsible for the

phenotype associated with *NOTCH4* LN susceptibility alleles. Interestingly, they also reported expression of albumin in proximal tubular cells, raising questions about the source of urinary albumin which is characteristic of many renal diseases but which is often attributed to a loss of function in the glomerular filtration barrier rather than reflecting tubular damage and proteinuria from a tubular source.

LN has been a focus of several studies using scRNA-seq. Using an integrated fluidic chip platform, we recently characterized renal and skin biopsies from LN patients [6]. This analysis revealed an interferon response signature that correlated with clinical parameters, including histological features such as glomerular IgG deposition, and was associated with poor response to conventional treatment. The prognostic importance of an interferon signature was validated in a separate cohort using a higher throughput platform. Moreover, a fibrotic signature in tubular cells was found which predicted response to treatment at 6 months post biopsy [7]. This signature indicated the early activation of a pro-fibrotic cascade in tubular cells in LN, which in many cases appeared before fibrosis in the tissue was visible by traditional histological evaluation. The increased number of cells analyzed and the larger patient cohort in this latter study allowed for biopsy class specific comparisons, which revealed several pathways which may contribute to the histological differences in the disease and provide a potential molecular basis for the biopsy classification system. For instance, tubular cells from biopsies of proliferative class disease (class III or class IV) had upregulated TNF signaling when compared to membranous (class V) or mixed class (class III/IV + V) biopsies. Arazi et al., as part of the same consortium but in a separate cohort, studied LN patient biopsies using a plate-based CEL-SEQ2 approach on sorted populations of leukocytes. They presented an immune landscape of the LN kidney and were able to identify several subsets of immune cells including T follicular helper regulatory cells, myeloid cells, NK cells, and B cells with both pro-inflammatory and inflammation resolving interferon signatures [8].

16.3 Pancreas

The pancreas is comprised of cells which have many important exocrine and endocrine functions. Acinar cells are the main exocrine cells of the pancreas, the main function of which is the secretion of digestive enzymes into the small intestine. The islets of Langerhans are micro-organs made up of five different endocrine cell types whose primary role is to regulate blood-glucose homeostasis by hormone secretion. Previously, gene expression in the islets of Langerhans (studied as whole islets) had been probed by bulk sequencing, providing essential insights into pancreatic function and diseases, such as type II diabetes (T2D). Nevertheless, despite the important discoveries made using bulk RNA-seq methods, global gene expression patterns are only representative of the most abundant cells found in the pancreas or islets [9]. This results in ambiguity as to the cellular source of the signals identified and does not account for the expression of rare cell types which may be critical for pathogenesis. Having a more complete inventory of different cell types within the pancreas along with individual expression patterns can provide major insight into pancreatic biology and diseases affecting this organ such as diabetes, cancer, and pancreatitis [10].

In vitro and animal models have long been used to study and advance our understanding of pancreatic physiology and pathophysiology. scRNA-seq allows researchers to examine the expression profiles of cells in vitro in parallel with in vivo, while also performing comparisons between different species. Xin et al. characterized the transcriptomic profiles of alpha and beta islet cells from non-diabetic and T2D human as well as mouse pancreas. Interestingly, there were 245 genes differentially expressed between non-diabetic and T2D islet cells, almost all of which were previously unreported. Moreover, although

they found a high correlation among the species among homologous genes, about 15% of genes were species-specific highlighting a drawback in using mouse models of complex human diseases [11]. Wang et al. also used scRNA-seq to explore T2D, and compared expression patterns between children and adults with and without T2D. Interestingly, they found that pancreatic islet cells from patients with T2D resembled those found in children, which they believe indicates a partial dedifferentiation of the islet cells during disease pathogenesis [12].

scRNA-seq has also facilitated transcriptomic profiling of the rarer cell types of the islet of Langerhans, such as delta and epsilon cells [13]. Of the five cell types of the islet of Langerhans, the most prevalent are the glucagon producing alpha cells and the insulin producing beta cells which together account for approximately 75–85% of the islet [14]. The remaining cell types: delta cells, epsilon cells, and pancreatic polypeptide cells secrete primarily somatostatin, ghrelin, and pancreatic polypeptide, respectively. One of scRNA-seq's strengths is the ability to study rare cell types. Segerstolpe et al. found when specifically examining the rare cell types of the islet that delta cells expressed ghrelin and leptin receptors and epsilon cells expressed receptors for a diverse range of molecules including neurotransmitters, endorphins, and glycoproteins. Together, these results suggest a role for delta and epsilon cells as islet sensor cells. Further, they found expression of MHC class II molecules in acinar cells, indicating a role for pancreatic exocrine cells as antigen presenting cells relevant to immune interfacing, which could have implications for type I diabetes (T1D) pathogenesis [15].

In addition to insights into pathogenesis, scRNA-seq offers potential to monitor and discover biomarkers to facilitate early diagnosis and intervention. In the context of T1D, diagnosis in children typically only occurs after there are detectable levels of autoantibodies and the immune mediated destruction of islet cells has already progressed. A longitudinal study of T1D using scRNA-seq to monitor PBMC expression profiles in children and their matched controls found upregulated genes including *IL32* well before autoantibodies were detectable [16]. Another study used scRNA-seq in evaluating stem cell differentiation as a potential therapy for T1D and found that WNT signaling pathways may induce proliferation and differentiation of a pluripotent stem cell population to replace damaged cells in the pancreas [17]. Similarly, in adult T2D Wang et al. found that the sonic hedgehog signaling pathway was active in a small population of proliferating alpha islet cells, providing a potential novel pathway to target in therapy.

16.4 Viral Infections

In viral diseases, understanding the heterogeneity of cell populations is key given that viruses have evolved to exploit cellular mechanisms for their own benefit. Better characterizing the cells that are the target of infection by scRNA-seq technology has led to a better understanding of the mechanisms by which viruses replicate and spread. Moreover, scRNA-seq provides a unique means of monitoring the change in transcriptome before, during, and after viral exposure, while offering a new way to identify novel targets of inhibition, treatment, and vaccination [18]. Indeed, scRNA-seq allows for single-cell resolution of both the cellular transcriptome as well as the specific virus that infected that cell. scRNA-seq has already been applied to a variety of viruses which infect humans including influenza, Zika, West Nile, and HIV [19–22].

16.5 HIV

HIV, a lentivirus, has posed formidable challenges to researchers since its appearance on the global stage. While much has been discovered and advances in treatment have vastly improved the prognosis and quality of life in HIV-infected patients, much also remains to be learned. Recent

studies using scRNA-seq have focused on better characterizing HIV permissive cells, the process of HIV latency, and creation of viral reservoirs.

The success of a virus depends on its ability to inhibit cellular defenses and exploit cellular machinery to replicate and effectively infect an organism. The interplay between cellular susceptibility and the viral ability to replicate determines the permissiveness of a cell [23]. Studies have indicated that this permissiveness is highly variable as evidenced by the finding that not all CD4+ T cells are equally permissive to HIV infection, both between patients and within the same patient [24]. This difference in permissiveness varies across tissue, cell line, activation state, and proliferation state. More activated and proliferating T cells appear to have a higher permissiveness to HIV infection; accordingly, the activation marker CD25 is a classic biomarker for HIV permissiveness [24]. While the existence of a permissiveness spectrum is generally accepted, the mechanism behind differential sensitivity to infection has not yet been fully explored. Rato et al. explored the basis of permissiveness using scRNA-seq analysis of non-infected CD4+ T cells patients with both high and low permissiveness (as determined by previous studies with cells isolated from these individuals). They found that a continuum of cell states mediated by T-cell receptor activation was strongly linked to permissiveness. Further, they identified surface proteins which were linked with infection permissiveness, including CD25, CD298, CD63, and CD317. A population expressing all of these markers showed an enrichment of HIV infection of up to 28-fold as compared to the negative population. FACS purification of these susceptible cells indicated downregulation of interferon-induced genes and the antiviral restriction factors *BST2*, *APOBEC3G*, *MX2*, *SAMHD1*, and *TRIM5* as potential mechanisms of increased permissiveness. This study identified biomarkers for cells that are more permissive to HIV, and validated increased cellular activation states as being an indication of permissiveness [24].

In a latent state, HIV is known to establish reservoir cells which are infected with HIV but do not actively produce viral particles. These latent cells have a low viral load, and therefore are not easily detected and can remain unnoticed without inducing patient symptoms. Efforts are being made to eliminate these reservoirs of latent cells, for they often persist despite treatment with anti-retroviral therapy or latency reversing agents and are often responsible for rebound infection [25]. Latent cells exist as multiple subtypes and during different stages of differentiation; hence, elucidating specific gene expression patterns in host cells might provide clues into the viral replication processes that are being inhibited in latent reservoir cells.

Bradley et al. used a single-cell approach to explore the expression of viral and host cell genes in latently infected cell lines and primary cells. They found that latently infected cellular subsets occur in many cellular environments, and in several subsets of T cells from naive to central memory T cells. Additionally, Bradley et al. reported that a specific host transcriptional signature was associated with the latent state, and that latency induction was most prevalent in cells with higher proliferation potential or cells returning from an activated state to a resting state [26]. This finding was confirmed by Golumbeanu et al. who used human primary T cells and scRNA-seq to study gene expression signatures in latent cells, revealing a gene expression pattern including *IL32*, *GAPDH*, and *CD96* associated with latency [27]. Thus, scRNA-seq has provided important insights into gene expression of both the host cells and the viruses themselves, allowing investigators to recognize patterns that may determine disease progression and perhaps facilitate the development of more directed and effective therapeutic strategies to target specific cells.

16.6 Zika Virus

Zika virus is an RNA flavivirus that is prevalent in certain parts of the world including Africa and the Americas. Clinically the presentation can be relatively mild, often resembling flu-like symptoms. However, Zika is thought to hijack autophagic processes, subsequently leading to

abnormal fetal development and microcephaly in babies born to pregnant mothers who are infected with the Zika virus [28]. Despite the Zika epidemic of 2015, Zika biology remains mostly unknown, including key mechanisms underlying viral entry and disease.

Two recent studies have used scRNA-seq to investigate Zika virus pathogenesis. Nowakosky et al. employed scRNA-seq to characterize the gene expression profile of multiple cell types of the developing human cortex to identify Zika susceptible cell types. They identified *AXL,* a tyrosine kinase receptor that transmits signals from the extracellular matrix to the cytoplasm, as a highly expressed gene in astrocytes, microglia, and endothelial cells. Moreover, increased *AXL* expression was also found in radial glial cells (cortical neuron progenitors) in the ventricular zone and neural stem cells, suggesting it as a candidate viral entry receptor [19]. A second study used scRNA-seq to investigate neurodevelopment effects of Zika infection on cell lines of human spinal cord neuroepithelial stem cells and radial glial cells. They also confirmed previous findings in that there was a higher expression of *AXL* in neuroepithelial stem cells and radial glial cells, further supporting *AXL* as a candidate viral entry receptor [20].

16.7 Influenza

While influenza virus is a commonly occurring virus in the general population, it remains a major cause of morbidity and mortality among immune-compromised groups such as infants, the elderly, and immunosuppressed individuals. Even though influenza is common, intracellular viral replication and the variation in host responses across different cell types remain uncharacterized. Using scRNA-seq, Steuerman et al. investigated the heterogeneity of lung cell responses to influenza infection in vivo. They analyzed both the host and viral transcriptomes in individual cells, and further compared exposed but uninfected cells (bystander cells) to host cells infected by influenza. Epithelial cells are generally understood to be the main cell type infected by influenza. In this study, however, they found that at lower viral loads non-epithelial cell types, including endothelial cells, NK cells, and macrophages, were infected as well. While infected cells showed some evidence for a specific antiviral type I interferon response, bystander cells were more heterogeneous in their interferon expression pattern which was accompanied by a more effective antiviral response. This finding suggests the importance of heterogeneity in cell-type-specific responses. Additionally, Steuerman et al. found upregulation of mitochondrial suppression genes in influenza-infected cells, pointing to mitochondrial suppression as another possible therapeutic target [21].

16.8 Conclusion

We have provided here an overview of some of the impressive strides in the understanding of complex disease pathogenesis accomplished using scRNA-seq. scRNA-seq is quickly facilitating the discovery of previously unknown cellular subtypes and cell states, revealing important contributions of overlooked cell types and subsets in human diseases. The increasing throughput of droplet-based platforms and sensitivity of improving chemistry is rapidly cementing scRNA-seq as an invaluable asset for the study of human diseases.

In this chapter, we discussed some of the specific uses of scRNA-seq in exploring viral infections and diseases of the kidney and pancreas. This review, however, is by no means exhaustive, and indeed this technology has advanced the study of pulmonary and cardiac diseases, transplant immunology, cancer, and many others as well. Nevertheless, the above reviewed studies do illustrate the utility and resolution of scRNA-seq in understanding exact cellular compositions, discovering heterogeneity within cellular expression patterns, and uncovering clues that may eventually lead to the development of more targeted and personalized therapies. Additionally, the increasing availability

of whole tissue cellular atlases in both health and disease as a result of scRNA-seq studies provides an important resource to better understand complicated molecular signaling patterns and events that are similar and different between human diseases.

References

1. Der E, Suryawanshi H, Buyon J, Tuschl T, Putterman C (2019) Single-cell RNA sequencing for the study of lupus nephritis. Lupus Sci Med 6:e000329
2. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z et al (2019) Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. Mol Cell 73(1):130–142.e5
3. Kriz W, Bankir L (1988) A standard nomenclature for structures of the kidney. The Renal Commission of the International Union of Physiological Sciences (IUPS). Kidney Int 33(1):1–7
4. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M et al (2018) Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science 360(6390):758–763
5. Sivakamasundari V, Bolisetty M, Sivajothi S, Bessonett S, Ruan D, Robson P (2017) Comprehensive cell type specific transcriptomics of the human kidney. bioRxiv 238063
6. Der E, Ranabothu S, Suryawanshi H, Akat KM, Clancy R, Morozov P et al (2017) Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. JCI Insight 2(9):e93009
7. Der E, Suryawanshi H, Morozov P, Kustagi M, Goilav B, Ranabathou S et al (2019) Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. Nat Immunol 20(7):915–927
8. Arazi A, Rao DA, Berthier CC, Davidson A, Liu Y, Hoover PJ et al (2019) The immune cell landscape in kidneys of patients with lupus nephritis. Nat Immunol 20(7):902–914
9. Ebrahimi A, Jung MH, Dryfuss JM, Pan H, Sgroi D, Bonner-Weir S, Weir GC (2017) Evidence of stress in β cells obtained with laser capture microdissection from pancreases of brain dead donors. Islets 9 (2):19–29
10. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E et al (2016) A single-cell transcriptome atlas of the human pancreas. Cell Syst 3(4):385–394.e3
11. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C et al (2016) RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab 24 (4):608–615
12. Wang YJ, Schug J, Won K-J, Liu C, Naji A, Avrahami D et al (2016) Single-cell transcriptomics of the human endocrine pancreas. Diabetes 65(10):3028–3038
13. Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieux C et al (2016) Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep 17(2):178–187
14. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren P-O, Caicedo A (2006) The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 103 (7):2334–2339
15. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X et al (2016) Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab 24(4):593–607
16. Kallionpää H et al (2019) Early detection of peripheral blood cell signature in children developing β-cell autoimmunity at a young age. Diabetes 68:2024. https://doi.org/10.2337/db19-0287
17. Sharon N, Vanderhooft J, Straubhaar J, Mueller J, Chawla R, Zhou Q et al (2019) Wnt signaling separates the progenitor and endocrine compartments during pancreas development. Cell Rep 27 (8):2281–2291.e5
18. Cristinelli S, Ciuffi A (2018) The use of single-cell RNA-Seq to understand virus–host interactions. Curr Opin Virol 29:39–50. https://doi.org/10.1016/j.coviro.2018.03.001
19. Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR (2016) Expression analysis highlights AXL as a candidate Zika virus entry receptor in human neural stem cells. Cell Stem Cell 18(5):591–596
20. Onorati M, Li Z, Liu F, Sousa AMM, Nakagawa N, Li M et al (2016) Zika virus disrupts phospho-TBK1 localization and mitosis in human neuroepithelial stem cells and radial glia. Cell Rep 16(10):2576–2592
21. Steuerman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E et al (2018) Dissection of influenza infection in vivo by single-cell RNA sequencing. Cell Syst 6(6):679–691.e4
22. Tsioris K, Gupta NT, Ogunniyi AO, Zimnisky RM, Qian F, Yao Y et al (2015) Neutralizing antibodies against West Nile virus identified directly from human B cells by single-cell analysis and next generation sequencing. Integr Biol 7(12):1587–1597
23. Rausell A, Muñoz M, Martinez R, Roger T, Telenti A, Ciuffi A (2016) Innate immune defects in HIV permissive cell lines. Retrovirology 13(1):43
24. Rato S, Rausell A, Muñoz M, Telenti A, Ciuffi A (2017) Single-cell analysis identifies cellular markers of the HIV permissive cell. PLoS Pathog 13(10): e1006678
25. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD et al (2015) Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. J Infect Dis 212(9):1361–1365

26. Bradley T, Ferrari G, Haynes BF, Margolis DM, Browne EP (2018) Single-cell analysis of quiescent HIV infection reveals host transcriptional profiles that regulate proviral latency. Cell Rep 25(1): 107–117.e3
27. Golumbeanu M, Cristinelli S, Rato S, Munoz M, Cavassini M, Beerenwinkel N et al (2018) Single-cell RNA-seq reveals transcriptional heterogeneity in latent and reactivated HIV-infected cells. Cell Rep 23 (4):942–950
28. Plourde AR, Bloch EM (2016 Jul) A literature review of Zika virus. Emerg Infect Dis 22(7):1185–1192

Sayra Garcia is PhD graduate student in the Sue Golding Division of the Albert Einstein College of Medicine.

Evan Der is PhD graduate student in the Sue Golding Division of the Albert Einstein College of Medicine.

Chaim Putterman is Professor of Medicine and Microbiology & Immunology, and Chief of the Division of Rheumatology at the Albert Einstein College of Medicine and Montefiore Medical Center (Bronx, New York). Dr Putterman's major research interests are in the field of immunology and autoimmune diseases, and specifically the identification and characterization of novel mechanisms, biomarkers, and treatment approaches to immune mediated nephritis, neuropsychiatric lupus, and systemic lupus erythematosus. Dr Putterman's laboratory has received funding from the National Institute of Health, Alliance for Lupus Research, Lupus Research Institute, Arthritis Foundation, Boehringer Ingelheim, and Biogen Idec, and he has published more than 200 articles and book chapters in the medical and scientific literature. He was elected to the American Society of Clinical Investigation, and is a member of the American College of Rheumatology and the American Association of Immunologists. Dr Putterman has served on multiple NIH and international study sections, and is currently a member of several editorial boards.

Single-Cell Sequencing in Human Genital Infections

17

Reema Singh

Abstract

Human genital infections are one of the most concerning issues worldwide and can be categorized into sexually transmitted, urinary tract and vaginal infections. These infections, if left untreated, can disseminate to the other parts of the body and cause more complicated illnesses such as pelvic inflammatory disease, urethritis, and anogenital cancers. The effective treatment against these infections is further complicated by the emergence of antimicrobial resistance in the genital infection causing pathogens. Furthermore, the development and applications of single-cell sequencing technologies have open new possibilities to study the drug resistant clones, cell to cell variations, the discovery of acquired drug resistance mutations, transcriptional diversity of a pathogen across different infection stages, to identify rare cell types and investigate different cellular states of genital infection causing pathogens, and to develop novel therapeutical strategies. In this chapter, I will provide a complete review of the applications of single-cell sequencing in human genital infections before discussing their limitations and challenges.

Keywords

Single-cell sequencing · Bioinformatics · Genital Infections · Sexually transmitted infections · Urinary tract Infections

17.1 Introduction

Genital Infections (GIs) are one of the most common problems worldwide, which are commonly transmitted via sexual contacts (vaginal, anal, and oral), mother to fetus or mother to child during birth, and imbalance in the vaginal microbiota [1]. It would not be wrong to classify GIs into three different categories i.e., Sexually Transmitted Infections (STIs) [2], Urinary Tract Infections (UTIs), [3] and Vaginal Infections (VIs) [4]. STIs are gonorrhea, chlamydia, syphilis, genital herpes, genital warts, Acquired Immunodeficiency Syndrome (AIDS), and hepatitis B. Bacterial vaginosis, trichomoniasis, and candidiasis are the commonly known VIs and UTIs, respectively, collectively called as vulvovaginitis. UTIs are described as uncomplicated and complicated. Uncomplicated UTIs mostly affect the healthy individuals whereas complicated UTIs are linked with factors that compromise the host defense [5]. GIs are caused by different types of bacteria,

R. Singh (✉)
Department of Biochemistry, Microbiology and Immunology, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada

Vaccine and Infectious Disease Organization-International Vaccine Centre, Saskatoon, SK, Canada
e-mail: res498@mail.usask.ca

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_17

viruses, and parasites, and associated with infertility, acute illnesses, and complications in the upper genital part. Most curable GIs are caused by bacteria belonging to the gram-negative strain category (*Neisseria gonorrhea*, *Chlamydia trachomatis*, *Escherichia coli*, *Treponema pallidum, Haemophilus ducreyi, and Klebsiella granulomatis* (also known as *Calymmatobacterium granulomatis*) [6]. While virus-mediated infections such as genital herpes, hepatitis B, AIDS, and genital warts are incurable. The complete information about each infection including symptoms and global estimates is given in Table 17.1.

Several difficulties are associated with genital infections. Firstly, life-threatening complications can happen when the untreated GIs disseminate into other parts of the body. For example, untreated gonorrhea, chlamydia, and bacterial vaginosis can cause pelvic inflammatory disease (PID), which can result in long-term reproductive disabilities such as ectopic pregnancy, infertility, and pelvic pain [21]. Furthermore, sexually acquired infections with hepatitis B and certain types of human papillomavirus are considered to be the leading cause of liver cirrhosis, hepatocellular carcinoma, and cervical cancer development, respectively [22, 23]. Some of the STIs induced diseases in the human host are shown in Fig. 17.1. Finally, coinfection with multiple pathogens is another obstacle. A growing body of evidence suggested that both bacterial and viral STIs, as well as bacterial vaginosis (BV), are known to be associated with the probability of HIV acquisition [24–26], and these infections appear to facilitate a vicious cycle of each other's transmission [27–31]. Furthermore, the chances of getting HPV-related cancers are higher in a person living with HIV and HPV coinfection [32].

The treatment therapy for GIs is based on the use of antibiotics such as penicillins, cephalosporins, quinolone, amoxicillin, erythromycin, azithromycin, doxycycline, macrolides, fluoroquinolones, and tetracycline [33]. However, the continuous rise of antibiotic resistance impeded the therapeutical strategies, especially due to the emergence of multidrug-resistant (MDR), extensive drug-resistant (XDR), and pan-drug-resistant (PDR) bacteria, which are resistant against a wide range of available classes of antimicrobial agents [34]. There are several well-defined genetics and mechanistic ways that these disease causing pathogens use to withstand the drug effects i.e., (1) resistance acquired by mutations and horizontal gene transfer and (2) biochemical resistance mechanisms e.g., target modification, antibiotic inactivation and changes in outer membrane permeability [35, 36]. Several studies have reported the increasing cases of antibiotic resistance in genital infection causing pathogens such as *Neisseria gonorrhea*, *Escherichia coli*, *Candida albicans*, Herpes simplex virus (HSV), HIV, etc. [37–42].

Recent advancements in next-generation sequencing technologies and analysis tools to study genomes and transcriptomes in bulk tissues/body fluids have provided new insights into the epidemiology and antimicrobial resistance/susceptibility pattern of genital infection causing pathogens. Reports are published on the applications of single-cell sequencing to investigate various human diseases such as cancer, infectious diseases, brain disorders, etc. [43, 44], however, in the case of GIs, limited information is available. Therefore, the focus of this chapter is to highlight the current applications of single-cell sequencing in human genital infection studies. As single-cell sequencing and analysis methods have already been reviewed in detail elsewhere [45, 46] so I will discuss them briefly in this chapter. Challenges and limitations are also highlighted along with the conclusions.

17.2 Single-Cell Sequencing: Different Types and Analysis

Single-cell sequencing has proven an important milestone in diverse fields of biology such as cancer research, metagenomics, developmental biology, reproductive health, drug discovery, immunology, microbiology, and neurobiology [43, 44]. Until the present, this platform has applied to study cell heterogenicity, host–pathogen interactions, clonal structure, clonal expansion and evolution, copy-number variants (CNVs), target

Table 17.1 Types of genital infections in humans, causative pathogens, symptoms, and global estimates (adapted Passos [7])

Types	Causative pathogens	Symptoms	Global estimates	References
Genital herpes	*Herpes simplex virus*	Genital ulcers	19.2 million	Looker et al. [8]
Syphilis	*Treponema pallidum*	Genital ulcers	19.9 million	Rowley et al. [9]
Gonorrhea	*Neisseria gonorrhoeae*	Urethral discharge, Endocervicitis/pelvic pain, testical pain/ swelling, proctitis, ophthalmia	30.6 million	Rowley et al. [9]
Chlamydia	*Chlamydia trachomatis*	Urethral discharge, Endocervicitis/pelvic pain, testical pain/ swelling, proctitis, ophthalmia	124.3 million	Rowley et al. [9]
Bacterial vaginosis	*Gardnerella vaginalis*, *Mobiluncus* sp., *Mycoplasma hominis*, *Bacteroides*, *Prevotella* sp., *Peptostreptococcus* sp., *Atopobium vaginae*	Vaginal discharge	Limited	Kenyon et al. [10]
Candidiasis	*Candida albicans*	Vaginal discharge	138 million	Cai and Chen [11]
Trichomoniasis	*Trichomonas vaginitis*	Vaginal discharge, vulvar irritation	110.4 million	Rowley et al. [9]
Chancroid	*Haemophilus ducreyi*	Papule	7 million (cases are decreasing)	Steen [12], González-Beiras et al. [13]
Lymphogranuloma venereum	*Chlamydia trachomatis,* serovars L1,L2 and L3	Painless papule	Limited	de Vries [14], Cole et al. [15]
Donovanosis	*Calymmatobacterium granulomatis*	Ulcerative lesions	Limited	O'Farrell [16]
Genital warts	*Human papillomavirus*	Papular or flat lesions	160–289 cases per 100,000 person	Patel et al. [17]
Acquired immune deficiency syndrome (AIDS)	*Human immunodeficiency virus*	Genital sores	36.9 million	UNAIDS 2017 global estimate
Hepatitis B	*Hepatitis B virus*	Dark urine, abdominal pain	257 million	WHO 2015 global estimate
Scabies	*Sarcoptes scabiei*	Genital papular and itching	>200 million	Chandler and Fuller [18]
Urinary tract infection	*Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, Acinetobacter baumannii*	Pain/burning sensation on urination,	92 million	Global Burden of Disease Study 2013 Collaborators [19]
Genital mycoplasmas	*Mycoplasma hominis, M. primatum, M. genitalium, M. spermatophilum, M. penetrans, Ureaplasma urealyticum*	Vaginal itching, pain during sex, burning sensation on urination	Limited	Baumann et al. [20]

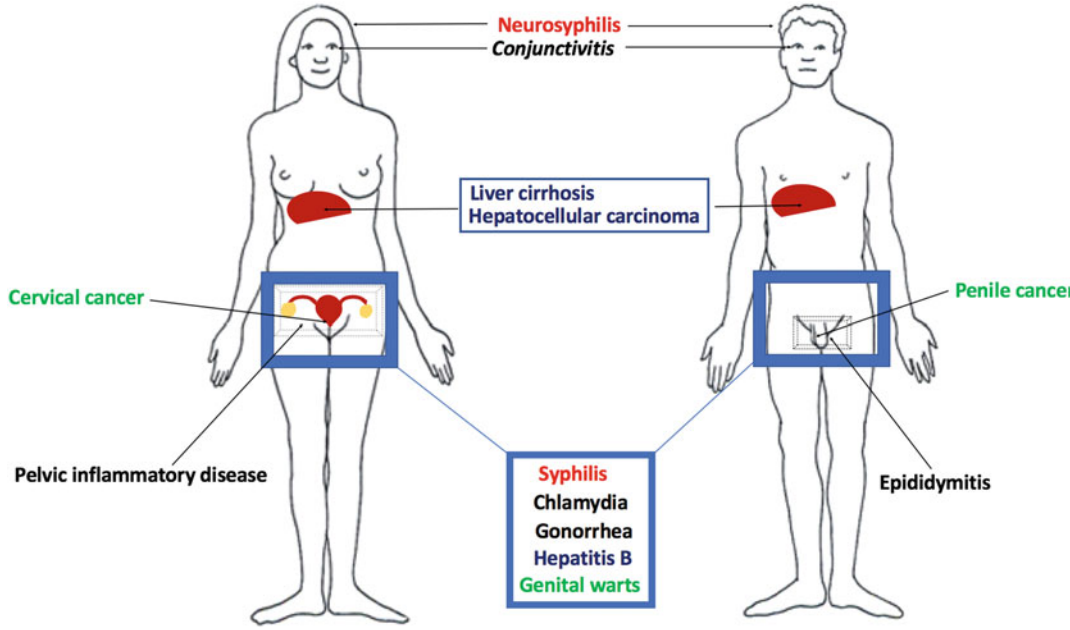


Fig. 17.1 Different types of diseases that occur in the human host when the untreated STIs disseminate to the other parts of the body. Diseases (neurosyphilis, conjunctivitis, pelvic inflammatory diseases, epididymitis, liver cirrhosis, hepatocellular carcinoma, penile cancer, and cervical cancer) and genital infections (syphilis, chlamydia, gonorrhea, hepatitis B, and genital warts), caused by the same pathogens, are highlighted in similar colors. The basic human body drawings are copied from https://www.clipartqueen.com/human-body-diagram.html (these are free to use for academic purpose). Please note—chlamydia (caused by *Chlamydia trachomatis*) and gonorrhea (caused by *Neisseria gonorrhoeae*) are highlighted in the same color because they both can cause conjunctivitis, epididymitis, and pelvic inflammatory disease

therapy, expression profiling, and RNA splicing [47].

To understand biological processes and how they contribute to a particular disorder, it is important to study genetic regulation at all levels including genome, epigenome, protein, and metabolite. Single-cell sequencing provides the oppertunity to construct a multiomics profile that can give insights to link genotypes to phenotypes of the individual cells. The analysis using single-cell sequencing in any biological and/or clinical study starts with a research question, depending on which the most suitable technology is selected and subsequently sequencing is performed. Finally, the resulting raw sequencing data gets analyzed before interpreting the meaningful information from results (Fig. 17.2).

Single-cell sequencing approaches can be described into two categories i.e., low-throughput and high-throughput. Single-cell microscope, single-cell flow cytometry, single-cell imaging, and single-cell PCR (qPCR and RT-PCR) are some of the low-throughput traditional techniques, with a diverse spectrum of applications, which have been frequently used to detect certain markers of single cells until now [48–50]. However, these methods are only able to uncover the gene expression pattern of a very small set of genes. Thus, these limitations can be easily bypassed by using high-throughput modern global analytical technologies available for single-cell sequencing including single-cell genomics, single-cell epigenomics, single-cell transcriptomics, single-cell proteomics, single-cell metabolomics, and single-cell omics (Box

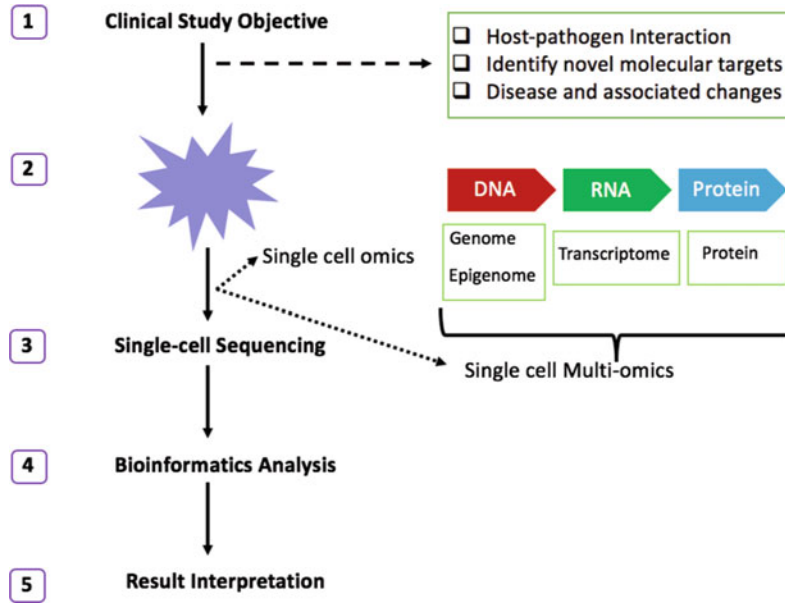


Fig. 17.2 General flowchart showing the single-cell sequencing use in a clinical application. A typical analysis includes the following steps: (1) Selecting an objective, (2) Single-cell isolation, (3) Selection of "omics" technology and Sequencing, (4) Use of bioinformatics software to analyze the data, and (5) Inferring meaningful information

17.1) [48–50]. A variety of wet-lab and dry-lab analysis tools are chosen depending upon the nature of the research hypothesis [49, 51]. For example, if the objective of the study is to understand the cell heterogenicity and/or lineage tracing in early development then single-cell transcriptomics is the most popular choice (Box 17.1). Besides, the complete single-cell sequencing analysis procedure can be divided into three different steps, (1) selecting and isolating individual cells from the bulk population, (2) single-cell sequencing, and (3) downstream computational analysis. The complete single-cell sequencing workflow and several methods (traditional and modern) for cell isolation, sequencing as well as for bioinformatics analysis are explained somewhere else in detail [44, 52].

Box 17.1 Summary of Available Single-Cell Sequencing Technologies and Their Applications

Technologies	Applications
Single-cell genomics	• To resolve variation between individual cells • To study genetic alteration of rare cell types
Single-cell epigenomics	• To understand DNA modification as regulatory epigenetic mark • To understand developmental process
Single-cell transcriptomics	• To identify novel cellular subtypes • Cell heterogenicity analysis and lineage tracing • Identification of diagnostic primers • To detect emergence of resistance clone during chemotherapy • Gene regulatory network construction
Single-cell proteomics	• To study protein–protein interactions • To determine post-translational modifications
Single-cell metabolomics	• To understand the phenotypical variations between cells

Technologies	Applications
Single-cell omics	• To discover novel regulatory mechanisms • To reveal relationship between different omics data types

Multiple challenges exist for the bioinformatics and statistical analysis of single-cell sequencing data, which needs to fix using two important steps i.e., quality check and normalization, before applying any specialized algorithms. To decide whether the data generated from single-cell sequencing is valid for further analysis or not, several factors need to consider during quality check. For instance, removal of contaminated cells, genes that are always present in low quantity and cells with an elevated mitochondrial gene expression can tremendously increase the quality of the raw data [45, 53, 54]. Furthermore, the normalization step is very crucial to eliminate the batch effects generated due to technical variations.

Moreover, a plethora of open-source computational tools have been developed, often free for academic users, and reviewed previously [46, 54–57]. To check the performance of different single-cell RNA-Seq analysis pipeline and complete workflow describing the best-practice recommendations, please see the recently published literature [58, 59]. Besides, a list of these tools is given in Table 17.2.

17.3 Applications of Single-Cell Sequencing to Study Human Genital Infections

Recently, single-cell sequencing technologies have exploited to investigate the bacteria and viruses that cause genital infections. Despite the dominance of bacteria in the STIs, scientists have mainly applied single-cell sequencing to study viral-mediated infections. The most direct applications of single-cell sequencing in genital infections are outlined in Fig. 17.3. In most of the studies reported in this chapter, high-throughput single-cell transcriptomic analysis has utilized. In the following sections, major single-cell sequencing applications are in.

17.3.1 Sexually Transmitted Infections

17.3.1.1 Acquired Immunodeficiency Syndrome (AIDS)

HIV is one of the research areas that is greatly benefited from single-cell sequencing. Until the present, this technology has been widely applied to understand cell heterogenicity, clonal expansion, HIV-1 replication cycle, and expressed gene signature in HIV (Fig. 17.3A).

HIV is transmitted sexually and attacks the $CD4^+$ lymphocytes cells that fight against infections. The final stage of HIV infection (if left untreated) is called AIDS. The virus elects some strategies for its survival in the host. One of these strategies is latent reservoirs, using which the virus does not only persist in the infected person but also reactivate the infection if treatment is stopped. While antiretroviral therapy (ART) is the widely accepted treatment to inhibit the HIV replication in the host, ART has not been able to eradicate the virus. Thus, most single-cell sequencing studies of HIV have applied to research on HIV-infected individuals on ART.

Reactivation of latent reservoirs is one of the leading research focus to study viral persistence in host and to identify novel therapeutical strategies for HIV annihilation. To report the heterogeneous effect of latency reserving agents (LRA) on HIV-1 activation, Yucha et al., developed a microfluidic single-cell-in-droplet PCR assay [70]. The assay was then applied to measure the number of transcriptionally active $CD4^+$ T cells from HIV-1 infected individuals on ART [70]. Furthermore, in another study, working on the Interleukin-7 (IL-7) authors indicated that IL-7 stimulates the spread of latently infected cells by enhancing the residual level of viral production in HIV-infected subjects. Therefore, IL-7 is not a suitable candidate for future therapeutical strategies to eradicate HIV [71].

Table 17.2 List of bioinformatics software and tools available for the single-cell sequencing data analysis

Types	Software/tools	References
Pipelines	Scruff, Granatum, BALDR, ScPipe, SINCERA	Wang et al. [56], Zhu et al. [60], Upadhyay et al. [61], Tian et al. [62], Guo et al. [63]
R-packages	Seurat, Scran, SingleSplice, OncoNEM, batchelor, infercnv, scds, scMerge, scRecover, scater, monocle, MAST, SC3, AUCell, clusterExperiment, splatter, M3Drop,SCnorm, scmap, TSCAN, scDD, BASiCS, slingshot, cellTree, AneuFinder, scfind, cellity, switchde, MetaNeighbor, DEsingle, cicero, BEARsec, singleCellTK, MIMOSA, sincell, slalom, LineagePulse, cellscape, celaref, bayNorm, scFeatureFilter, scTensor, mbkmeans, Melissa, phemd, cellBench, scds, scRecover, scAlign, Oscope	https://www.bioconductor.org
Others	ASAP, SSCC, iS-CellR, ESAT,SC1, bigSCale	Gardeux et al. [64], Ren et al. [65], Patel [66], Derr et al. [67], Moussa and Măndoiu [68], Iacono et al. [69]

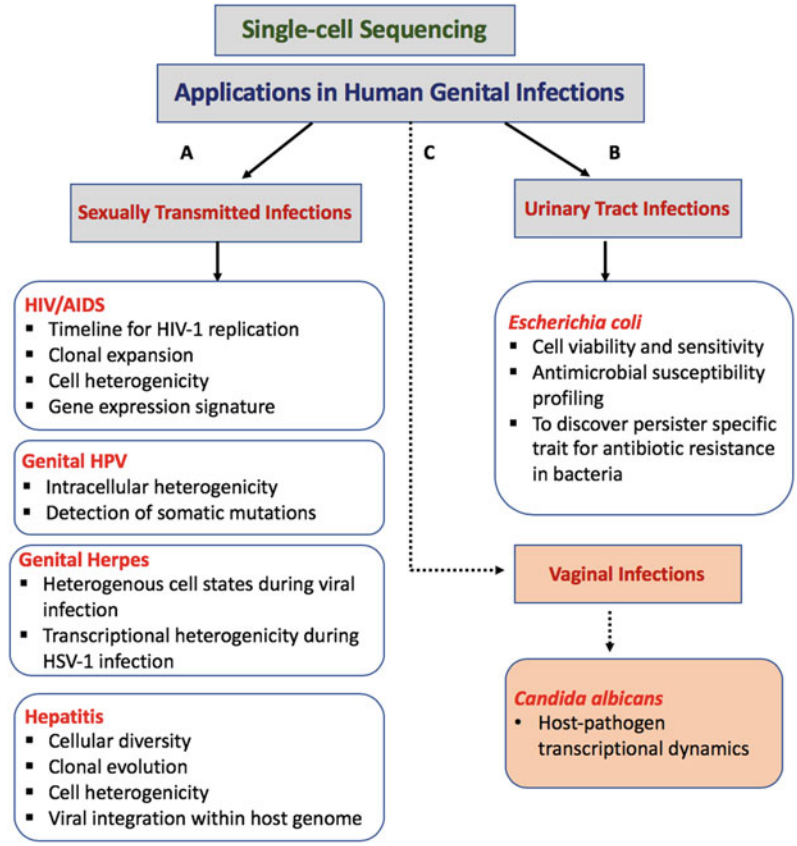


Fig. 17.3 Applications of single-cell sequencing in human genital infections: Sexually transmitted infections (A), Urinary tract infections (B) and Vaginal infections (C). In the case of vaginal infections (dotted lines with orange box), single-cell sequencing technologies have not directly applied to the samples extracted from human vagina, instead, the experiment was performed on samples extracted from mouse bone marrow

Moreover, clonal expansion is one of the key mechanisms using which the HIV-1 reservoir persists in the latently infected cells [72]. An integral part of the latent reservoir, called provirus, reactivates the dormant infection by producing the unspliced RNA. Evidence of this clonal expansion using single-cell sequencing came from a previous study [73]. In this study, authors

developed a method, called CARD-SGS (cell-associated HIV RNA and DNA single-genome sequencing), to detect the proviral expression of HIV RNA in a single cell isolated from individuals with viremia or on long-term suppressive ART. Interestingly, this study identified different cells producing identical RNA molecules which indicates that these molecules could originate either from single infected cells undergoing clonal expansion or multiple cells infected by the same HIV variant before ART. Therefore, HIV-infected cells can perpetually express HIV RNA during suppressive ART through clonal expansion [73].

HIV-1 replication is another important field that has benefited from single-cell sequencing research. Using single-cell FACS-based and quantitative fluorescence microscopy-based analysis, Holmes et al., revealed the timeline of the key events in the HIV-1 replication cycle and measured the gene expression level in infected cells [74]. Also, significant heterogenicity in the HIV-1 replication cycle length and infected cells generate virions for only a few hours are some of the important observations of this study.

Another area in HIV research where single-cell sequencing has been a major asset is cell heterogenicity. Despite the presence of technical and computational challenges, high-throughput single-cell sequencing approaches opened up new perspectives in HIV latency research. Figure 17.4 highlights the types of different experiments that have conducted using high-throughput single-cell sequencing in the HIV research field.

Using a combination of experiments, Cohn et al., (Fig. 17.4a) purified and characterized single reactivated latent cells from HIV-1 infected individuals on suppressive antiviral therapy [75]. Furthermore, by performing single-cell RNA sequencing (scRNA-Seq) on 227 cells (Control/uninfected; 109, LURE purified $Gag^{+}Env^{+}$; 85, YU2 infected; 33) from three subjects, authors demonstrated several conclusions. Firstly, read sequences from LURE purified $Gag^{+}Env^{+}$ mapped both to the human genome and HIV, and recovery of full-length HIV-1 from the reactivated latent cells is possible by using scRNA-Seq. Secondly, hierarchical clustering segregated the gene expression data into three distinct clusters, indicating the heterogeneous nature of the cells. Moreover, genes associated with HIV-1 latency are highly expressed in LURE purified $Gag^{+}Env^{+}$ cells as compared to control cells. Finally, gene set enrichment analysis of eight (among 240) overlapping genes between clustering and differential gene expression analysis are related to significantly enriched biological processes (immune system function), indicated that in response to the pathogen LURE purified $Gag^{+}Env^{+}$ and control cells are different in their gene expression. However, to reveal more differentially expressed genes and pathways contributing to maintaining HIV latency in the host cells, additional confirmation by performing this analysis on more donors are needed.

Working with the $CD4^{+}$ T cells, Golumbeanu et al., (Fig. 17.4b) reported transcription heterogenicity during the HIV reinfection and latency stage. They also found similar in-vivo expression pattern of 134 gene-specific signature identified among untreated, SAHA (vorinostat), and TCR (T-cell receptor)-treated $CD4^{+}$ T cells isolated from HIV+ individuals, a finding that could benefit LRA (latency-reversing agents) identification [76]. Moreover, a second study in which exploring cellular heterogenicity has been a primary focus includes Rato et al., (Fig. 17.4c), suggesting that the main factor of transcriptional heterogenicity at the single-cell level is cellular activation state, which results into diverse levels of HIV permissiveness. Besides, HIV-permissive cell-specific signature is identified, most of which are involved in innate immunity [77]. Furthermore, using single-cell transcriptomics, Bradley et al., (Fig. 17.4d) revealed a specific set of cellular genes that are associated with the HIV expression during latency. They further identified the latency-associated signature and concluded that viral transcript expression during latency is influenced by the host cell transcriptional program [78].

Another study including virologically suppressed HIV individuals by Farhadian et al., (Fig. 17.4e), identified potentially novel myeloid

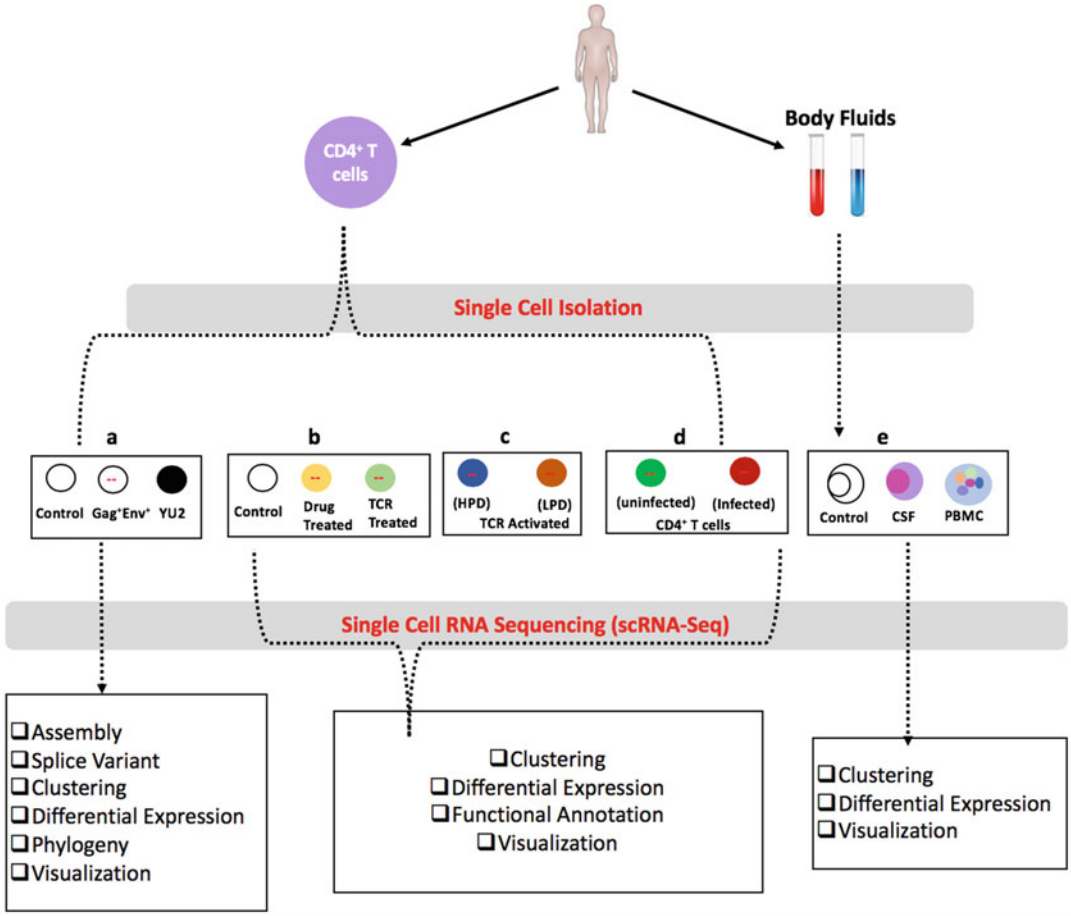


Fig. 17.4 Types of experiments conducted using high-throughput single-cell sequencing in HIV studies: (a) Cohn et al. [75]; (b) Golumbeanu et al. [76]; (c) Rato et al. [77]; (d) Bradley et al. [78]; (e) Farhadian et al. [79] (only single-cell sequencing transcriptomics experiments are reported from all these studies). $gag^{+}Env^{+}$ cells expressing both *env* and *gag*, *YU2* HIV-1_{YU2}, *HPD* high permissive TCR-activated $CD4^{+}$ T cells, LPD low permissive TCR-activated $CD4^{+}$ T cells, *CSF* cerebrospinal fluid, *PBMC* peripheral blood mononuclear cell, *TCR* T-cell receptor

cell populations that are associated with central nervous system (CNS) immune activation by performing single-cell RNA sequencing of blood cells and cerebrospinal fluid (CSF) samples collected at different time points. Furthermore, the gene expression signature present in these myeloid cells intersects with neurodegenerative disease-associated microglia in an animal model. Therefore, indicating a common mechanistic link between neuron injury in HIV and other neurodegenerative disease pathways. The small number of participants is the biggest limiting factor of this study and thus requires a large cohort study to confirm the association between CSF immune cells and markers of neurodegenerative disease [79].

HIV infection impaired the host immune system by attacking the cells such as $CD4^{+}$ T and memory B cells (MBC) that fight against infections. As an example of this, de Armas et al., applied single-cell RT-PCR methods to interrogate the cellular state of MBC from HIV-infected individuals. As a result, gene expression data revealed the overexpression of

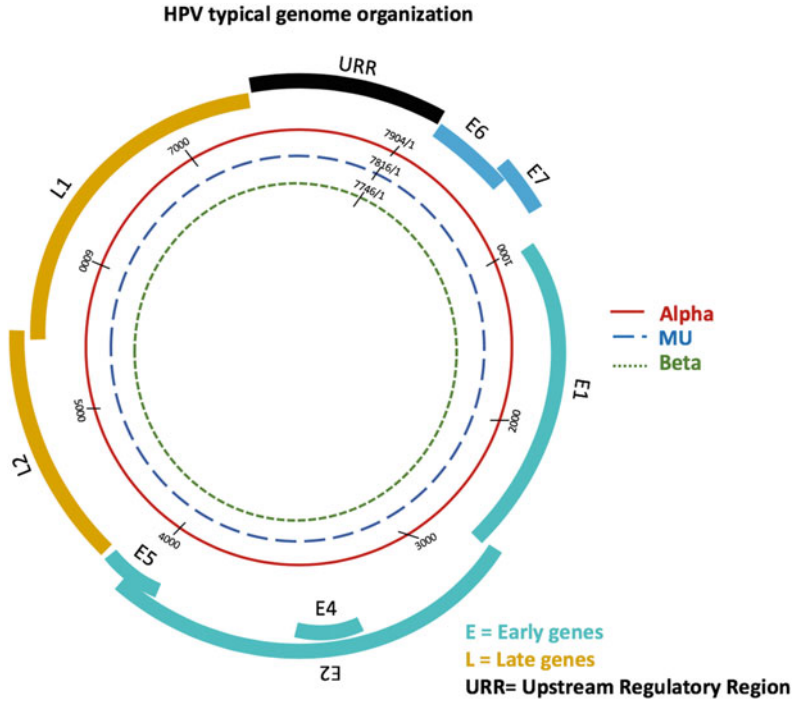


Fig. 17.5 Typical HPV genome organization for an Alpha, Mu, and Beta. (Adapted Doorbar [81]; Harari [82])

PTEN in H1N1-specific MBC from HIV-infected as compared to healthy individuals. Moreover, the authors concluded that as compared to age-matched healthy controls, H1N1-specific MBC demonstrates diverse gene expression signature in HIV, ART suppressed individuals [80].

17.3.1.2 Genital Human Papillomavirus Infection (Genital Warts)

Human papillomavirus (HPV) is transmitted mainly via sexual contacts, but the infected individual's immune system rapidly clear the infection in the majority of cases. There are more than 200 HPV types that have been identified and characterized. Based on their evolutionary analysis, HPV types are categorized into five diverse genera, the majority of which reside in two largest groups (Alpha and Beta) and one smallest group (Mu) [81] (Fig. 17.5). Moreover, depending on their carcinogenicity, HPV types are classified as low-risk and high-risk subtype [83]. Low-risk HPV subtypes are associated with low-grade cervical abnormalities and genital warts, while high-risk HPV subtypes are associated with six anogenital cancers (cervical, penile, anal, vaginal, vulvar, and oropharyngeal) [84]. Few studies have employed single-cell sequencing strategies (Fig. 17.3A) to characterize cell heterogenicity and mutations in human cervical cancer, however, none of them examine genital warts and low-grade cervical abnormalities.

Using HeLa cells, heterogenicity of viral alternative transcripts is revealed by Wu et al. In brief, using a novel platform, called MIRALCS (microwell full-length mRNA amplification and library construction system), single-cell RNA sequencing were performed on 40 HeLa S3 cells to understand the heterogenicity of HPV+ cancer cell lines in gene expression, alternative splicing, and fusion events. Also, at the single-cell level, they classified cells based on cell cycle states and identified that E6 and E7 viral oncogenes were residing in a cluster of the co-expressed gene [85]. Another study in which heterogenicity of HPV status has exposed at a single-cell level include Shen et al., validated the likelihood of single-cell analysis of HPV infection. Moreover,

to verify the coexistence of multiple HPV types, the authors measured the viral load in single cells using qPCR and confirmed the results by performing PCR-based DNA sequencing [86].

Furthermore, Yang et al., investigated the intracellular heterogenicity in tumour cells earlier and after radiotherapy. They performed single-cell whole-genome sequencing on 25 cervical tumour cells including the detection of somatic mutations, virus integration site mapping and clustering. Consequently, they concluded that the HPV integration site in *POU5FIB* might be responsible for radiotherapy resistance [87].

17.3.1.3 Genital Herpes or Genital Ulcer

Herpes Simplex Virus (HSV) is the main cause of genital herpes, a chronic life-long infection in the sexually active population. It belongs to the herpesviridae family of the virus with a linear double-stranded genome ranging from 152 to 155 kb in size. There are two serotypes of this virus i.e., HSV-1 (the main cause of oral lesions) and HSV-2 (the main cause of genital lesions). A growing number of pieces of evidence concluded that genital ulcer cases due to HSV-1 have risen in high-income countries [88, 89].

In the case of HSV, single-cell sequencing has used to highlight the heterogeneous cell states in viral infections (Fig. 17.3A). For instance, in a recent study, Wyler et al. used single-cell transcriptomic analysis to quantify transcripts of human fibroblast at various time points during early HSV-1 infection. Consequently, authors identified cell cycle phases S/G2/M carried more viral transcripts and as compared to cells in the G1 phase provide a more favourable environment to establish the infection [90]. Furthermore, in a separate study, transcriptional heterogenicity has also characterized during HSV-1 infection [91]. By observing differences at the level of viral infection dynamics, gene expression and cellular host response in infected cells, authors concluded that highly infected cells activate multiple developmental pathways due to transcriptional reprogramming [91].

17.3.1.4 Hepatitis

Hepatitis B virus (HBV) causes an acute and chronic infection which transmitted very efficiently through sexual contacts (heterosexual and/or male homosexual contacts). Coinfection with other viral forms i.e., Hepatitis C virus (HCV) and Hepatitis D virus (HDV) have become chronic as well [92]. Like HBC, HCV can be sexually transmitted [93], and both are associated with liver cirrhosis and hepatocellular carcinoma [23].

In the case of hepatitis, single-cell sequencing has proven a powerful tool to study viral integration within the host genome, cellular diversity, clonal evolution, and cell heterogenicity (Fig. 17.3A). In an early study, single-cell RNA viral sequencing was performed to study viral–host interactions during HCV infection to reveal quasispecies diversity in individual cells [94]. Also, the existence of independent evolution at the cellular level indicated by the diverse (in bulk cells population) and unique (in single cells) quasispecies in infected cells.

Furthermore, Eltahla et al., proposed a novel computational pipeline, called VDJPuzzle, to analyze the single-cell transcriptome of flow-sorted Ag-specific $CD8^+$ T cells, and to reconstruct the native TCRαβ [95]. Moreover, to compare the differences in $CD8^+$ T cell exhaustion, Wang et al., performed the reanalysis of the publicly available single-cell sequencing data from chronic hepatitis B (CHB) and Hepatocellular carcinoma (HCC). The authors further confirmed that cell exhaustion exists in both CHB and HCC along with genetic and phenotypic differences [96].

Another study performed single-cell whole-genome sequencing of 96 tumour cells and 15 normal liver cells from patients with HBV-associated hepatocellular carcinoma to gain insight into the intratumor heterogenicity [97]. Also, Chen et al., used single-cell viral capture sequencing to identify tumour heterogenicity from a patient with multifocal HCC without active replication and successfully identified viral integration within the host genome [98].

17.3.2 Urinary Tract Infections

Urinary tract infections are caused by both grams negative and positive bacteria. Uropathogenic *Escherichia coli* (UPEC) is the most frequent cause of UTI in humans, affecting mostly women. Treating UPEC has become challenging due to emerging resistance to the commonly used antibiotics.

Although there are other pathogens such as *Pseudomonas aeruginosa* [99], *Klebsiella pneumonia* [100] that cause UTI, *Escherichia coli* is the only bacteria that have benefited with the single-cell sequencing technologies. In the case of UTIs (caused by *Escherichia coli*), single-cell sequencing has applied to study cell viability and sensitivity, antimicrobial susceptibility profiling, and bacterial persister specific trait for antibiotic resistance (Fig. 17.3B).

A previous study [101] used an infected mouse model to understand the host and bacterial characteristics leading to UTI and recurrent UTI. Duraiswamy et al., developed a simple method to isolate pure IBCs (Intracellular Bacterial Communities) from mouse bladder during acute experimental UTI. Using micro pipetting they isolated the individual IBCs and further verified the sensitivity and purity based on microscopy, gene expression, and culture-based methods. Besides, they found 10^3 viable bacteria in an early IBC (6 h post-infection) [101]. Furthermore, Yang et al., reported the complete procedure to extract IBCs from a mouse infected experimentally in the urinary tract [102].

While antibiotics are the most effective treatment for UTI, recurrent infections (due to antibiotic-resistant bacteria) complicate the situations. Incorrect prescription is one of the key factors involved in the development and spread of antimicrobial-resistant bacterial strains. Certainly, identifying the antibiotic resistance/susceptibility profile of the infecting bacteria could facilitate the use of effective antibiotics. The previous study showed that the susceptibility of an antibiotic can be detected in less than 30 min using single-cell technology (such as imaging, microscope, etc.) [103]. Thereby, these findings further support the development of a point-of-care test to direct the correct UTI treatment.

Bacterial persistence plays a very important role in the evolution of antibiotic resistance and the reoccurrence of infections. Goormaghtigh and Melderen devised a single-cell approach and applied it to analyze the cell persistence in wild-type *E.coli* to ofloxacin in steady-state growth conditions using microfluidics together with fluorescence microscopy. Consequently, after antibiotic treatment, the analysis discovered persister specific traits during recovery [104]. Therefore, this finding could provide valuable insights to tackle the AMR issue in UPEC.

17.3.3 Vaginal Infections

Despite the involvement of several pathogens in vaginal infections, researchers have only applied single-cell sequencing to study *Candida albicans* (a commensal residence of healthy human gastrointestinal and urogenital tract, mainly casing candidiasis). Until the present, single-cell sequencing has applied to study host–pathogen transcriptional dynamics (Fig. 17.3C), however, this technology has not directly applied to study *Candida albicans* extracted from human vaginal samples.

Macrophages play an essential role in detecting, engulfing and destroying a pathogen. In a recent study, single-cell transcriptomics is employed to access gene expression variability between host and *C.albicans* [105]. Using four different infected murine macrophages i.e., infected macrophages with dead, phagocytosed *C.albicans*, infected macrophages with live *C. albicans*, macrophages exposed to *C.albicans* that remained uninfected, and *C.albicans* exposed to macrophages that remained unengulfed, authors examined pathogen interactions with host cells. They further observed a tightly coordinated shift in transcriptional dynamics and revealed bimodality in expression and changes in splicing patterns that may direct infection outcomes. Although this study was performed using bone-marrow-derived macrophages from mice model, this data provides useful insight

that could be applied to study host–pathogen interactions during vaginal infections.

17.4 Limitations and Challenges

While pure bacterial culture is one of the essential requirements to study pathogenic virulence and antibiotic resistance/susceptibility pattern, inability to grow all of the microorganisms in a manner that enough genetic material can be preserved is one of the main challenges in applying single-cell sequencing technologies to study human genital infections [106]. Low amount of nucleic acid content in a single cell and contamination hinders their sequencing [49], especially in the case of clinical samples where microbial culturing is more difficult and requires specialized training. Sexually transmitted infection causing pathogens such as *Chlamydia trachomatis* and *Treponema pallidum* require special culture conditions. For example, *Treponema pallidum* grows slowly and it proliferates only in laboratory animals (rabbits) and *Chlamydia trachomatis* requires specialized level 3 laboratories [106]. Furthermore, other factors such as single-cell lysis, cDNA synthesis and amplification in the case of microorganisms are particularly challenging due to the presence of a rigid cell wall that inhibits lysis [107]. Thus, appropriate cell isolation strategies need to be selected carefully.

Another significant challenge in single-cell sequencing is to correct technical and biological variations during expression analysis which raises substantial questions throughout the computational analysis. Differences in tissue sampling and handling, and during sequencing when the cells from one condition are cultured and sequenced separately than another are the potential causes of technical variations. Furthermore, the key reasons for the biological variations are the random biochemical reactions, transcriptional bursting causing stochastic gene expression, and uneven genome coverage due to stochastic primer binding and chimeric fragments [108].

Artifacts introduced during library preparation and PCR amplification are responsible for artificial mutations and sequencing bias. These challenges are further exacerbated by the presence of missing gene expression values in the case of some cells which could lead to spurious results, including incorrect data integration and interpretation. Without proper study designs, results can be significantly affected by batch effects. Thus, correcting technical and biological variations is crucial for accurately analyzing the single-cell sequencing data. Finally, it also quite challenging to make sure that only data from single and live cells are included in the downstream analysis so that data from compromised cells do not negatively affect the result interpretation [53].

Moreover, another major limitation is the lack of scRNA-Seq computational analysis pipelines with a graphical user interface that could be easily accessible or useful for users with limited bioinformatics skills.

Finally, the key challenge is related to resource and infrastructure requirements. The emergence of high-throughput single-cell sequencing technologies is responsible for an exponential increase in the amount of data. Efficient approaches are required to store, distribute and analyze this data. Besides, along with storage, data confidentiality, security, and integrity are equally important and needs to be handled properly

17.5 Conclusions

This chapter introduced single-cell sequencing technologies and its applications in human genital infections i.e., sexually transmitted, urinary tract, and vaginal infections. Although in the field of viral causing sexually transmitted infections single-cell sequencing has paved the way to a broad range of applications, its use in bacterial causing genital infections is still in infancy. These technologies that are used to analyze genital infections from a single-cell perspective can deliver valuable insights into the discovery of acquired drug resistance mutations, cell to cell variations, deeper interrogation of cellular states and identify rare cell types, where the field of bulk analysis present difficulties. Furthermore, several

limitations such as complications in microbial culturing, biological and technical variations, unavailability of user-friendly computational tools, and proper storage slow down the research.

Glossary

Antimicrobial Resistance The ability of a microorganism to withstand the effect of an antibiotic.

Antiretroviral Therapy A treatment to control AIDS with the use of a combination of antiretroviral agents.

Chronic hepatitis B When the host body's immune system is not able to clear the infection and the HBV remains in the blood and liver.

Clonal expansion The process of unpredictable increase in the number of cells.

Extensive Drug Resistance The ability of a microorganism to withstand the effect of at least one agent in all but two or fewer antimicrobial categories.

Genital Infection Infection related to human reproductive organs such as vulva, vagina, cervix, urethra, penis, fallopian tube, pelvic, testicles, scrotum, epididymis.

Hepatocellular carcinoma Type of liver malignancy that occurs in people with chronic liver diseases caused by hepatitis B and C infection.

HIV permissiveness The feature of a cell that supports the growth of HIV.

Horizontal gene transfer The process of transfer genetic material between unrelated organisms.

Latent reservoir These are the collections of infected immune cells that go into a dormant state and not actively produce new HIV.

Multidrug Resistance The ability of a microorganism to withstand the effect of at least one agent in three or more antimicrobial categories.

Pan-drug Resistance The ability of a microorganism to withstand the effect of all agents in all antimicrobial categories.

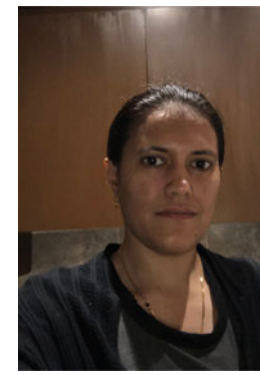
References

1. Mendling W (2016) Vaginal microbiota. Adv Exp Med Biol 902:83–93
2. Smith L, Angarone MP (2015) Sexually transmitted infections. Urol Clin North Am 42(4):507–518
3. Karam MRA, Habibi M, Bouzari S (2019) Urinary tract infection: pathogenicity, antibiotic resistance and development of effective vaccines against uropathogenic Escherichia coli. Mol Immunol 18:56–57
4. Mashburn J (2012) Vaginal infections update. J Midwifery Womens Health 57(6):629–634
5. Flores-Mireles AL, Walker JN, Caparon M et al (2015) Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13(5):269–284
6. Chen MY, Tabrizi SN (2015) Challenges to the management of curable sexually transmitted infections. BMC Infect Dis 15:337
7. Passos MR (ed) (2017) Atlas of sexually transmitted diseases: clinical aspects and differential diagnosis. Springer, Cham
8. Looker KJ, Magaret AS, Turner KME et al (2015) Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. PLoS One 10(1):e114989
9. Rowley J, Vander Hoorn S, Korenromp E et al (2019) Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. Bull World Health Organ 97(8):548
10. Kenyon C, Colebunders R, Crucitti T (2013) The global epidemiology of bacterial vaginosis: a systematic review. Am J Obstet Gynecol 209(6):505–523
11. Cai YM, Chen XS (2018) Concern regarding bias in estimates of recurrent vulvovaginal candidiasis. Lancet Infect Dis 18(11):1180
12. Steen R (2001) Eradicating chancroid. Bull World Health Organ 79:818–826
13. González-Beiras C, Marks M, Chen CY et al (2016) Epidemiology of Haemophilus ducreyi infections. Emerg Infect Dis 22(1):1
14. de Vries HJ (2019) Lymphoganuloma venereum in the Western world, 15 years after its re-emergence: new perspectives and research priorities. Curr Opin Infect Dis 32(1):43–50
15. Cole MJ, Field N, Pitt R, Amato-Gauci AJ, Begovac J, French PD, Keše D, Klavs I, Lepej SZ, Pöcher K, Stary A (2019) Substantial underdiagnosis of lymphogranuloma venereum in men who have sex with men in Europe: preliminary findings from a multicentre surveillance pilot. Sex Transm Infect 96 (2): sextrans-2019
16. O'Farrell N (2002) Donovanosis. Sex Transm Infect 78:452–457
17. Patel H, Wagner M, Singhal P et al (2013) Systematic review of the incidence and prevalence of genital warts. BMC Infect Dis 13:39

18. Chandler DJ, Fuller LC (2019) A review of scabies: an infestation more than skin deep. Dermatology 235:79–90
19. Global Burden of Disease Study 2013 Collaborators (2015) Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 386 (9995):743–800. https://doi.org/10.1016/S0140-6736(15)60692-4
20. Baumann L, Cina M, Egli-Gany D et al (2018) Prevalence of mycoplasma genitalium in different population groups: systematic review and meta-analysis. Sex Transm Infect 94(4):255–262
21. Brunham RC, Gottlieb SM, Paavonen J (2015) Pelvic inflammatory disease. N Engl J Med 372:2039–2048
22. Ciotti M, Paba P, Bonifacio D et al (2006) Single or multiple HPV types in cervical cancer and associated metastases. Oncol Rep 15:143–148
23. Ringehan M, McKeating JA, Protzer U (2017) Viral hepatitis and liver cancer. Phil Trans R Soc B 372:20160274
24. Juzlova K, Rob F, Zakoucka H et al (2018) The first case of lymphogranuloma venereum in a woman in East-Central Europe and its multiple co-infections. Folia Microbiol 63:93–95
25. Mutagoma M, Nyirazinyoye L, Sebuhoro D et al (2017) Syphilis and HIV prevalence and associated factors to their co-infection, hepatitis B and hepatitis C viruses prevalence among female sex workers in Rwanda. BMC Infect Dis 17:525
26. Peebles K, Velloza J, Balkus JE et al (2019) High global burden and costs of bacterial vaginosis: a systematic review and meta-analysis. Sex Transm Dis 46(5):304–311
27. Dreyer G (2018) Clinical implications of the interaction between HPV and HIV infections. Best Pract Res Clin Obstet Gynaecol 47:95–106
28. Mungati M, Machiha A, Mugurungi O et al (2018) The etiology of genital ulcer disease and coinfections with chlamydia trachomatis and Neisseria gonorrhoeae in Zimbabwe: results from the Zimbabwe STI etiology study. Sex Transm Dis 45 (1):61–68
29. Orlando G, Bianchi S, Fasolo MM et al (2017) Cervical human papillomavirus genotypes in HIV-infected women: a cross-sectional analysis of the VALHIDATE study. J Prev Med Hyg 58:E259–E265
30. Ortiz AP, Tamayo V, Scorsone A et al (2017) Prevalence and correlates of cervical HPV infection in a clinic-based sample of HIV-positive Hispanic women. Papillomavirus Res 4:39–44
31. Reda S, Gonçalves FA, Mazepa MM et al (2018) Women infected with HIV and the impact of associated sexually transmitted infections. Int J Gynecol Obstet 142:143–147
32. Badial RM, Dias MC, Stuqui B et al (2018) Detection and genotyping of human papillomavirus (HPV) in HIV-infected women and its relationship with HPV/HIV co-infection. Medicine 97:14
33. Wagenlehner FME, Brockmeyer NH, Discher T et al (2016) The presentation, diagnosis and treatment of sexually transmitted infections. Dtsch Arztebl Int 113:11–22. https://doi.org/10.3238/arztebl.2016.0011
34. Theuretzbacher U (2016) Global antimicrobial resistance in gram-negative pathogens and clinical need. Curr Opin Microbiol 39:106–112
35. Kapoor G, Saigal S, Elongavan A (2017) Action and resistance mechanisms of antibiotics: a guide for clinicians. J Anaesthesiol Clin Pharmacol 33:300–305
36. Munita JM, Arias CA (2016) Mechanisms of antibiotic resistance. Microbiol Spectr 4(2). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015
37. Bodie M, Gale-Rowe M, Alexandre S et al (2019) Addressing the rising rates of gonorrhea and drug resistant gonorrhea: there's no time like the present. Can Commun Dis Rep 45(2/3):54–62
38. Hamers LR, Rinke de Wit TF, Holmes BC (2018) HIV drug resistance in low-income and middle-income countries. Lancet HIV 5:e588–e596
39. Jiang YC, Feng H, Lin YC et al (2016) New strategies against drug resistance to herpes simplex virus. Int J Oral Sci 8:1–6
40. Mazzariol A, Bazaj A, Cornaglia G (2017) Multi-drug-resistance gram-negative bacteria causing urinary tract infections: a review. J Chemother 29:S1
41. Sheikh AF, Veisi H, Shahin M et al (2019) Frequency of quinolone resistance gene among extended-spectrum ß-lactamase (ESBL)-producing *Escherichia coli* strains isolated from urinary tract infections. Trop Med Health 47:19
42. Zhang JY, Liu JH, Liu FD et al (2014) Vulvovaginal candidiasis: species distribution, fluconazole resistance and drug efflux pump gene overexpression. Mycoses 57:584–591
43. Strzelecka PM, Ranzoli AM, Cvejic A (2018) Dissecting human disease with single-cell omics: application in model systems and in the clinic. Dis Models Mech 11:dmm036525
44. Wang Y, Navin NE (2015) Advances and applications of single-cell sequencing technologies. Mol Cell 58:598–609
45. AlJanahi AA, Danielsen M, Dunbar CE (2018) An introduction to the analysis of single-cell RNA-sequencing data. Mol Ther Methods Clin Dev 10:189–196
46. Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50:96
47. Tang X, Huang Y, Lei J et al (2019) The single-cell sequencing: new developments and medical applications. Cell Biosci 9:53
48. Brehm-Stecher BF, Johnson EA (2004) Single-cell microbiology: tools, technologies, and applications. Microbiol Mol Biol Rev 68(3):538–559

49. Chen Z, Chen L, Zhang W (2017) Tools for genomic and transcriptomic analysis of microbes at single-cell level. Front Microbiol 8:1831
50. Liang J, Cai W, Sun Z (2014) Single-cell sequencing technologies: current and future. J Genet Genomics 14:513–528
51. Chen G, Ning B, Shi T (2019) Single-cell RNA-seq technologies and related computational data analysis. Front Genet 10:317
52. Valihrach L, Androvic P, Kubista M (2018) Platforms for single-cell collection and analysis. Int J Mol Sci 19(3):807
53. Ilicic T, Kim JK, Kolodziejczyk AA et al (2016) Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17:29
54. Yalcin D, Hakgunder ZM, Otu HH (2016) Bioinformatics approaches to single-cell analysis in developmental biology. Mol Hum Reprod 22:182–192
55. Ji F, Sadreyev RI (2019) Single-cell RNA-seq: introduction to bioinformatics analysis. Curr Protoc Mol Biol 127:e92. https://doi.org/10.1002/cpmb.92
56. Wang Z, Hu J, Evan W (2019) scruff: an R/Bioconductor package for preprocessing single-cell RNA-sequencing data. BMC Bioinformatics 20:222
57. Wang T, Li B, Nelson CE et al (2019) Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics 20:40
58. Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15:e8746
59. Vieth B, Parekh S, Ziegenhain C et al (2019) A systematic evaluation of single cell RNA-seq analysis pipelines. Nat commun 10(1):1–1
60. Zhu X, Wolfgruber TK, Tasato A et al (2017) Granatum: a graphical single-cell RNA-seq analysis pipeline for genomics scientists. Genome Med 9 (1):108
61. Upadhyay AA, Kauffman RC, Wolabaugh AN et al (2018) BALDR: a computational pipeline for paired heavy and light chain immunoglobulin reconstruction in single-cell RNA-seq data. Genome Med 10:20
62. Tian L, Su S, Dong X et al (2018) scPipe: A flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. PLoS Comput Biol 14 (8):e1006361
63. Guo M, Wang H, Potter SS et al (2015) SINCERA: a pipeline for single-cell RNA-seq profiling analysis. PLoS Comput Biol 11(11):e1004575
64. Gardeux V, David FPA, Shajkofci A et al (2017) ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. Bioinformatics 33(19):3123–3125
65. Ren X, Zheng L, Zhang Z (2019) SSCC: a novel computational framework for rapid and accurate clustering large-scale single-cell RNA-seq data. Genomics Proteomics Bioinformatics 17:201–210
66. Patel MV (2018) iS-CellR: a user-friendly tool for analyzing and visualizing single-cell RNA sequencing data. Bioinformatics 34(24):4305–4306
67. Derr A, Yang C, Zilionis R et al (2016) End sequence analysis toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. Genome Res 26(10):1397–1410
68. Moussa M, Măndoiu II (2018) SC1: a web-based single cell RNA-seq analysis pipeline. In: 2018 IEEE 8th international conference on computational advances in bio and medical sciences (ICCABS) 2018 Oct 18, IEEE, p 1
69. Iacono G, Mereu E et al (2018) bigSCale: an analytical framework for big-scale single-cell data. Genome Res 28(6):878–890
70. Yucha RW, Hobbs KS, Hanhauser E et al (2017) High-throughput characterization of HIV-1 reservoir reactivation using a single-cell-in-droplet PCR assay. EBioMedicine 20:217–229
71. Vandergeeten C, Fromentin R, DaFonseca S et al (2013) Interleukin-7 promotes HIV persistence during antiretroviral therapy. Blood 121(21):4321–4329
72. Kwon KJ, Siliciano RF (2017) HIV persistence: clonal expansion of cells in the latent reservoir. J Clin Invest 127(7):2536–2538
73. Wiegand A, Spindler J, Hong FF et al (2017) Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. Proc Natl Acad Sci U S A 114(18):E3659–E3668
74. Holmes M, Zhang F, Bieniasz PD (2015) Single-cell and single-cycle analysis of HIV-1 replication. PLoS Pathog 11(6):e1004961
75. Cohn LB, da Silva IT, Valieris R, Huang AS, Lorenzi JC, Cohen YZ, Pai JA, Butler AL, Caskey M, Jankovic M, Nussenzweig MC (2018) Clonal CD4+ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat Med 24(5):604
76. Golumbeanu M, Cristinelli S, Rato S et al (2018) Single-cell RNA-seq reveals transcriptional heterogeneity in latent and reactivated HIV-infected cells. Cell Rep 23:942–950
77. Rato S, Rausell A, Muñoz M et al (2017) Single-cell analysis identifies cellular markers of the HIV permissive cell. PLoS Pathog 13(10):e1006678
78. Bradley T, Ferrari G, Haynes BF et al (2018) Single-cell analysis of quiescent HIV infection reveals host transcription profiles that regulate proviral latency. Cell Rep 25:107–117
79. Farhadian SF, Mehta SS, Zografou C et al (2018) Single-cell RNA sequencing reveals microglia-like cells in cerebrospinal fluid during virologically suppressed HIV. JCI Insight 3(18):e121718
80. de Armas LR, Pallikkuth S, Pan L et al (2019) Single cell profiling reveals PTEN overexpression in influenza-specific B cells in aging HIV-infected individuals on antiretroviral therapy. Sci Rep 9:2482
81. Doorbar J (2018) Host control of human papillomavirus infection and disease. Best Pract Res Clin Obstet Gynaecol 47:27–41
82. Harari A, Chen Z, Burk RD (2014) Human papillomavirus genomics: past, present and future. In: Human papillomavirus, vol 45. Karger Publishers, Basel, pp 1–18

83. Harden ME, Munger K (2017) Human papillomavirus molecular biology. Mutat Res 772:3–12
84. Forman D, Martel CD, Lacey CJ et al (2012) Global burden of human papillomavirus and related diseases. Vaccine 305:F12–F23
85. Wu L, Zhang X, Zhao Z et al (2015) Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells. GigaScience 4:51
86. Shen Z, Liu X, Morihara J et al (2015) Detection of human papillomavirus infections at the single-cell level. Intervirology 58:324–331
87. Yang D, Zhang W, Liu Y et al (2018) Single-cell whole-genome sequencing identifies human papillomavirus integration in cervical tumour cells prior to and following radiotherapy. Oncol Lett 15:9633–9640
88. Lafferty WE, Downey L, Celum C et al (2000) Herpes simplex virus type 1 as a cause of genital herpes: impact on surveillance and prevention. J Infect Dis 181:1454–1457
89. Parra-Sánchez M (2019) Genital ulcers caused by herpes simplex virus. Enferm Infecc Microbiol Clin 37(4):260–264
90. Wyler E, Franke V, Menegatti J et al (2019). Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. Nat Commun 10(1):1–4
91. Drayman N, Patel P, Vistain L et al (2019) HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. elife 8:e46339
92. Brook MG (2002) Sexually acquired hepatitis. Sex Transm Infect 78:235–240
93. Chan PCD, Sun HY, Wong THH et al (2016) Sexually acquired hepatitis C virus infection: a review. Int J Infect Dis 49:47–58
94. Leitch ECM, McLauchlan J (2013) Determining the cellular diversity of hepatitis C virus quasispecies by single-cell viral sequencing. J Virol 87 (23):12648–12655
95. Eltahla AA, Rizzetto S, Pirozyan MR et al (2016) Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol Cell Biol 94:604–611
96. Wang X, He Q, Shen H et al (2019) Genetic and phenotypic difference in CD8+ T cell exhaustion between chronic hepatitis B infection and hepatocellular carcinoma. J Med Genet 56:18–21
97. Duan M, Hao J, Cui S et al (2018) Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. Cell Res 28:359–373
98. Chen XP, Long X, Jia W et al (2019) Viral integration drives multifocal HCC during the occult HBV infection. J Exp Clin Cancer Res 38:261
99. Paz-Zarza VM, Mordani SM, Maldonado AM et al (2019) Pseudomonas aeruginosa: pathogenicity and antimicrobial resistance in urinary tract infection. Rev Chil Infectol 36(2):180–189
100. Behzadi P, Behzadi E, Yazdanbod H et al (2010) A survey on urinary tract infections associated with the three most common uropathogenic bacteria. Maedica 5(2):111
101. Duraiswamy S, Chee LYJ, Chen S et al (2018) Purification of intracellular bacterial communities during experimental urinary tract infection reveals an abundant and viable bacterial reservoir. Infect Immun 86 (4):e00740–e00717
102. Yang E, Chee JL, Duraiswamy S et al (2019) Isolation of single intracellular bacterial communities generated from a murine model of urinary tract infection for downstream single-cell analysis. J Vis Exp 146:e58829
103. Baltekin Ö, Boucharin A, Tano E et al (2017) Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. PNAS 114 (34):9170–9175
104. Goormaghtigh F, Melderen LV (2019) Single-cell imaging and characterization of Escherichia coli persistence cells to ofloxacin in experimental cultures. Sci Adv 5:eaav9462
105. Muñoz JF, Delorey T, Ford CB et al (2019) Coordinated host-pathogen transcriptional dynamics revealed using sorted subpopulations and single macrophages infected with Candida albicans. Nat Commun 10:1607
106. Lagier J-C, Edouard S, Pagnier I et al (2015) Current and past strategies for bacterial culture in clinical microbiology. Clin Microbiol Rev 28:208–236
107. Zhang Y, Gao J, Huang Y et al (2018) Recent developments in single-cell RNA-seq of microorganisms. Biophys J 115(2):173–180
108. Yuan GC (ed) (2019) Computational methods for single-cell data analysis. Humana Press, Totowa



Reema Singh is a postdoctoral fellow with the college of medicine at the University of Saskatchewan (Canada). She earned a master's degree in "Bioinformatics" from Chaudhary Charan Singh Haryana Agricultural University (Hisar, India) in 2006, and her Ph.D. in "Computational

Biology and Bioinformatics" from Jawaharlal Nehru University (JNU, New Delhi, India) in 2015. Dr. Singh worked at GJU Science & Technology University (India) as a Guest Faculty in the Department of Bio and Nanotechnology in 2006 for some time. She also served as a Scientist-1 in the Biomedical Informatics Centre at the Indian Council of Medical Research (New Delhi, India) from 2006 to 2012. During her stay in ICMR, she developed an antimicrobial resistance (AMR) gene database of ß-lactamases (Dlact) from 814 bacterial genomes and identified ß-lactamases group-specific signatures. She moved to Scotland (United Kingdom) in 2013 for her first postdoc in "Bioinformatic analysis of next-generation sequencing data from *Dictyostelium* social amoebae" at the University of Dundee. In 2017, she joined the University of Saskatchewan for her second postdoc to work on a research project entitled "A Disruptive Whole Genome Sequencing Platform for the Simultaneous Identification and Characterization of Multiple Sexually Transmitted Pathogens." As a part of the project, she developed a computational Whole Genome Sequencing (WGS) pipeline, named Gen2Epi, to link full genomes to antimicrobial susceptibility and molecular epidemiological data in *Neisseria gonorrhoeae*. Using this pipeline she has analyzed thousands of *Neisseria gonorrhoeae* WGS datasets. Her main research interests focus on analyzing large-scale next-generation sequencing data to predict and track antimicrobial resistance, identify novel mutations associated with AMR, and develop an integrated framework to link the novel AMR mutations with the phenotypic changes using transcriptomics, proteomics, and metabolomics data in genital infection causing pathogens. Dr. Singh has published 12 scientific papers in international well-recognized journals (Total Impact factor: 46). For more information visit https://www.researchgate.net/profile/Reema_Singh3.

Emerging Strategies for Therapeutic Antibody Discovery from Human B Cells

18

Saravanan Rajan and William F. Dall'Acqua

Abstract

Monoclonal antibodies from human sources are being increasingly recognized as valuable options in many therapeutic areas. These antibodies can show exquisite specificity and high potency while maintaining a desirable safety profile, having been matured and tolerized within human patients. However, the discovery of these antibodies presents important challenges, since the B cells encoding therapeutic antibodies can be rare in a typical blood draw and are short-lived ex vivo. Furthermore, the unique pairing of V_H and V_L domains in each B cell contributes to specificity and function; therefore, maintaining antibody chain pairing presents a throughput limitation. This work will review the various approaches aimed at addressing these challenges with an eye to next-generation methods for high-throughput discovery from the human B-cell repertoire.

Keywords

Antibody therapeutics · Next-generation sequencing · Microfluidics · Phage display · Hybridoma · High-throughput screening

18.1 Introduction

In the past few years, the use of monoclonal antibodies has seen an explosion as therapeutics, diagnostics, and tools in biomedical research. This is because antibodies carry exquisite specificity to their respective target, persist in the body for many weeks (particularly if endowed with half-life extension technology) and can elicit responses through multiple mechanisms of action, particularly as relates to interactions with their Fc domains. With several improvements in R&D for antibody discovery and development, the number of antibody therapeutics has dramatically increased, with over 570 molecules in clinical development and 12 new molecules approved in 2018 alone [1]. The therapeutic use of monoclonal antibodies spans the breath of therapeutic areas, including infectious disease, cancer, and autoimmune disorders and increasingly ingenious delivery methods, from inhaled to gene therapy, are constantly improving the convenience of administering these therapeutics such that we expect their use to continue increasing with time.

Although monoclonal antibody research began with mouse hybridoma technology [2] and many approved antibodies are mouse-derived, the immunogenicity issues that ensued have led to an interest in antibody discovery from other sources, particularly human. This review will therefore focus on technologies that derive therapeutic antibodies from human sources, though it is acknowledged that many of the next-generation

S. Rajan (✉) · W. F. Dall'Acqua
Antibody Discovery and Protein Engineering, AstraZeneca R&D, Gaithersburg, MD, USA
e-mail: sarav.rajan@astrazeneca.com

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_18

B cell approaches could be applied to mouse or any species for which we have antibody sequence information.

18.2 Antibody Discovery from Synthetic Libraries

Antibody specificity and activity are the result of sequence evolution for a starting set of germline immunoglobulin sequences. One approach to artificially simulate this evolution is through random mutation of a human germline sequence using error-prone PCR [3, 4] or specific mutation of the complementarity determining regions (CDR) using randomized primers [5, 6] and selecting variants with desirable characteristics through one of the display-based methods such as phage display [7]. These approaches can screen through libraries of massive diversity for binders, though the size of most naïve libraries is often dwarfed by the theoretical diversity of the introduced mutations, making full coverage of the sequence space impossible. The resulting antibody candidates therefore tend to be partially evolved sequences with mid-level affinities and modest therapeutic efficacy. They can however serve as a template for secondary evolution through more targeted affinity maturation processes before final therapeutic leads can be chosen.

18.3 Antibody Discovery from Natural Repertoires

Another approach is to use the extraordinary power of natural systems to evolve antibodies, through immunizations of mice or other species, including humans. Here, sequence evolution and selection take place in germinal centers where the antibody expressing B cells compete for limited antigen binding and growth factors. Once isolated, these antibodies can display very high-affinity (often picomolar K_D or lower) for their cognate antigens and target a variety of epitopes, including functional ones. Moreover, antibodies evolved within humans may be better tolerated as therapeutics, having edited out immunogenic sequence variants during evolution. The vast majority of antibodies on the market and in development have been isolated from natural repertoires, most notably in the infectious disease areas, with many prominent reviews on the discovery of neutralizing antibodies against HIV [8], Influenza [9], Ebola [10], Zika [11], and many more. While not as extensive, studying B cells from cancer patients is a burgeoning field and recent analyses of the B cell repertoires from non-progressing cancer patients have led to the identification of specific tumor-inhibiting antibodies that have therapeutic potential [12–14]. Similarly, antibody-mediated autoimmune diseases are benefiting from the analysis and screening of the patient B cell repertoire, including in myasthenia gravis [15, 16], Celiac disease [17, 18], multiple sclerosis [19] and rheumatoid arthritis [20]. While the identification of pathogenic antibodies in these diseases does not directly represent a therapeutic option, the targets they bind can point to potential avenues for valuable therapies that may emerge in the future.

18.4 Challenges and Opportunities

Despite many successes, recovering antigen-specific antibody sequences from humans is challenging for reasons listed below and in response to these there has been a steady evolution of technologies to further improve and simplify the process (Fig. 18.1). Each of these approaches have successfully led to the identification of valuable antibodies, some seminal to studying of the disease in question.

18.5 Antigen-Specific B Cells Can Be Very Rare

B cells producing antibodies against any particular antigen tend to be rare in the blood of a healthy or convalescent individual, amidst the vast

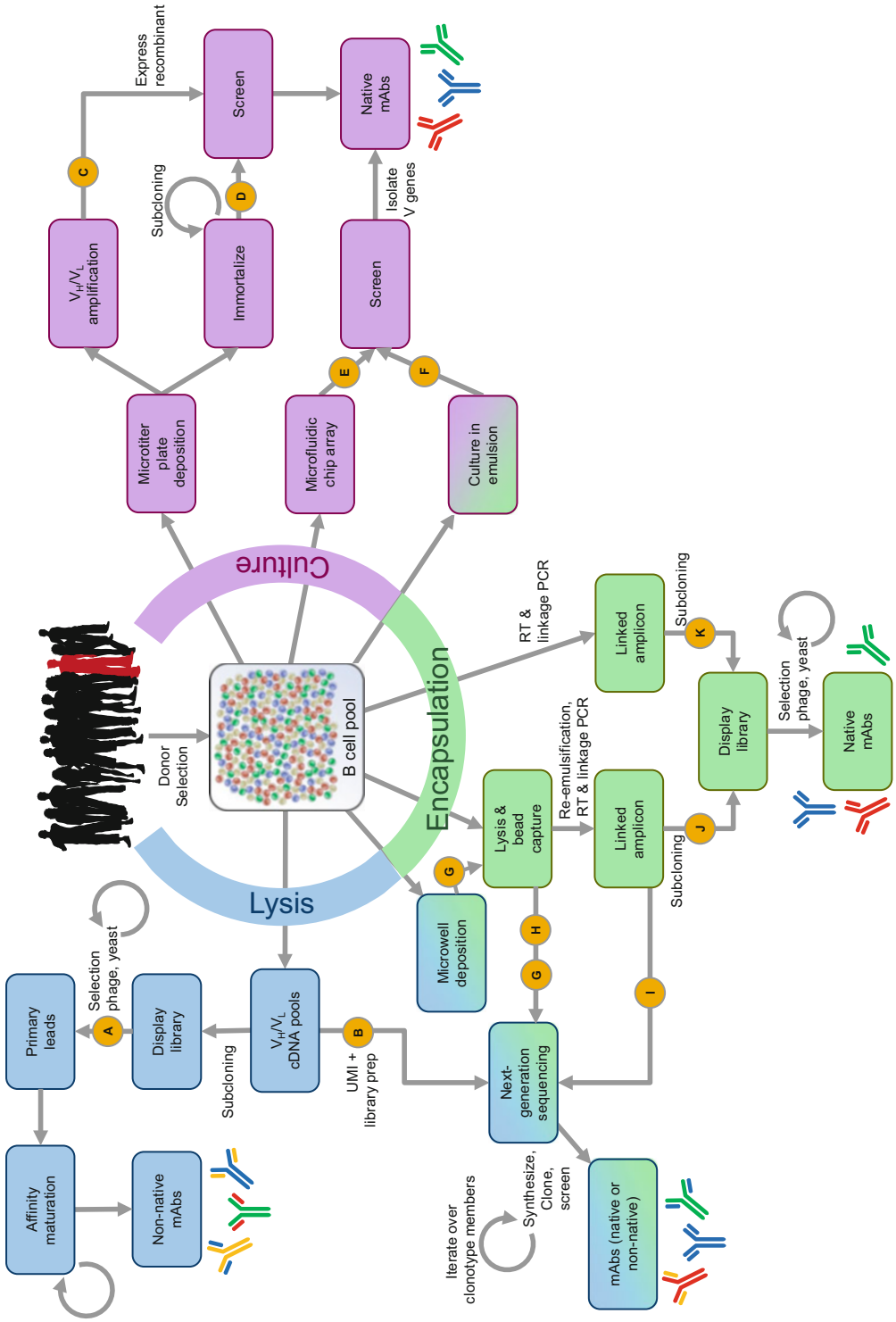


Fig. 18.1 Overview of technologies used to discover antibodies from human B cells. Following selection of reactive donors, polyclonal B cells are isolated from blood or lymphoid tissues and the antibodies which they encode can be screened for antigen specificity through many approaches. Cells can be lysed in bulk and enriched through

number of cells encoding irrelevant antibodies. Finding these cells using the standard practice of depositing single B cells in each well of a microtiter plate therefore becomes very challenging. In some cases, this problem can be circumvented by immunizing the host and collecting blood when antigen-specific B cells reach peak abundance and indeed most discovery campaigns using single B cell cloning have used this approach. Additionally, if therapeutically relevant domains can be purified, these can be used not only for immunization, but also selection of antigen-specific memory B cells through FACS, for instance in the identification of broadly neutralizing anti-HIV antibodies of high therapeutic potential [21]. However, in the absence of a specific domain of interest, immunodominant antibodies elicited through vaccination may not be against epitopes with therapeutic potential, such as sites that mediate neutralization or conserved sites useful for cross-reactivity. Moreover, the availability of a suitable immunogen, adjuvant and relevant host can limit the diseases for which this approach can be used.

18.6 B Cells Are Short-Lived Ex Vivo

Campaigns using primary B cells are significantly time-bound, as the ex-vivo viability of primary B cells is limited to 1–2 weeks, with antibody expression waning prior to that, particularly if grown in isolated cultures. Moreover, the proportion of B cells actively secreting IgG within the blood is very low, so methods need to be in place to differentiate non-secreting cells (i.e., memory B cells) in culture. One method would be to perform RT-PCR on lysed single B cells, followed by amplification of V_H/V_L sequences, and reconstitution of the antibody in recombinant format for screening. The approach works well to find antibodies from immunized donors where the proportion of antigen-specific B cells is high but can be a laborious process if most B cells express irrelevant antibodies [22–24]. Several approaches have focused on immortalizing B cells using viral infection [25–27] or hybridoma generation [28], coupled with cytokine stimulation for cells to secrete antibody for screening. However, each of these steps (immortalization, stimulation, fusion, single-cell cloning) carry inherent inefficiencies and biases that when put together may limit the number of single B cells that can be screened. Successful campaigns would also need large B cell populations to be kept in culture for extended periods with considerable manipulations.

18.7 Antibody Chain Pairing Is Often Important for Function

Antibodies are heterodimeric proteins encoded by uniquely mutated heavy and light chain transcripts whose pairing is often necessary for specificity and activity. Ideally this information

Fig. 18.1 (continued) combinatorial libraries using display technologies (A) or barcoded using unique molecular identifiers (UMI) and sequenced using next-generation sequencing (NGS—B). Various analysis methods can be used to identify dominant clonotypes which need to be paired and synthesized for screening. Individual members of each clonotype can then be subsequently screened for improved function. To preserve the native V_H/V_L pairing, B cells are deposited in microtiter plates through FACS cloning and isolation of individual V_H/V_L sequences (C) by reverse-transcription (RT) and polymerase chain reaction (PCR), followed by recombinant expression and screening. B cells can also be immortalized, stimulated and cultured to allow conditioned media to be screened (D). This can be miniaturized using commercial platforms, either using nanoliter-sized chambers (e.g., AbCellera, Berkeley Lights—E) or picoliter-sized water-in-oil emulsions (e.g., HiFiBiO—F) whereby single B cells can be screened without immortalization, recovered and sequenced. If B cells are colocated with poly-dT beads, they can be lysed and cognate V_H/V_L mRNA species paired for NGS. This can be done using microwells on chips (G) or using emulsions (10x Genomics—H). The beads can also be re-emulsified to generate a linked product suitable for NGS (I). Finally, this linked amplicon can be generated in-frame using beads (J) or directly in droplets (K) to create a natively paired library which can be enriched for antigen specificity using display approaches, thereby combining the advantages of most of the above-mentioned approaches. Donor image designed by Kjpargeter/Freepik (www.freepik.com)

can be captured by sequestering individual B cells in microtiter plates for extraction and cloning of their V genes [23, 24, 29], with the limitation that again only a fraction of the full B cell repertoire of a typical blood draw (one to ten million cells) can be covered. Conversely, it has been a common approach to collect total RNA from a large pool of lysed B cells, separately amplify heavy and light chains, then pair them randomly to form a library of exponentially larger complexity [30–32]. In theory, the diversity of such immune libraries is lower than that of the synthetic libraries mentioned above and could be harnessed using phage display to identify the originally paired and functional sequence and this has been shown with immunized mice with restricted diversity [33]. However, in more diverse libraries, for instance from healthy donors, recovering the original pair is a challenge. Biases in expression and V_H/V_L pairing preferences [34] can lead to the selection to nonnatural solutions and require secondary optimization screens, for instance using light chain shuffling.

Another approach is to sequence the B cell repertoire and synthesize candidate antibodies for screening, for instance after hierarchical clustering of sequences to identify phylogenetic lineages. The application of next-generation sequencing (NGS) has been well described for the characterization of the antibody repertoire, particularly regarding separately prepared heavy and light chain libraries obtained from mRNA isolated from thousands to millions of B cells [35–37]. A recent NGS analysis of the antibody repertoires from ten individuals revealed that their repertoires were largely unique and that the overall diversity of antibody sequences in the human population is extremely large, on the order of 10^{12} unique paired V_H/V_L sequences [37]. While the amplification of B cell mRNA using 5′ RACE provides an unbiased representation of the expressed repertoire for sequencing, current NGS length limitations make assembling such a fragment from paired-end sequencing a challenge. As a result, libraries are often made using multiplex V-gene specific primers to remove the 5′ untranslated region and leader sequences and reduce amplicon size [38] which can introduce bias. Additionally, given that antibodies undergo somatic hypermutation, a significant challenge in the field was to determine if a given mutation was due to natural antibody diversification or a result of PCR and/or sequencing-related artifacts. This issue has been elegantly solved through the addition of unique molecular identifier (UMI) barcodes, where the initial template cDNA can be ligated to a unique tag that is also sequenced to enable error correction at the analysis stage [39, 40]. Using these methods, a population of B cells can be profiled to identify phylogenetic lineages [41] that indicate the maturation of specific clonotypes as evidence of antigen specificity. However, given that maturation at the heavy and light chains occur independently, it is not possible to accurately predict chain pairing based on NGS data and heuristics need to be used to down-select panels of heavy and light chain sequences to synthesize and combinatorially pair for functional testing. Again, in cases where subjects are immunized and B cells harvested at optimal times, this approach can be quite effective, as the most abundant heavy and light chain clonotypes may represent the original pairs. However, for cases involving the identification of antigen-specific antibodies from healthy donors or patients with chronic diseases such as cancer, the selection of antibody function from sequence information alone is likely to be incomplete.

18.8 Next-Generation Microfluidic Technologies

Though successful in many instances, these two broad approaches (display versus B cell cloning/sequencing) suffer from conflicting issues. On the one hand, display-based systems can screen through vast synthetic or combinatorial libraries to identify antigen-specific antibodies of mid to low quality. Conversely, B cell discovery platforms start from B cell pools encoding potentially high-quality antibodies but lack the screening power to identify antigen-specific antibodies,

especially if the B cells encoding them are rare. Over the past 5 years a range of new technologies have emerged promising to solve this tradeoff by miniaturizing the vessel into which B cells are sequestered. Several growing companies (AbCellera, Berkeley Lights, and HiFiBiO) have been successful in directly screening antibody secreting cells within these vessels for binding or even functional activity then exporting antigen-specific B cells into defined locations for V gene capture. Alternatively, several approaches described below have detailed using droplet microfluidics to capture the natively paired repertoire from B cells into a format suitable for next-generation sequencing technology. Finally, the repertoire can be captured via microfluidics in an expressible format to display and/or screen as recombinant protein, effectively combining the benefits of natural antibody evolution with the screening power of display-based approaches.

18.9 Paired Ig Sequencing

In 2013, DeKosky and colleagues devised a method to have B cells deposited within microwells on a microfabricated chip along with magnetic beads conjugated to poly-dT oligonucleotides [42]. The chip could be sealed with lysis reagents such that the cognate heavy and light mRNA strands would be recovered and linked in a format suitable for next-generation sequencing. The repertoire from 68,000 B cells could be captured in a single run, an improvement in throughput of one order of magnitude over traditional 96-well formats. In a follow-up paper, the group expanded the method to have the B cells and magnetic beads encapsulated into water-in-oil droplets, further raising the throughput to one million B cells per run [43]. This falls within the range of B cells obtained from a typical blood draw and enabled the first comprehensive evaluations of the paired antibody repertoire for therapeutic antibodies. As with the single chain NGS studies however, it is a challenge to determine antigen reactivity from antibody sequence alone. An elegant addition to this method therefore has been to overlay paired sequencing data from circulating B cells with proteomic sequence analysis of serum antibodies immunoprecipitated with antigen [44, 45]. Using immunized donors, the authors were able to identify potent neutralizing antibodies targeting influenza and HIV. The advent of 10x Genomics now provides a commercial option for obtaining paired immunoglobulin sequences from primary B cells, albeit from a smaller number of cells (approximately 10,000 cells) and this system has recently been used to sequence B cells from immunized mice to identify antigen-specific antibodies [46].

18.10 Native Library Screening

A natural evolution of these technologies has led to the combination of miniaturizing B cell capture into microfluidic emulsions with paired immunoglobulin capture into a format that can be expressed. Recently, three independent groups have reported in short succession microfluidic methods to capture the repertoire from millions of B cells and rapidly screen them for antigen-specific antibodies.

Adler and colleagues at GigaGen devised an approach to co-encapsulate one to two million B cells in a co-flow setup with poly-dT magnetic beads suspended in lysis/binding buffer [47]. Following bead capture of the mRNA, the emulsions are broken and recovered beads re-emulsified with RT-PCR buffer and a cocktail of primers to generated linked heavy-light amplicons in scFv format. The authors then expressed this library of natively paired scFv amplicons on the surface of yeast and used multiple rounds of fluorescence-activated cell sorting (FACS) to enrich for yeast cells displaying antibodies specific to influenza A and pneumococcal polysaccharide antigens. A subset of recovered scFv-s confirmed to be antigen-specific and functional when expressed in IgG format, and based on NGS analysis, they were estimated to be present at 0.001% of the starting library diversity. The approach was also validated with immunized mice (having a higher

proportion of antigen-specific B cells) for the isolation of antibodies blocking checkpoint inhibitors [48].

Wang and coworkers also reported a similar two-step emulsification strategy to generate a natively paired Fab library, a more aggregation resistant antibody fragment with biophysical properties closer aligned with IgG, that was also displayed the library on the surface of yeast [49]. They used this method to construct libraries from immunized or convalescent patients and panned them over successive rounds to isolate functional antibodies specific to Ebola, HIV, and influenza antigens.

Finally, our group has developed a method of capturing the native repertoire from millions of B cells into natively paired scFv fragments displayed on the surface of phage [50]. Here, B cells are not co-encapsulated with magnetic beads but rather with a highly optimized reaction mix that performs sequential reactions for B cell lysis, amplification of V_H and V_L segments and their pairing by overlap-extension PCR, all within the same droplets. This streamlines the process and obviates the need to handle beads, where captured mRNA species can exchange. The approach was used to rapidly identify very rare and cross-reactive antibodies targeting influenza hemagglutinin.

18.11 Future Directions

Since all three native screening methods use emulsions for amplification of template, this may have an added benefit of normalizing for mRNA expression levels within individual B cells, as the limiting reagents within droplets should saturate with enough cycles of PCR. It would be interesting to see if this bears out in future studies using NGS analysis. On the other hand, since these methods require priming at specific regions to be in-frame (i.e., the start of framework 1 and the end of framework 4), the multiplex primer sets may not be ideally suited to perform as well as other amplification methods (i.e., 5′ RACE) and would benefit from continued development. While not extensively mentioned in this text, these new methods have the potential for screening natively paired repertoires from other species simply by changing primer sets. This has been demonstrated with immunized mice [48], though other species such as rat, rabbit, and even nonhuman primates could provide B cells from which valuable antibodies can be derived.

The success of antibody therapeutics has led to increasing numbers of molecules in clinical trials and as approved medicines and this trend is expected to continue. Fueling this growth is a continued evolution of methods for mining new therapeutic antibodies, both through synthetic and natural repertoires. We can expect that the technologies of the future will continue to harness the natural antibody repertoire with increasing throughput, breadth, speed, and fidelity to reliably generate therapeutic candidates against a continually expanding list of targets.

References

1. Kaplon H, Reichert JM (2018) Antibodies to watch in 2019. mAbs 11(2):219–238
2. Köhler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256(5517):495–497
3. Liu J-L, Hu Z-Q, Xing S, Xue S, Li H-P, Zhang J-B et al (2012) Attainment of 15-fold higher affinity of a Fusarium-specific single-chain antibody by directed molecular evolution coupled to phage display. Mol Biotechnol 52(2):111–122
4. Unkauf T, Hust M, Frenzel A (1701) Antibody affinity and stability maturation by error-prone PCR. Methods Mol Biol 2018:393–407
5. Fellouse FA, Li B, Compaan DM, Peden AA, Hymowitz SG, Sidhu SS (2005) Molecular recognition by a binary code. J Mol Biol 348(5):1153–1162
6. Fellouse FA, Barthelemy PA, Kelley RF, Sidhu SS (2006) Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. J Mol Biol 357(1):100–114
7. Rajan S, Sidhu SS (2012) Simplified synthetic antibody libraries. Methods Enzymol 502:3–23
8. Sok D, Burton DR (2018) Recent progress in broadly neutralizing antibodies to HIV. Nat Immunol 19 (11):1179–1188
9. Krammer F (2019) The human antibody response to influenza A virus infection and vaccination. Nat Rev Immunol 19(6):383–397
10. King LB, Milligan JC, West BR, Schendel SL, Ollmann Saphire E (2019) Achieving cross-reactivity

with pan-ebolavirus antibodies. Curr Opin Virol 34:140–148
11. Campos JLS, Mongkolsapaya J, Screaton GR (2018) The immune response against flaviviruses. Nat Immunol 19(11):1189–1198
12. Bushey RT, Moody MA, Nicely NL, Haynes BF, Alam SM, Keir ST et al (2016) A therapeutic antibody for cancer, derived from single human B cells. Cell Rep 15(7):1505–1513
13. DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B et al (2018) Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. Clin Immunol 187:37–45
14. McDaniel JR, Pero SC, Voss WN, Shukla GS, Sun Y, Schaetzle S et al (2018) High-throughput discovery of antitumor antibodies mined from sentinel lymph node B cells and circulating IgG of breast cancer patients. Cancer Immunol Immunother 67(5):729
15. Makino T, Nakamura R, Terakawa M, Muneoka S, Nagahira K, Nagane Y et al (2017) Analysis of peripheral B cells and autoantibodies against the anti-nicotinic acetylcholine receptor derived from patients with myasthenia gravis using single-cell manipulation tools. PLoS One 12(10):e0185976. [cited 2019 Sep 19]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5645109/
16. Stathopoulos P, Kumar A, Heiden JAV, Pascual-Goñi E, Nowak RJ, O'Connor KC (2018) Mechanisms underlying B cell immune dysregulation and autoantibody production in MuSK MG. Ann N Y Acad Sci 1412(1):154
17. Di Niro R, Mesin L, Zheng N-Y, Stamnaes J, Morrissey M, Lee J-H et al (2012) High abundance of plasma cells secreting transglutaminase 2-specific IgA autoantibodies with limited somatic hypermutation in celiac disease intestinal lesions. Nat Med 18(3):441–445
18. Roy B, Neumann RS, Snir O, Iversen R, Sandve GK, Lundin KEA et al (2017) High-throughput single-cell analysis of B cell receptor usage among autoantigen-specific plasma cells in celiac disease. J Immunol 199 (2):782–791
19. Lomakin YA, Zakharova MY, Stepanov AV, Dronina MA, Smirnov IV, Bobik TV et al (2014) Heavy-light chain interrelations of MS-associated immunoglobulins probed by deep sequencing and rational variation. Mol Immunol 62(2):305–314
20. Ge C, Holmdahl R (2019) The structure, specificity and function of anti-citrullinated protein antibodies. Nat Rev Rheumatol 15(8):503–508
21. Krebs SJ, Kwon YD, Schramm CA, Law WH, Donofrio G, Zhou KH et al (2019) Longitudinal analysis reveals early development of three MPER-directed neutralizing antibody lineages from an HIV-1-infected individual. Immunity 50(3):677–691. e13
22. Meijer P-J, Andersen PS, Haahr Hansen M, Steinaa L, Jensen A, Lantto J et al (2006) Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. J Mol Biol 358 (3):764–772
23. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H (2008) Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. J Immunol Methods 329(1–2):112–124
24. Busse CE, Czogiel I, Braun P, Arndt PF, Wardemann H (2014) Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. Eur J Immunol 44(2):597–603
25. Bernasconi NL, Traggiai E, Lanzavecchia A (2002) Maintenance of serological memory by polyclonal activation of human memory B cells. Science 298 (5601):2199–2202
26. Traggiai E, Becker S, Subbarao K, Kolesnikova L, Uematsu Y, Gismondo MR et al (2004) An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med 10(8):871–875
27. Kwakkenbos MJ, Diehl SA, Yasuda E, Bakker AQ, van Geelen CMM, Lukens MV et al (2010) Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. Nat Med 16(1):123–128
28. Yu X, McGraw PA, House FS, Crowe JE (2008) An optimized electrofusion-based protocol for generating virus-specific human monoclonal antibodies. J Immunol Methods 336(2):142–151
29. von Boehmer L, Liu C, Ackerman S, Gitlin AD, Wang Q, Gazumyan A et al (2016) Sequencing and cloning of antigen-specific antibodies from mouse memory B cells. Nat Protoc 11(10):1908–1923
30. McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: filamentous phage displaying antibody variable domains. Nature 348 (6301):552–554
31. Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD, Winter G (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage. J Mol Biol 222 (3):581–597
32. Vaughan TJ, Williams AJ, Pritchard K, Osbourn JK, Pope AR, Earnshaw JC et al (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. Nat Biotechnol 14(3):309–314
33. Wang B, Lee C-H, Johnson EL, Kluwe CA, Cunningham JC, Tanno H et al (2016) Discovery of high affinity anti-ricin antibodies by B cell receptor sequencing and by yeast display of combinatorial VH: VL libraries from immunized animals. mAbs 8 (6):1035–1044
34. Jayaram N, Bhowmick P, Martin ACR (2012) Germline VH/VL pairing in antibodies. Protein Eng Des Sel 25(10):523–529
35. Menzel U, Greiff V, Khan TA, Haessler U, Hellmann I, Friedensohn S et al (2014)

Comprehensive evaluation and optimization of amplicon library preparation methods for high-throughput antibody sequencing. PLoS One 9(5): e96727
36. Turchaninova MA, Davydov A, Britanova OV, Shugay M, Bikos V, Egorov ES et al (2016) High-quality full-length immunoglobulin profiling with unique molecular barcoding. Nat Protoc 11 (9):1599–1616
37. Briney B, Inderbitzin A, Joyce C, Burton DR (2019) Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature 566 (7744):393–397
38. Bernat NV, Corcoran M, Hardt U, Kaduk M, Phad GE, Martin M et al (2019) High-quality library preparation for NGS-based immunoglobulin germline gene inference and repertoire expression analysis. Front Immunol 10:660. [cited 2019 Sep 19]. https://www-ncbi-nlm-nih-gov.iscape.idm.oclc.org/pmc/articles/PMC6459949/
39. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 108(23):9530–9535
40. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S et al (2011) Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods 9(1):72–74
41. Briney B, Le K, Zhu J, Burton DR (2016) Clonify: unseeded antibody lineage assignment from next-generation sequencing data. Sci Rep 6:23901. [cited 2019 Sep 19]. https://www-ncbi-nlm-nih-gov.iscape.idm.oclc.org/pmc/articles/PMC4840318/
42. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM et al (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotechnol 31 (2):166–169
43. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD et al (2015) In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat Med 21(1):86–91
44. Lee J, Boutz DR, Chromikova V, Joyce MG, Vollmers C, Leung K et al (2016) Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. Nat Med 22(12):1456
45. Williams LD, Ofek G, Schätzle S, McDaniel JR, Lu X, Nicely NI et al (2017) Potent and broad HIV-neutralizing antibodies in memory B cells and plasma. Sci Immunol. 2(7). [cited 2019 Sep 18]. https://www-ncbi-nlm-nih-gov.iscape.idm.oclc.org/pmc/articles/PMC5905719/
46. Goldstein LD, Chen Y-JJ, Wu J, Chaudhuri S, Hsiao Y-C, Schneider K et al (2019) Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. Commun Biol 2:304
47. Adler AS, Mizrahi RA, Spindler MJ, Adams MS, Asensio MA, Edgar RC et al (2017) Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics. mAbs 9(8):1282–1296
48. Adler AS, Mizrahi RA, Spindler MJ, Adams MS, Asensio MA, Edgar RC et al (2017) Rare, high-affinity mouse anti-PD-1 antibodies that function in checkpoint blockade, discovered using microfluidics and molecular genomics. mAbs 9(8):1270–1281
49. Wang B, DeKosky BJ, Timm MR, Lee J, Normandin E, Misasi J et al (2018) Functional interrogation and mining of natively paired human VH:VL antibody repertoires. Nat Biotechnol 36 (2):152–155
50. Rajan S, Kierny MR, Mercer A, Wu J, Tovchigrechko A, Wu H et al (2018) Recombinant human B cell repertoires enable screening for rare, specific, and natively paired antibodies. Commun Biol 1(1):5

Saravanan Rajan is a Senior Scientist in the department of Antibody Discovery and Protein Engineering at AstraZeneca. Dr Rajan obtained his PhD in Human Genetics from McGill University and a postdoctoral fellowship from the University of Toronto, specializing in using high-throughput molecular biology in genomics and the discovery of therapeutic antibodies. Over the past 8 years at AstraZeneca he has contributed to multiple discovery campaigns and earned several distinctions, including major awards for innovation and entrepreneurship and large external grants tied to new technology he developed. He currently manages a team in Gaithersburg, USA that uses a multidisciplinary approach to push the boundaries of antibody discovery.

William F. Dall'Acqua is VP, Research and Development at AstraZeneca, a biopharmaceutical company headquartered in Cambridge, UK. Dr. Dall'Acqua has been studying protein structure, function, and engineering for over 20 years and has coauthored over 70 papers and three book chapters in the field. He has also been named inventor on more than 80 patent applications and on 18 issued patents. He has Bachelor's and Master's Degrees in Biochemistry from University of Paris, and a PhD in Immunology from University of Paris in conjunction with University of Maryland. He also spent 2½ years as a postdoctoral fellow at Genentech in the Molecular Oncology Department before joining AstraZeneca. At AstraZeneca, Dr. Dall'Acqua supervises a group of over 30 scientists in the Department of Antibody Discovery and Protein Engineering, working on discovery of new therapeutic antibody drugs and establishing novel technology platforms.

Values of Single-Cell RNA Sequencing in Development of Cerebral Cortex

19

Enqiang Chang, Xiaoguo Ruan, Ruilou Zhu, Yangyang Wang, and Jiaqiang Zhang

Abstract

The single-cell RNA sequencing (scRNA-seq) is a powerful tool for exploring the complexity, clusters, and specific functions of the brain cells. Using scRNA-seq, the heterogeneity and changes in transcriptomic profiles of a single neuron were defined during dynamic development and differentiation of cells in cerebral cortex regions, and in the pathogenesis of neurological diseases. One of the great challenges is that the brain sample is susceptible to interference and confounding. More advanced methodologies of computational systems biology need to be developed to overcome the inherent interference and technical differences in the detection of single-cell signals. It is expected that scRNA-seq will be extended to metabolic profiles of the single neuron cell on basis of transcriptional profiles and regulatory networks. It is also expected if the transcriptional profiles can be integrated with molecular and functional phenomes in a single neuron and with disease-specific phenomes to understand molecular mechanisms of brain development and disease occurrence. scRNA-seq will provide the new emerging neurological disciple of the artificial intelligent single neuron for deep understanding of brain diseases.

Keywords

Single-cell RNA sequencing · Cerebral cortex · Neurons · Brain · Anesthesia

19.1 Introduction

The genome, epigenome, and microenvironment of each single cell in the organism are unique. The gene expression of the single-cell is based on fluctuations in the mechanisms of transcription and translation. The heterogeneity of cells is the basic nature of the homeostasis and development of the living body system to perform specific tasks and functions. It is necessary to define similarities and differences of different cells from morphology, protein level, and even gene level to understand the differences between cells. Single-cell RNA sequencing will become a new approach to monitor gene expression in clinical practice to explore gene expression profiles at the single-cell level using single-cell RNA sequencing (scRNA-seq). The scRNA-seq is a powerful tool to classify and identify cell subtypes [1], characterize rare or small cell populations, and track dynamics of cell-to-cell variations [2].

E. Chang · X. Ruan · R. Zhu · Y. Wang · J. Zhang (✉)
Department of Anesthesiology and Perioperative Medicine, Center of Clinical Single-Cell Biomedicine, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, P.R. China
e-mail: jqzhang@henu.edu.cn

B. Yu et al. (eds.), *Single-cell Sequencing and Methylation*, Advances in Experimental Medicine and Biology 1255, https://doi.org/10.1007/978-981-15-4494-1_19

The brain is one of the most complex tissues and is intensively investigated on brain cells, and current studies focus on location, morphology, electrophysiological property, target specificity, molecular biomarker, and gene expression pattern [3–6]. Thus, scRNA-seq becomes more important for the understanding of the brain contributions to learning, memory, and other cognitive functions [7]. scRNA-seq can make it possible to understand the heterogeneity and regulatory networks in brain cells at the single-cell level [8]. The present chapter will review recent studies on the use of scRNA-seq in brain cells and summarize the values of analysis method and the significance of results from scRNA-seq in brain cells. We emphasize the importance of clinical application of scRNA-seq in brain cells and potential challenges to be faced in future. We will comprehensively discuss the application of scRNA-seq in the development of the cerebral cortex to better understand the development and function of cerebral nervous system.

19.2 Single-Cell RNA Sequencing Technologies

The hotspot technology of scRNA-seq methods are summarized in Fig. 19.1 to assure the high-resolution analysis of individual cells unbiased and profound. The scRNA-seq contributes to reveal the heterogeneity, dynamics of transcription, and regulatory relationships between genes in a cell [9]. Developed scRNA-seq is applied to investigate the rare cell clusters, which may be omitted by traditional analysis. Specific features of these cells as well as the events of the interactions among cells are unveiled by scRNA-seq, rather than by previous high-throughput analysis [10]. scRNA-seq revealed the dynamic function of individual cell in developmental biology [11–16], neurobiology [17–20], immunology [21–24], and cancer research [25–29]. The landscape of single-cell tumor immune map accelerates the immune treatment on basis of high heterogeneity of immune cells in cancer and identifies molecular characterization of tumors in symptomatic and asymptomatic patients [30].

Using scRNA-seq, the differentiation fate of progenitor cells and the progress of individual cell are defined in the development, during which new cell clusters are distinguished by scRNA-seq [31–33]. The complexity of brain structure enables the delicate regulation in the developmental progress of brain. The nervous system is the most complex organ in mammal, where the cerebral cortex development is the main model system for neural developmental investigations and shares many consistent mechanisms with the developing brain and spinal cord. The Dll1/Notch, Nrg1/ErB, and Fgf10/Fgfr2 pathways were involved in this transformation of neuroepithelial stem cells into radial glial (RG) stem cells. Other transcriptions factors (Ap2γ, Ngn2, Insm1, or Tbr2) are discovered to activate the generation of basal progenitors from RGs, which was inhibited through the Notch and FGF pathways and the epigenetic regulator Ezh2, to differentiate RGs into astrocytes and lead to the termination of neurogenesis. Multiple signaling pathways (Jack/Stat, Notch, BMP, FGF) promote the neurogenic-to-glycogenic switch, although other signalings are still unclear. scRNA-seq is applied to investigate the novel mechanisms implicated in the cerebral cortex development (Fig. 19.2) and to characterize the cellular composition of the mouse cortex at development-embryonic day 14.5, representing a progenitor-driven stage and birth, when neurons corresponding to all six cortical layers were born and gliogenesis has begun. Distinct cortical layer-specific cell types and the spatial and temporal expression patterns of hallmark genes were assigned to 22 cell clusters and described (Fig. 19.3).

The distinct sensory neuronal types were dissected by scRNA-seq and Notch signaling is indispensable for brain development [17, 34]. Combining the scRNA-seq with electrophysiology, the development of embryonic hippocampal neurons and the neonatal cortical neuron cells were mapped and classified. The

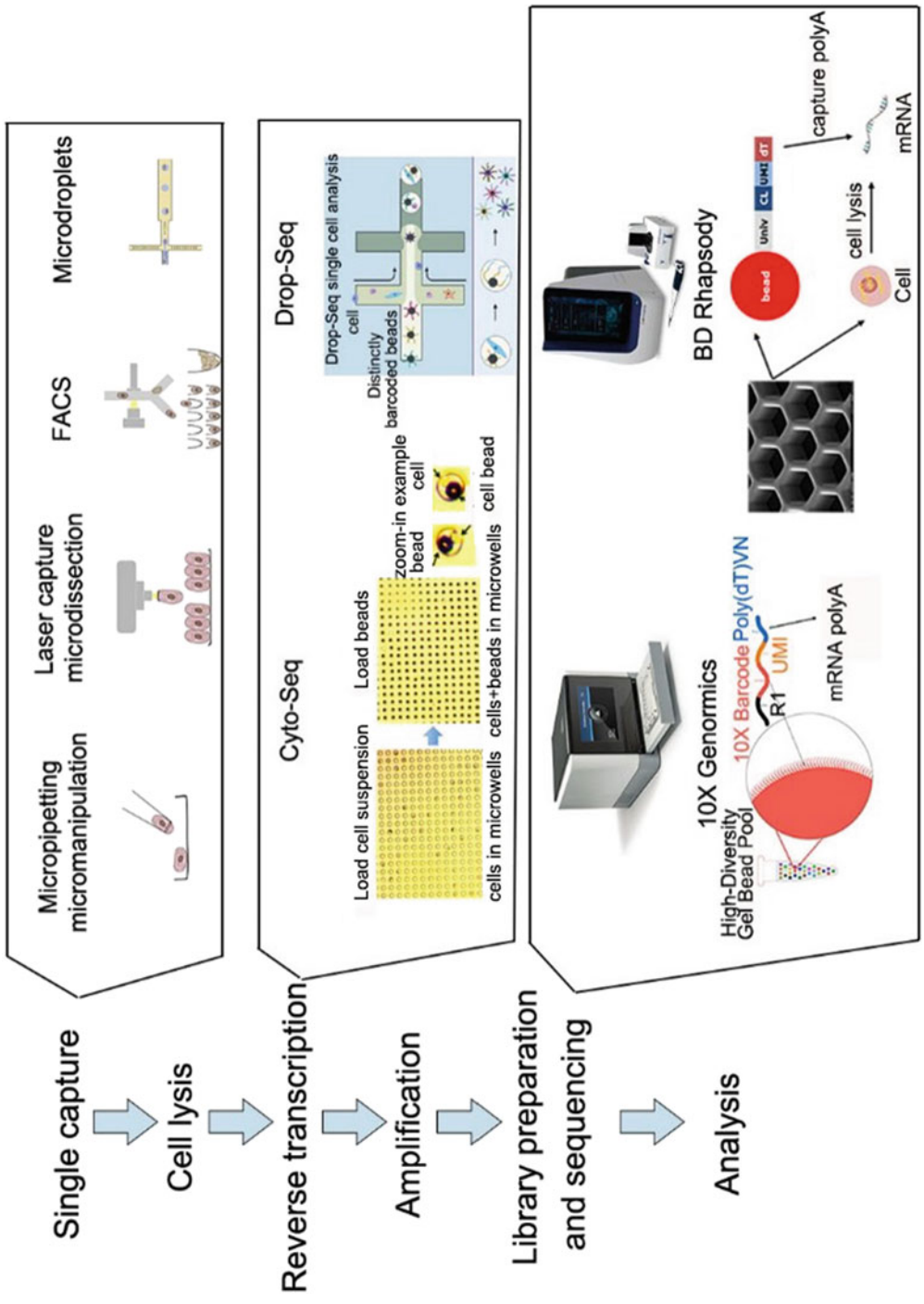


Fig. 19.1 Schematic of single-cell RNA-seq experiment showing each step and different experimental approaches [9]

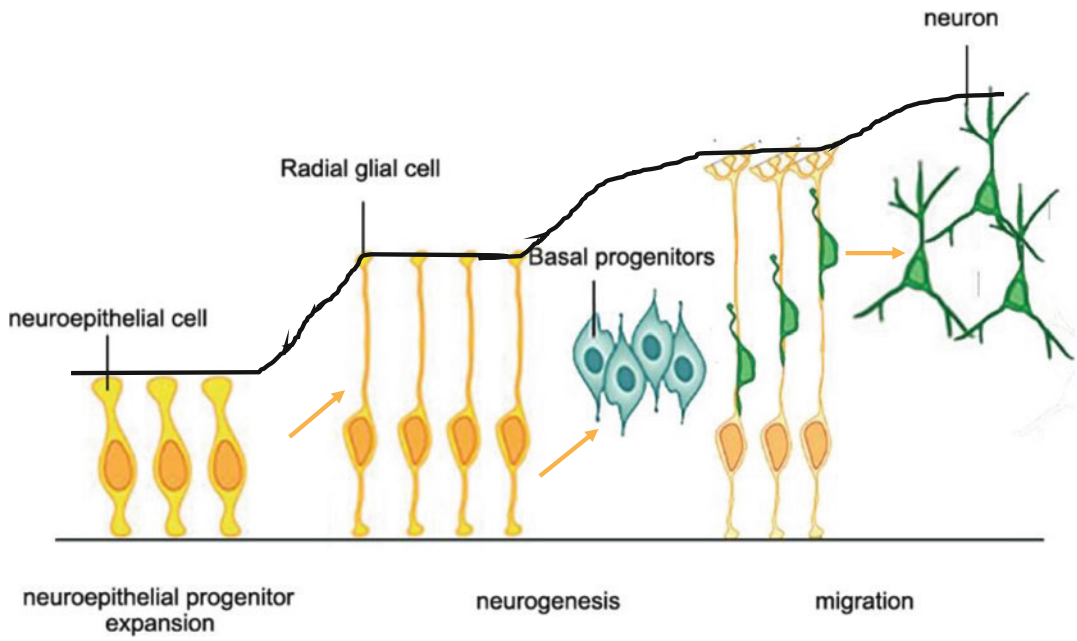


Fig. 19.2 Development of mouse cerebral cortex. Cerebral cortex develops via a complex sequence of cell proliferation, differentiation, and migration events [93]

new marker genes of the cerebral cortex during human development and the specific developmental characteristics, as well as the developmental timeline of excitatory neurons were found using scRNA-seq [19]. By the scRNA-Seq in situ, the neurons location and the difference between the neighbor neurons spatially were defined [35].

The single-cell transcriptomic profiles developed along with the qPCR development and single-molecule fluorescence in situ hybridization are the primary method of the transcripts analysis [36–39]. The whole-transcriptome analysis and the subsequent RNA sequencing are adapted for analyzing single cells [40–43]. The scRNA-Seq was applied to investigate the early embryonic development and global patterns of gene expression variations [44, 45], although the amount of biological materials were a limiting factor [2]. The unbiased analyzing of scRNA-Seq is adapted for the hundreds of thousands of cells endowed of heterogeneity [46].

The methods for capturing single cells from enormous cells include mouth pipetting, serial dilution, robotic micromanipulation, and flow-assisted cell sorting; while the methods for isolating rare single cells still lag behind, including Nanofilters [47], MagSweeper [17], Laser-capture microdissection, CellSearch [48], CellCelector [49], and DEP-Array [50]. It is necessary to amplify the RNA sequence for scRNA-seq due to that the total RNA in a mammalian cell is only 10 pg and the mRNA is only 0.1 pg. SMART-Seq is a whole-transcriptome amplification (WTA) method performed using Moloney Murine Leukemia Virus (MMLV) reverse transcriptase to protect the full-length amplification from the strong plague of the 3′ mRNA bias. MMLV has both template-switching and terminal transferase activity, which leads to the addition of contemplated cytosine residues to the 5′ end of the cDNA [1]. The templates can be switched by MMLV and transcribe the other strand to amplify the full-length cDNA transcripts by adding a poly

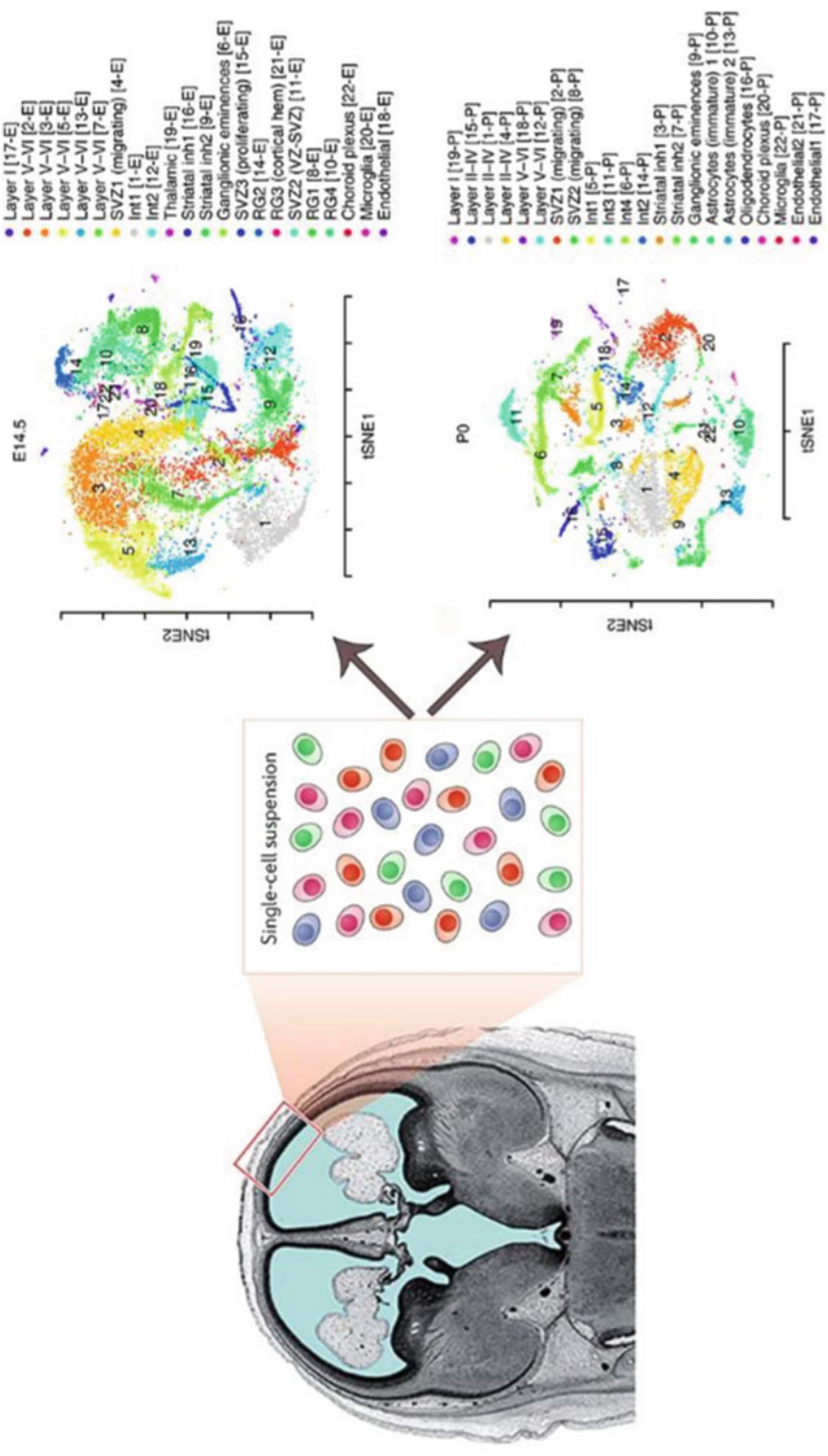


Fig. 19.3 Cell clusters analysis of developing cerebral cortex by scRNA-seq. t-SNE visualization of the 22 clusters identified in the E14.5 and P0 cortex [94, 95]

(G) template with an adaptor sequence. To overcome the distortion of the mRNA strand during the amplification progress, cDNA is labeled with barcodes thus specific cDNA sequence is assigned to the specific cells [51–54]. With the revolutionary development of technologies, MARS-Seq, Cyto-Seq, Drop-Seq, inDrop, and the scRNA-Seq are promoted extensively.

19.3 scRNA-seq and Developing Cerebral Cortex

The diversity, function and the range of transcriptional regulation are studied on brain cells from the cerebral cortex with RNA-seq. As an emerging tool, scRNA-seq is gradually applied to study the complexity of brain cells, new cell populations, specific genetic characteristics, and potential regulatory networks [55, 56]. Brain cells include highly complex nerve cell types/subtypes with special morphology, excitability, connectivity, and cell location [57]. Different neuronal cell types and subtypes, and new cell-specific markers were found using scRNA-seq. For example, Amit et al. extracted 3005 single cells from the cerebral cortex and hippocampus from mice, performed scRNA-seq analysis, and found nine major cell populations, including S1 and CA1 vertebral neurons, transfer neurons, oligodendrocytes, astrocytes, microglia, vascular endothelial cells, parietal cells, and ependymal cells. Novel and specific molecular markers of different cell types were also discovered, e.g., Gmll549 for S1 vertebral neurons, Pnoc for transferred neurons.

The individual adult neurons are freshly isolated from a limited regional sample of neurosurgical tissue [58]. Freshly separated neurosurgical tissue is better for analyzing individual neurons, while more samples of postmortem tissues are available in clinical practice. Lake et al. developed a new method of the neuron nucleus and RNA sequencing for the brain, separated the 3227 individual neurons from six different regions of the cerebral cortex for RNA sequencing, and found 16 subtypes of neurons in the cortex cells with molecular biomarkers. The cerebral cortical neurons are evaluated and developed originally from the subependymal neural progenitors, and neural progenitor cells complete the development process of the cerebral cortex through the proliferation, differentiation, and migration. The temporal and spatial characteristics of cerebral cortex development from rodents to primates include differentiation characteristics from progenitors to various types of neurons, with a clear relationship between mental disorders and target gene expression.

The developmental mechanisms of the cortex with different functional divisions are involved in intrinsic and extrinsic biological mechanisms. For example, Gbx2 contributes to the development of normal cortical regions, rather than thalamic cortical projections. Providing the first clear evidence that thalamic innervation is not necessary for the basic generation of cortical area maps, and the formation of cortical regions depends primarily on the mechanisms within the telencephalon [59]. The intrinsic mapping center of the telencephalon controls the size and location of the cortical region. The morphine and fibroblast growth factor 8 (FGF8) released from the commissural plate at the distal prefrontal tip of the brain periphery can be adjusted along the frontal tail. Other factors play complementary roles, such as Fgf17 [60], BMP [61], and Wnt [62]. The thalamic input is required to establish genetic and functional divisions between the primary and adjacent high-grade sensory cortexes [63]. In addition, in-cortical self-generating activities also contribute to the formation of cortical columns [64] (Fig. 19.4).

The precise gene spatiotemporal expression profile of the cerebral cortex is important for the evolution, development, and function of the nervous system [65–67]. The temporal and spatial characteristics of gene expression during cortical development are different from genetic characteristics of different brain regions in the same period, or developmental stages of the same brain region. Using immunofluorescence technology, the diversity of Drosophila neurons was found to be dependent upon the integration of time and space patterns [68]. The temporal and

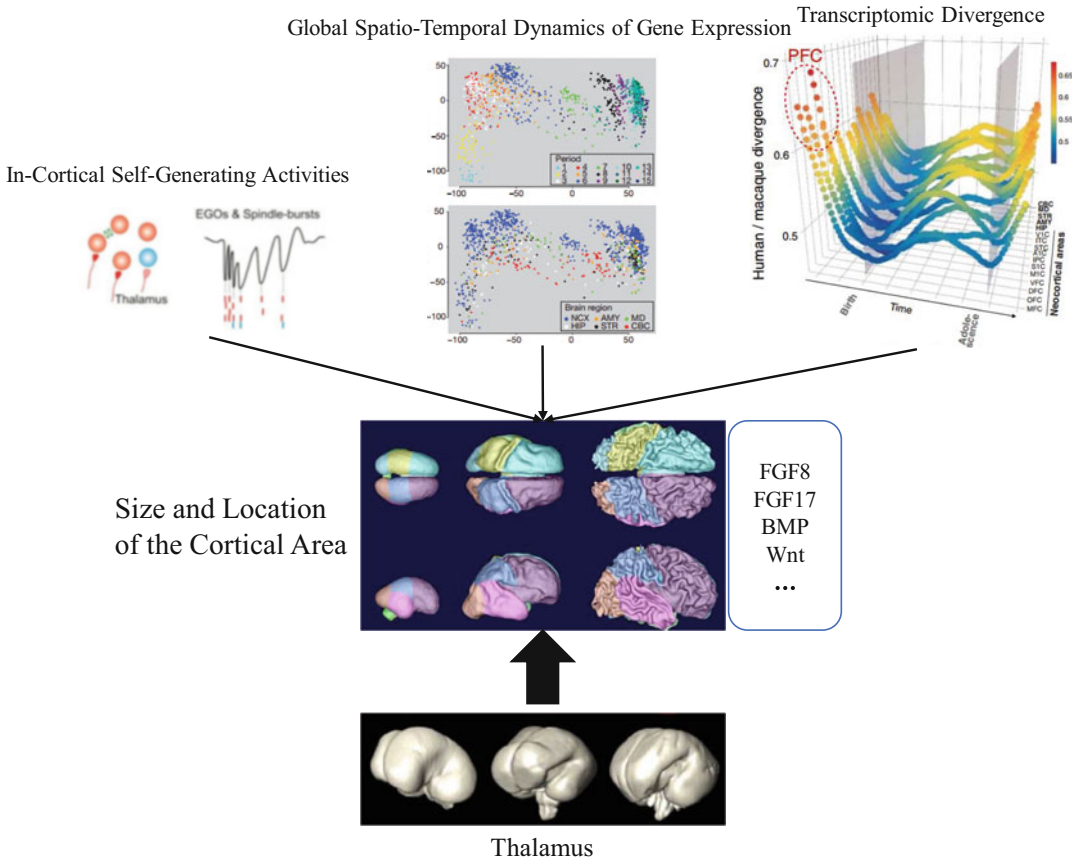


Fig. 19.4 Factors involved in cerebral cortex development. The formation of cortical regions depends mainly on the mechanisms within the cortex, self-generating activities, temporal-spatial differences in gene expression, and the involvement of the thalamus in cortical development [18, 60–64, 96, 97] (Credit: Studholme Lab/UWMC)

spatial characteristics of mouse brain development were defined with a new algorithm by detecting repetitive patterns in spatiotemporal gene expression data of developing mouse brains. Expression patterns can reveal regional differences in brain development [69]. Previous studies analyzed exon-level transcriptome data from multiple brain regions and neocortex regions of developmental brain and adult brain by transcriptome sequencing. About 90% of expressed genes were in the whole-transcriptome or exogenous sublevels are different before birth, and then the similarity of transcriptomes in the same region increases, forming different co-expression networks [70].

Systematic analysis of temporal and spatial gene expression trajectories during cerebral cortical development because of the coexistence of multiple cell types in emerging tissues at different stages of maturation and differentiation. The scRNA-seq of primary and medial ganglionic eminence (MGE) micro-dissected from germinal zone and cortical plate samples was performed at various stages of peak neuronal firing from progenitor cells to neuronal differentiation during post-mitotic neuronal differentiation [71]. Those cortex areas mainly include DFC, HIP, AMY, STR, MD, and CBC from the 60 days post-conception to the 11-year-old macaque. Human-specific transcriptomics and

spatiotemporal transcription specificity were associated with neurological diseases such as autism and schizophrenia, synapse formation, and neuronal development [72].

The classification of neurons is based on morphological, chemical, and electrophysiological differences [73, 74, 75], as well as the different molecular features that researchers use to study. The large-scale characterization of approximately 1000 neurologically important gene expressions was defined with in situ hybridization in the visual and adult temporal cortex of the adult brain prior to the application of scRNA-seq. Changes in gene expression profiles distinguish cortical functions between species [74]. Compared with scRNA-seq, those methods have significantly lower flux and higher workload. For example, the isolation 49 cells from mouse cerebral cortex and hippocampus, including 23 GABAergic neurons, 19 glutamatergic neurons, and 7 non-neuronal cells, took much longer time in the previous study [76], as compared with clarification of mouse somatosensory cortex and hippocampal CA1 region cells using scRNA-seq. About 47 different subclasses of molecules were identified and corresponded to cell types, morphologies, and locations. The authors found a layer I neuron expressing Pax6 and a unique mitotic oligodendrocyte subclass labeled by Itpr2. The diversity of cortical cell types on basis of transcription factors forms a complex hierarchical surveillance to demonstrate mechanisms for maintaining adult cell type identity [73]. Three thousand two hundred and twenty-seven data sets of single nerves from six different regions of the cerebral cortex were generated using a scalable approach to sequencing and quantifying RNA molecules in neuronal nuclei isolated from postmortem brains. Using iterative clustering and classification methods, 16 neuronal subtypes were identified by annotating biomarkers and cortical cell structures [77]. The cortex of different brain regions of human embryos was analyzed using scRNA-seq at 22 and 23 weeks. The distribution characteristics of local gene expression and neuronal maturation were evaluated using the modified STRT-seq method in human cerebral cortex [18]. With the advancement of scRNA-seq, more than 20,000 cells in the primary visual cortex and motor cortex of adult mice were sequenced and 133 cell types were identified. Excitatory neurons are regionally specific, and one subtype is only distributed in a certain cortex, where different subtypes also show different long-range projection modes by retrograde labeling [78].

Cortical development is experienced from progenitors to complex neural networks. The mechanisms by which neuronal diversity and the connections form complex neural networks can improve the discovery of pathogenic targets of neurodevelopmental disorders [79]. As the major components of a complex neural network, excitatory neurons migrate to the cortical plate, form six cortical layers with a stereotyped connection pattern, and contribute to the configurations of the functional circuit [71]. Those neurons account for approximately 80% of neurons in the cortex and interact with a small number of inhibitory cortical neurons in important ways [79]. In the cerebral cortex, GABAergic interneurons are evolved into high heterogeneity of cell types with unique temporal and spatial capabilities to affect neuronal circuits. Up to 50 different types of GABAergic neurons are distributed in the cerebral cortex and derived from the subcutaneous progenitors in the ventral surface of the lower cerebral ventricle. Internal neuronal diversity occurs through the implementation of intracellular genetic processes in progenitor cells over a longer period of time until the internal neurons acquire mature features [80]. A few precursor cells of inhibitory neurons are present in the early prefrontal cortex using scRNA-seq, of which the most are in the cell cycle, except for during the rest period [19]. Glial cells, including astrocytes, oligodendrocyte Schwann cells, and microglia, do not transmit transmitters like neurons, but form the environment where neurons survive and form neural networks [81]. Their dysfunction is associated with neurological diseases [82]. Human glial cells wrap more than single astrocytes in mice which can wrap more than 100,000 synaptic structures [82], support the role of neurons, and participate in the

development of the nervous system, plasticity, and disease. Glial cells are also involved in synapse formation, regulating synaptic function, and blood flow [83].

Many long-chain noncoding RNAs (lncRNA) express in brain tissue and regulate neuronal function, responsible for the development of diseases. lncRNA exhibits stronger in the tissue and cell specificity, as compared to mRNA [84]. The subtype-dependent enrichment of lncRNA was noticed in cortical progenitors isolated and sequenced from in human fetal brain cerebral cortex within 4 h after autopsy [85]. Liu et al. deeply dissected lncRNA from polyadenylated and total RNA obtained at different developmental stages in human neocortex using strand-specific RNA-seq and analyzed the transcriptome of individual cells. Single-cell transcriptomics of hundreds of neocortical cells revealed that many lncRNAs abundantly expressed in a single cell and are cell type specific. Among those, LOC646329 is a lncRNA rich in single radial glial cell and regulates cell proliferation [86]. A variety of lncRNAs are involved in the cellular processes of brain development and the spatiotemporal expression of lncRNA in a cohort of 13 lncRNA null mutant mouse models showed different between developing and adult brains, between transcriptomes and phenomes, between temporal and spatial brain development, and between selected and non-selected brain regions. Among those, a variety of cellular pathways and processes changed after deletion of the lncRNA locus, and four of the lncRNAs affected the expression of adjacent protein-coding genes in a cis-like manner [87]. In addition to lncRNA, microRNAs play an important role in posttranscriptional regulation and complexity during brain development and are considered as important triggers of brain development and neurological or psychiatric diseases [88]. The limited understanding of in vivo miRNA targets and their intensity in single cells makes it difficult to define miRNA-mRNA networks. Single-cell analysis using binary and co-expression networks is carried out by combining high-throughput sequencing of RNA and immunoprecipitation-cross-linked immunoprecipitation with AGO2 antibody (AGO2-HITS-CLIP) The miRNA-mRNA interaction as a functional module undergoes dynamic transformation during brain development and shows cell-specific and highly dynamic during development and throughout the evolution process. For example, the interaction between ORC4 and miR-2115 abundant in radial glial cells, can control the proliferation rate of radial glial cells during human brain development [89] (Fig. 19.5). Those studies on brain cells based on scRNA-seq enable us to better understand the similarities and differences between brain cells.

19.4 Application of scRNA-seq in Neurologic Diseases

The scRNA-seq as a powerful tool can provide new insights for understanding molecular mechanisms of the functional and dysfunctional regulations in a single neuron. It is the time to deeply understand the occurrence and development of neuro-related diseases, e.g., degenerative diseases and brain tumors, at the level of single-cell transcriptome, find disease-specific cell populations, and discover target genes for therapy.

Alzheimer's disease (AD) is a harmful neurodegenerative disease without effective treatments, due to the heterogeneity among neurocytes and among immune cells associated with the development and progression of AD. A microglial cell type with molecular markers, spatial locations, and signaling pathways was identified to contribute to neurodegenerative disease, accompanied with AD-associated brain immune cell populations. Such specific microglia specifically present in neurodegenerative diseases, with great potential for the future treatment of AD and other neurological diseases.

Glial cells and stem cells undergo genetic mutations during development and may develop human gliomas, due to the composition of different apparent states and cell types. In recent years, more and more single-cell RNA sequencing has

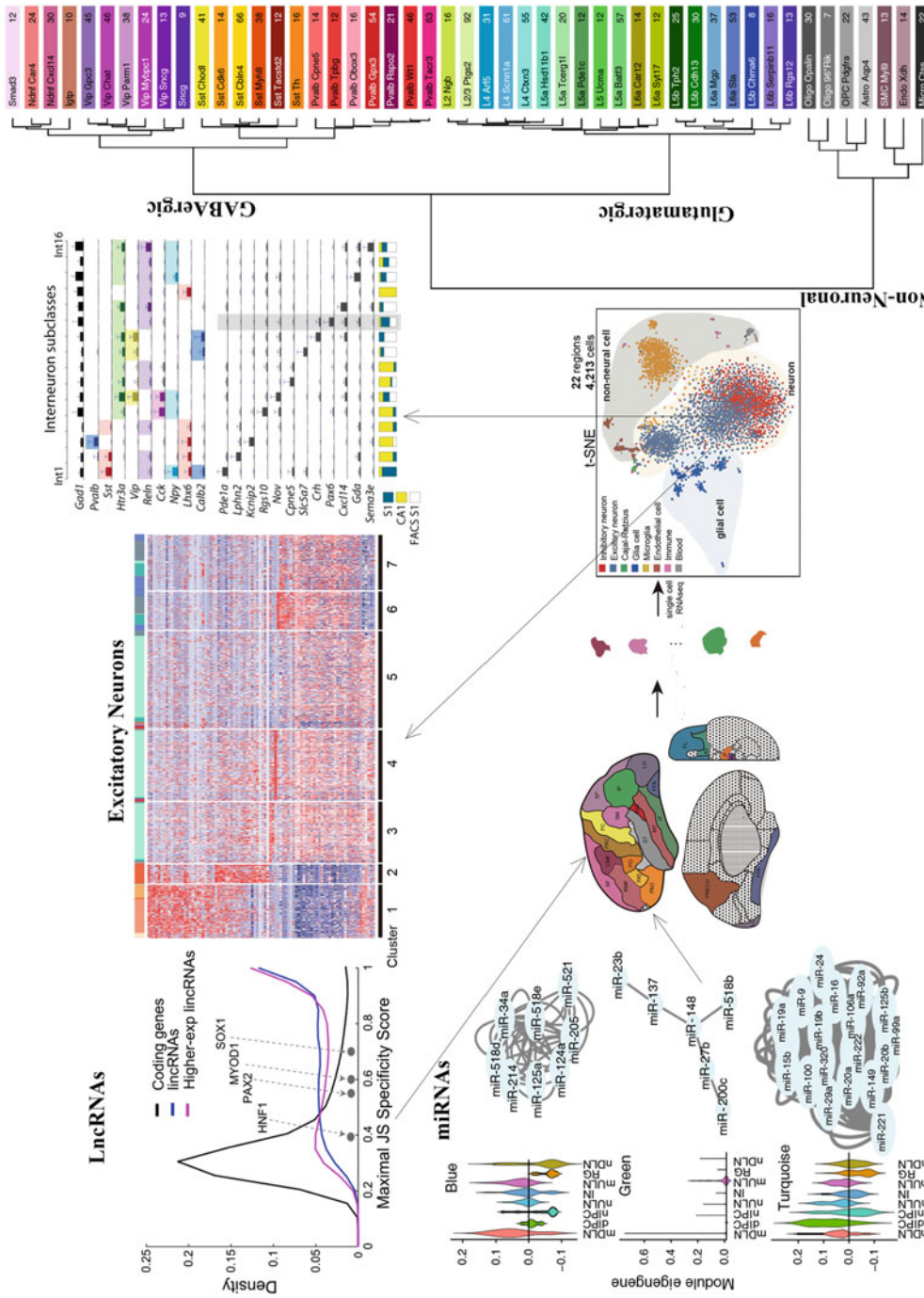


Fig. 19.5 Noncoding RNA (LncRNA and miRNA) characteristics and differentiation of cerebral cortical cells excitatory neurons and interneurons [19, 73, 76, 84, 89]

been applied in the study of glioma. Patel et al. [90] found that 430 cells from 5 primary glioma samples were associated with oncogene signaling, proliferation, immune response, and hypoxic signaling using scRNA-seq. The clear heterogeneity among intra-glioma cell subtypes was identified, as an important breakthrough in understanding the heterogeneity, biology, prognosis, and treatment of glioma.

Tumor stem cells can drive the growth of tumor cells, but there is no particularly good evidence for the existence of tumor stem cells in human solid tumors. Tirosh et al. analyzed 4347 individual oligodendrocytes using scRNA-seq, and found that the majority of tumor cells differentiated into two specific glia oligodendrocytes and astrocytes as well as a small number of cells in the undifferentiated state ranged from genome expression level to the development process and associated with cancer stem cell signaling pathways. Proliferating characteristics of gene defined were consistent with the tumor stem cells promoting tumor growth in human oligodendroglioma. This scRNA-seq provides insight into the developmental structure of oligodendroglioma at the single-cell level and strong support for disease treatment.

19.5 Limitation

Transcriptome features of glutamatergic neurons vary widely among cortical regions. It is questioned whether each of those transcriptome features represents a unique cell type or reflects a heterogeneous transcriptional state in a single projection neuron that may be affected by a series of nerve activity and other factors [91, 92]. By combining the transcriptome profile to other phenotypes, e.g., morphology, electrophysiological properties, and function, more molecular subtypes and phenomes of neurons are characterized using scRNA-seq. The gene expression in the cortical cellular and molecular networks at a single cell will generate important information to determine the molecular interactions between the connected genome and transcriptome within a cell. There are urgent needs to furthermore explore intercellular and intermolecular heterogeneity, the degree of selectivity and differentiation of cortical projections, and disease-specific biomarkers and mechanisms of circuit development and maturity.

19.6 Summary and Prospect

scRNA-seq is a powerful tool for exploring the complexity, clusters, and specific functions of the brain cells. Using scRNA-seq, the heterogeneity and changes in transcriptomic profiles of a single neuron were defined during dynamic development and differentiation of cells in cerebral cortex regions, and in the pathogenesis of neurological diseases. One of the great challenges is that the brain sample is susceptible to interference and confounding. More advanced methodologies of computational systems biology need to be developed to overcome the inherent interference and technical differences in the detection of single-cell signals. It is expected that scRNA-seq will be extended to metabolic profiles of the single neuron cell on basis of transcriptional profiles and regulatory networks. It is also expected if the transcriptional profiles can be integrated with molecular and functional phenomes in a single neuron and with disease-specific phenomes to understand molecular mechanisms of brain development and disease occurrence. scRNA-seq will provide the new emerging neurological disciple f of the artificial intelligent single neuron for deep understanding of brain diseases.

References

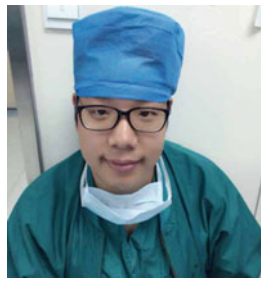
1. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30 (8):777–782. https://doi.org/10.1038/nbt.2282
2. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X et al (2010) Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell

RNA-Seq analysis. Cell Stem Cell 6(5):468–478. https://doi.org/10.1016/j.stem.2010.03.015
3. DeFelipe J, Lopez-Cruz PL, Benavides-Piccione R, Bielza C, Larranaga P, Anderson S et al (2013) New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat Rev Neurosci 14(3):202–216. https://doi.org/10.1038/nrn3444
4. Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci 8(6):427–437. https://doi.org/10.1038/nrn2151
5. Gelman DM, Marin O (2010) Generation of interneuron diversity in the mouse cerebral cortex. Eur J Neurosci 31(12):2136–2141. https://doi.org/10.1111/j.1460-9568.2010.07267.x
6. Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D et al (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 9(1):e1000582. https://doi.org/10.1371/journal.pbio.1000582
7. Cheah CS, Lundstrom BN, Catterall WA, Oakley JC (2019) Impairment of sharp-wave ripples in a murine model of Dravet syndrome. J Neurosci 39 (46):9251–9260. https://doi.org/10.1523/JNEUROSCI.0890-19.2019
8. Handley A, Schauer T, Ladurner AG, Margulies CE (2015) Designing cell-type-specific genome-wide experiments. Mol Cell 58(4):621–631. https://doi.org/10.1016/j.molcel.2015.04.024
9. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58(4):610–620. https://doi.org/10.1016/j.molcel.2015.04.005
10. Chattopadhyay PK, Gierahn TM, Roederer M, Love JC (2014) Single-cell technologies for monitoring immune systems. Nat Immunol 15(2):128–135. https://doi.org/10.1038/ni.2796
11. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ et al (2019) A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 566(7745):490–495. https://doi.org/10.1038/s41586-019-0933-9
12. Zeng Y, He J, Bai Z, Li Z, Gong Y, Liu C et al (2019) Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. Cell Res 29(11):881–894. https://doi.org/10.1038/s41422-019-0228-6
13. Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y et al (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500(7464):593–597. https://doi.org/10.1038/nature12364
14. Yan L, Yang M, Guo H, Yang L, Wu J, Li R et al (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol 20(9):1131–1139. https://doi.org/10.1038/nsmb.2660
15. Cheng S, Pei Y, He L, Peng G, Reinius B, Tam PPL et al (2019) Single-cell RNA-Seq reveals cellular heterogeneity of pluripotency transition and X chromosome dynamics during early mouse development. Cell Rep 26(10):2593–2607.e3. https://doi.org/10.1016/j.celrep.2019.02.031
16. Clark BS, Stein-O'Brien GL, Shiau F, Cannon GH, Davis-Marcisak E, Sherman T et al (2019) Single-cell RNA-Seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. Neuron 102(6):1111–1126.e5. https://doi.org/10.1016/j.neuron.2019.04.010
17. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH et al (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 32(10):1053–1058. https://doi.org/10.1038/nbt.2967
18. Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L et al (2018) Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-Seq analysis. Cell Res 28(7):730–745. https://doi.org/10.1038/s41422-018-0053-3
19. Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J et al (2018) A single-cell RNA-Seq survey of the developmental landscape of the human prefrontal cortex. Nature 555(7697):524–528. https://doi.org/10.1038/nature25980
20. Tiklova K, Bjorklund AK, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L et al (2019) Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. Nat Commun 10(1):581. https://doi.org/10.1038/s41467-019-08453-1
21. Jordao MJC, Sankowski R, Brendecke SM, Sagar, Locatelli G, Tai YH et al (2019) Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation. Science 363(6425):eaat7554. https://doi.org/10.1126/science.aat7554
22. Chapuy L, Bsat M, Sarkizova S, Rubio M, Therrien A, Wassef E et al (2019) Two distinct colonic CD14(+) subsets characterized by single-cell RNA profiling in Crohn's disease. Mucosal Immunol 12(3):703–719. https://doi.org/10.1038/s41385-018-0126-0
23. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A et al (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 20(2):163–172. https://doi.org/10.1038/s41590-018-0276-y
24. Ciucci T, Vacchio MS, Gao Y, Tomassoni Ardori F, Candia J, Mehta M et al (2019) The emergence and functional fitness of memory CD4(+) T cells require the transcription factor thpok. Immunity 50 (1):91–105.e4. https://doi.org/10.1016/j.immuni.2018.12.019
25. Cheng YH, Chen YC, Lin E, Brien R, Jung S, Chen YT et al (2019) Hydro-Seq enables contamination-free high-throughput single-cell RNA-sequencing for circulating tumor cells. Nat Commun 10(1):2163. https://doi.org/10.1038/s41467-019-10122-2

26. Caruso S, Calatayud AL, Pilet J, La Bella T, Rekik S, Imbeaud S et al (2019) Analysis of liver cancer cell lines identifies agents with likely efficacy against hepatocellular carcinoma and markers of response. Gastroenterology 157(3):760–776. https://doi.org/10.1053/j.gastro.2019.05.001
27. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ et al (2019) Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. Cell 176(4):775–789.e18. https://doi.org/10.1016/j.cell.2018.11.043
28. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q et al (2018) Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med 24(7):978–985. https://doi.org/10.1038/s41591-018-0045-3
29. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S et al (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell 174(5):1293–1308.e36. https://doi.org/10.1016/j.cell.2018.05.060
30. Ledergor G, Weiner A, Zada M, Wang SY, Cohen YC, Gatt ME et al (2018) Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. Nat Med 24(12):1867–1876. https://doi.org/10.1038/s41591-018-0269-2
31. Tang Q, Iyer S, Lobbardi R, Moore JC, Chen H, Lareau C et al (2017) Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. J Exp Med 214 (10):2875–2887. https://doi.org/10.1084/jem.20170976
32. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A et al (2016) A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst 3(4):346–360.e4. https://doi.org/10.1016/j.cels.2016.08.011
33. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J et al (2017) Single-cell RNA-Seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356 (6335):eaah4573. https://doi.org/10.1126/science.aah4573
34. Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, Lou D et al (2015) Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat Neurosci 18(1):145–153. https://doi.org/10.1038/nn.3881
35. Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S et al (2014) Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. Nat Methods 11(2):190–196. https://doi.org/10.1038/nmeth.2804
36. Femino AM, Fay FS, Fogarty K, Singer RH (1998) Visualization of single RNA transcripts in situ. Science 280(5363):585–590. https://doi.org/10.1126/science.280.5363.585
37. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4(10):e309. https://doi.org/10.1371/journal.pbio.0040309
38. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5(10):877–879. https://doi.org/10.1038/nmeth.1253
39. Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14 (3):303–308. https://doi.org/10.1038/nbt0396-303
40. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 5(7):613–619. https://doi.org/10.1038/nmeth.1223
41. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18(9):1509–1517. https://doi.org/10.1101/gr.079558.108
42. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320(5881):1344–1349. https://doi.org/10.1126/science.1158441
43. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I et al (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453(7199):1239–1243. https://doi.org/10.1038/nature07002
44. Deng Q, Ramskold D, Reinius B, Sandberg R (2014) Single-cell RNA-Seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343(6167):193–196. https://doi.org/10.1126/science.1245316
45. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2(3):666–673. https://doi.org/10.1016/j.celrep.2012.08.003
46. Fan HC, Fu GK, Fodor SP (2015) Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. Science 347 (6222):1258367. https://doi.org/10.1126/science.1258367
47. Adams DL, Martin SS, Alpaugh RK, Charpentier M, Tsai S, Bergan RC et al (2014) Circulating giant macrophages as a potential biomarker of solid tumors. Proc Natl Acad Sci U S A 111(9):3514–3519. https://doi.org/10.1073/pnas.1320198111
48. Yu M, Stott S, Toner M, Maheswaran S, Haber DA (2011) Circulating tumor cells: approaches to isolation and characterization. J Cell Biol 192(3):373–382. https://doi.org/10.1083/jcb.201010021
49. Choi JH, Ogunniyi AO, Du M, Du M, Kretschmann M, Eberhardt J et al (2010) Development and optimization of a process for automated recovery of single cells identified by microengraving.

Biotechnol Prog 26(3):888–895. https://doi.org/10.1002/btpr.374
50. Altomare L, Borgatti M, Medoro G, Manaresi N, Tartagni M, Guerrieri R et al (2003) Levitation and movement of human tumor cells using a printed circuit board device based on software-controlled dielectrophoresis. Biotechnol Bioeng 82(4):474–479. https://doi.org/10.1002/bit.10590
51. Macaulay IC, Voet T (2014) Single cell genomics: advances and future perspectives. PLoS Genet 10(1): e1004126. https://doi.org/10.1371/journal.pgen.1004126
52. Saliba AE, Westermann AJ, Gorski SA, Vogel J (2014) Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res 42(14):8845–8860. https://doi.org/10.1093/nar/gku555
53. Sandberg R (2014) Entering the era of single-cell transcriptomics in biology and medicine. Nat Methods 11(1):22–24. https://doi.org/10.1038/nmeth.2764
54. Potter SS (2018) Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol 14(8):479–492. https://doi.org/10.1038/s41581-018-0021-7
55. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S et al (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34 (36):11929–11947. https://doi.org/10.1523/JNEUROSCI.1860-14.2014
56. Yan Q, Weyn-Vanhentenryck SM, Wu J, Sloan SA, Zhang Y, Chen K et al (2015) Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. Proc Natl Acad Sci U S A 112 (11):3445–3450. https://doi.org/10.1073/pnas.1502849112
57. Rappaport B, Mellon RD, Simone A, Woodcock J (2011) Defining safe use of anesthesia in children. N Engl J Med 364(15):1387–1390. https://doi.org/10.1056/NEJMp1102155
58. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM et al (2015) A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci U S A 112(23):7285–7290. https://doi.org/10.1073/pnas.1507125112
59. Rubenstein JL, Anderson S, Shi L, Miyashita-Lin E, Bulfone A, Hevner R (1999) Genetic control of cortical regionalization and connectivity. Cereb Cortex 9 (6):524–532. PMID: 10498270
60. Cholfin JA, Rubenstein JLR (2007) Patterning of frontal cortex subdivisions by Fgf17. Proc Natl Acad Sci U S A 104(18):7652–7657. https://doi.org/10.1073/pnas.0702225104
61. Hébert JM, Mishina Y, Mcconnell SK (2002) BMP signaling is required locally to pattern the dorsal telencephalic midline. Neuron 35(6):1029–1041. https://doi.org/10.1016/s0896-6273(02)00900-5
62. Grove EA, Tole S, Limon J, Yip L, Ragsdale CW (1998) The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. Development 125(12):2315–2325. PMID: 9584130
63. Gabrielle P, Frédéric G, Camilla B, Ludovic T, Ilaria V, Christian L et al (2014) Modality-specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons. Nature 511(7510):471–474. https://doi.org/10.1038/nature13390
64. Luhmann HJ, Khazipov R (2017) Neuronal activity patterns in the developing barrel cortex. Neuroscience 368:256–267. https://doi.org/10.1016/j.neuroscience.2017.05.025
65. Hill RS, Walsh CA (2005) Molecular insights into human brain evolution. Nature 437(7055):64–67. https://doi.org/10.1038/nature04103
66. Fish JL, Dehay C, Kennedy H, Huttner WB (2008) Making bigger brains-the evolution of neural-progenitor-cell division. J Cell Sci 121 (Pt 17):2783–2793. https://doi.org/10.1242/jcs.023465
67. Lui JH, Hansen DV, Kriegstein AR (2011) Development and evolution of the human neocortex. Cell 146 (1):18–36. https://doi.org/10.1016/j.cell.2011.06.030
68. Erclik T, Li X, Courgeon M, Bertet C, Chen Z, Baumert R et al (2017) Integration of temporal and spatial patterning generates neural diversity. Nature 541(7637):365. https://doi.org/10.1038/nature20794
69. Chou SJ, Wang C, Sintupisut N, Niou ZX, Lin CH, Li KC et al (2016) Analysis of spatial-temporal gene expression patterns reveals dynamics and regionalization in developing mouse brain. Sci Rep 6(1):19274. https://doi.org/10.1038/srep19274
70. Hyo Jung K, Yuka Imamura K, Feng C, Ying Z, Xuming X, Mingfeng L et al (2011) Spatio-temporal transcriptome of the human brain. Nature 478 (7370):483. https://doi.org/10.1038/nature10523
71. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di LE et al (2017) Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358 (6368):1318. https://doi.org/10.1126/science.aap8809
72. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE et al (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87(6):1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016
73. Amit Z, Muñoz-Manchado AB, Simone C, Peter LN, Gioele LM, Anna J et al (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347 (6226):1138–1142. https://doi.org/10.1126/science.aaa1934
74. Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ et al (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell 149(2):483–496. https://doi.org/10.1016/j.cell.2012.02.052
75. Poulin JF, Tasic B, Hjerling-Leffler J, Trimarchi JM, Awatramani R (2016) Disentangling neural cell

diversity using single-cell transcriptomics. Nat Neurosci 19(9):1131–1141. https://doi.org/10.1038/nn.4366
76. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z et al (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 19(2):335–346. https://doi.org/10.1038/nn.4216
77. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R et al (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352(6293):1586. https://doi.org/10.1126/science.aaf1204
78. Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D et al (2018) Shared and distinct transcriptomic cell types across neocortical areas. Nature 563(7729):72–78. https://doi.org/10.1038/s41586-018-0654-5
79. Kast RJ, Levitt P (2019) Precision in the development of neocortical architecture: from progenitors to cortical networks. Prog Neurobiol 175:77–95. https://doi.org/10.1016/j.pneurobio.2019.01.003
80. Lim L, Mi D, Llorca A, Marín O (2018) Development and functional diversification of cortical interneurons. Neuron 100(2):294–313. https://doi.org/10.1016/j.neuron.2018.10.009
81. Simons M, Trajkovic K (2006) Neuron-glia communication in the control of oligodendrocyte function and myelin biogenesis. J Cell Sci 119(Pt 21):4381–4389. https://doi.org/10.1242/jcs.03242
82. Zuchero JB, Barres BA (2015) Glia in mammalian development and disease. Development 142 (22):3805–3809. https://doi.org/10.1242/dev.129304
83. Barres BA (2008) The mystery and magic of glia: a perspective on their roles in health and disease. Neuron 60(3):430–440. https://doi.org/10.1016/j.neuron.2008.10.013
84. Ma Q, Chang HY (2016) Single-cell profiling of lncRNAs in the developing human brain. Genome Biol 17(1):1–2. https://doi.org/10.1186/s13059-016-0933-0
85. Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL et al (2015) Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. Nat Neurosci 18(5):637. https://doi.org/10.1038/nn.3980
86. Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ et al (2016) Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol 17(1):67. https://doi.org/10.1186/s13059-016-0932-1
87. Goff LA, Groff AF, Sauvageau M, Trayes-Gibson Z, Sanchez-Gomez DB, Morse M et al (2015) Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 112(22):6855–6862. https://doi.org/10.1073/pnas.1411263112
88. Chen W, Qin C (2015) General hallmarks of microRNAs in brain evolution and development. RNA Biol 12(7):701–708. https://doi.org/10.1080/15476286.2015.1048954
89. Nowakowski TJ, Rani N, Golkaram M, Zhou HR, Alvarado B, Huch K et al (2018) Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. Nat Neurosci 21 (12):1784–1792. https://doi.org/10.1038/s41593-018-0265-3
90. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344(6190):1396–1401. https://doi.org/10.1126/science.1254257
91. Chen HH, Arlotta P (2016) Seq-ing the cortex one neuron at a time. Nat Neurosci 19(2):179. https://doi.org/10.1038/nn.4230
92. Chevée M, Robertson JDJ, Cannon GH, Brown SP, Goff LA (2018) Variation in activity state, axonal projection, and position define the transcriptional identity of individual neocortical projection neurons. Cell Rep 22(2):441–455. https://doi.org/10.1016/j.celrep.2017.12.046
93. Pilaz LJ, Silver DL (2015) Post-transcriptional regulation in corticogenesis: how RNA-binding proteins help build the brain. Wiley Interdiscip Rev RNA 6 (5):501–515. https://doi.org/10.1002/wrna.1289
94. Lehtinen MK, Walsh CA (2011) Neurogenesis at the brain-cerebrospinal fluid interface. Annu Rev Cell Dev Biol 27:653–679. https://doi.org/10.1146/annurev-cellbio-092910-154026
95. Loo L, Simon JM, Xing L, McCoy ES, Niehaus JK, Guo J et al (2019) Single-cell transcriptomic analysis of mouse neocortical development. Nat Commun 10 (1):134. https://doi.org/10.1038/s41467-018-08079-9
96. Hao H, Rong X, Jiangyang Z, Tianbo R, Richards LJ, Paul Y et al (2009) Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging. J Neurosci 29 (13):4263–4273. https://doi.org/10.1523/JNEUROSCI.2769-08.2009
97. Fukuchi-Shimogori T, Grove EA (2001) Neocortex patterning by the secreted signaling molecule FGF8. Science 294(5544):1071–1074. https://doi.org/10.1126/science.1064252



Enqiang Chang is a doctor in the Department of Anaesthesia and Perioperative Medicine at Henan Provincial People's Hospital. He graduated from Imperial College London. His main research is focused on mitochondria and neurobiology. He has published ten scientific publications.

Xiaoguo Ruan is a doctor in the Department of Anaesthesia and Perioperative Medicine at Henan Provincial People's Hospital. In addition to learning and memory, he has also been interested in pain and itch research. He mastered a number of basic behavioral and molecular biology experimental methods. His main research is focused on clinical bioinformatics, perioperative organ function protection, perioperative cognitive impairment, and chronic kidney disease. He has published more than ten scientific papers.

Ruilou Zhu is a researcher in the Department of Anesthesiology and Perioperative Medicine at Henan Provincial People's Hospital; Zhengzhou University People's Hospital; and Henan University People's Hospital. She graduated from Nanjing University. Her main research is focused on investigating the mechanism of postoperative cognitive dysfunction. She has published five scientific reports.

Yangyang Wang is a doctor in the Department of Anesthesia and Perioperative Medicine at Henan Provincial People's Hospital. His main research is focused on neurobiology. He has published two scientific publications.

Jiaqiang Zhang, Professor and Director of Department of Anaesthesiology and Perioperative Medicine, Henan Provincial People's Hospital, Member of Chinese medical association Anaesthesia Branch, National Committee of the Anaesthesia Branch of the Chinese Medical Association. Vice Chairman of the Anaesthesia Branch of the Henan Medical Association.

His main research is focused on investigating of the mechanism of postoperative cognitive dysfunction, clinical bioinformatics, Anesthesia and neurodevelopment, Physiology of pain. He is the author of more than 150 scientific publications.