



# Single-Cell Sequencing in Precision Medicine

9

Julia E. Wiedmeier, Pawan Noel, Wei Lin, Daniel D. Von Hoff and Haiyong Han

## Contents

9.1 Background .....	238
9.2 Single-Cell Sequencing Modalities .....	240
9.2.1 Sample Type and Preparation .....	240
9.2.2 Single-Cell DNA Sequencing .....	240
9.2.3 Single-Cell RNA Sequencing .....	243
9.2.4 Single-Cell Epigenetic Analysis .....	244
9.2.5 Other Single-Cell Analysis Techniques .....	245
9.3 Data Analysis .....	245
9.4 Application of Single-Cell Analysis in Precision Cancer Therapy .....	248
9.5 Perspectives .....	249
References .....	250

---

J. E. Wiedmeier · D. D. Von Hoff  
Mayo Clinic, Scottsdale, AZ, USA

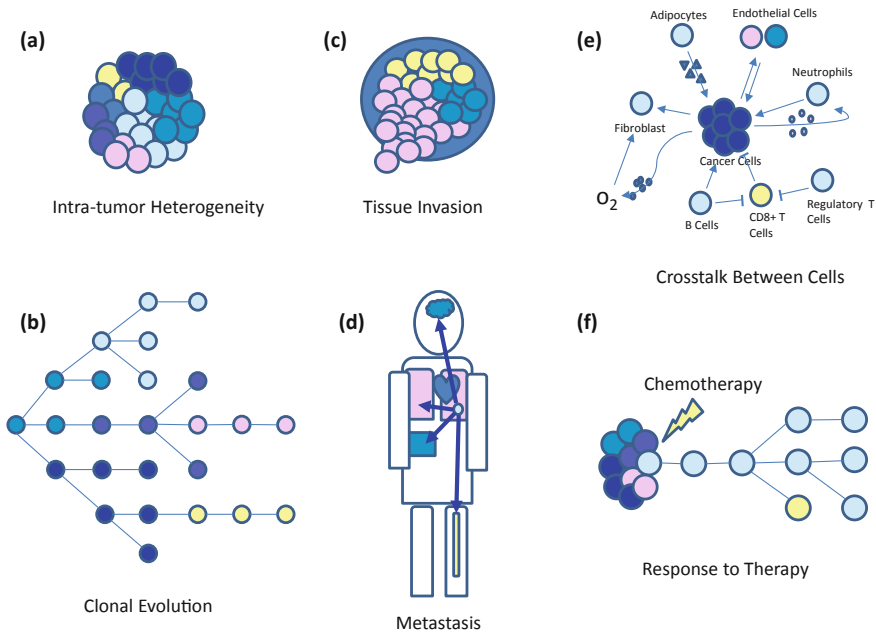
P. Noel · W. Lin · D. D. Von Hoff · H. Han (✉)  
Molecular Medicine Division, Translational Genomics Research Institute,  
Phoenix, AZ, USA  
e-mail: [hhan@tgen.org](mailto:hhan@tgen.org)

## 9.1 Background

The advent of next-generation sequencing (NGS) platforms made it possible to sequence DNA more efficiently and economically than Sanger sequencing. In addition, the application of NGS in cancer genomics allowed for a deeper understanding of the underlying genetics and pathogenesis of cancer. The human body is composed of trillions of cells that belong to approximately 200 different cell types; however, individual cells from defined cell types are diverse with unique expression profiles [1]. Many cell types/subtypes have few reliable markers that can be used for purification which is in part due to the fact that even cell types with well-established markers contain diversity [1]. Standard techniques for cancer analysis involve averaging signals from mixed populations of cells, which may mask or hide rare/small tumor clones and subclones that contribute to cell diversity [2]. However, with the use of single-cell genomics, the underlying genetics, expression levels, and epigenetics of every gene in the genome can now be analyzed across thousands of individual cells.

The first report of RNA transcriptome sequencing of single-cell mammalian cells occurred in 2009, and the first report of single human cancer cell DNA genome sequencing occurred in 2011 [2, 3]. Since that time, studies on single-cell exome and whole genome sequencing in varying types of cancers including renal, myeloproliferative, colon, lung, glioblastoma, breast, and prostate have been conducted [4–11]. The original single-cell sequencing method combined flow sorting, whole genome amplification, and NGS to generate genomewide datasets from single cancer cells; however, it had only ~10% physical coverage of a single cell's genome, sufficient for measuring large-scale copy number changes, but insufficient for resolving mutations at base-pair resolution [2]. Since that time, several other methods have been developed that can achieve high coverage (>90%) from single mammalian cells.

The data now available by single-cell sequencing has revolutionized cancer cell biology. The underlying mechanisms of how tumor and clonal diversity contribute to cancer biological processes remain largely unknown. Intratumor heterogeneity, clonal evolution, underlying mechanisms of tissue invasion, metastasis, and response to cancer-related therapies can potentially be elucidated by investigating molecular signatures at the single-cell level (Fig. 9.1). Tumor diversity is impacted by selection pressures, which can impact the underlying genetics of a cancer cell population. Examples of selection pressures include effects of the immune system, hypoxia, nutrient deprivation, geographical barriers, pH changes, and chemotherapy [2]. Understanding the underlying genetics of intratumor heterogeneity at the single-cell level has the potential to reveal cancer therapy resistance mechanisms that are lost at the bulk level. Single-cell tumor phylogenetic evolutionary trees have the potential to reveal driver mutations which can be used for targeted cancer therapies for those small populations of cells that harbor resistant mutations after treatment [12].



**Fig. 9.1** Applications of single-cell sequencing. Mechanisms behind **a** Intratumor heterogeneity, where different tumor cells show distinct genotypic and phenotypic variability (represented by different colored cells), **b** clonal evolution, or genetic diversification and clonal selection where different colored cells represent genetic changes, **c** tissue invasion, where mutant cells invade adjacent tissues (pink cells) with the potential to **d** travel to different sites, or metastasis, **e** crosstalk between cells with newer technologies, and **f** response to cancer-related therapies and subsequent clonal evolution can be elucidated by investigating the molecular signature at the single-cell level

Identification of cancer therapy resistance mechanisms at the single-cell level may also reveal novel mutations after induction of therapeutic agents (which generate clonal and subclonal populations of cells). In fact, NGS studies on single-cell mutations that drive tumorigenesis have revealed that resistance mutations vary from tumor to tumor [13]. Single-cell sequencing also has the potential to illuminate the mechanisms behind metastatic dissemination. For instance, several groups have used circulating tumor cells (CTCs) to study genomic and transcriptomic data from metastatic colon cancer, lung adenocarcinoma, and melanoma [2].

Bulk tumor gene expression studies are composites of transcriptional changes of heterogeneous cell populations; however, analysis of tumor cell expression at the single-cell level expands *average* tumor expression profiles of specific cell types, including non-malignant stromal, immune, and tissue-specific cells. In addition, single-cell transcriptomics aids in the detection of novel variants after treatments that may potentially drive drug resistance or serve as biomarkers of therapeutic success [14]. Furthermore, single-cell sequencing can detect low abundance of expression and/or novel RNA variants that are not detectable in bulk cell populations.

## 9.2 Single-Cell Sequencing Modalities

### 9.2.1 Sample Type and Preparation

Current single-cell technologies assay a single cell's gene expression, DNA variation, epigenetic state, and nuclear structure (see Table 9.1). In order to analyze genomics at the single-cell level, cells need to be isolated from extracellular matrix and cell–cell adhesion for downstream processes. One of the major limitations of single-cell sequencing from solid tissues includes unbiased disaggregating of the tissue into a suspension of single cells [15]. This is important as preferential processing or lysis of one cell type over another may skew data generation and thus analysis of results.

Single cells can be obtained from virtually any tissue, and current research focuses on circulating blood cancer cells, solid tumors, or circulating tumor cells (CTCs). Single cells from solid tumors can be obtained either following surgical removal, sampling/biopsy of the primary tumor or other organs with overt metastasis, or bone marrow aspiration. Solid tumors, especially in the invasive metastatic stages, are also known to shed cells into systemic circulation. These cells in the patient's blood stream are known as circulating tumor cells (CTCs), and those that disseminate to distant organs are termed disseminated tumor cells (DTCs).

Different methods for isolating single cells of interest from a suspension have been developed. Early studies used manual methods of cell isolation, using specialized pipettes or micromanipulation devices to isolate single cells [17]. This method of single-cell isolation has low throughput but can be used when a small number of cells are to be analyzed [22]. Other methods include fluorescence-activated cell sorting (FACS), magnetic-activated cell sorting (MACS), laser capture microdissection (LCM), and microfluidics, all of which can generally be used for larger numbers of cells. Microfluidics is a common method of single-cell isolation and allows for high throughput investigation of complex cellular systems using nanoliters of material. Microfluidics technologies isolate and encapsulate single cells in reaction chambers or droplets followed by standardized and automated nanoliter reactions, including barcoded sequence library prep for RNA and DNA sequencing. Commercially available microfluidic devices include Fluidigm C1 system, the 10× Genomics Chromium, and Illumina Biorad SureCell system [23].

### 9.2.2 Single-Cell DNA Sequencing

Whole genome amplification (WGA) followed by DNA sequencing identifies the underlying genetics and mutation frequencies of a single cell. Various methods for DNA whole genome amplification (WGA) at the single-cell level are available (multiple displacement amplification (MDA), PCR, or combination of both, see Table 9.1). These methods can now achieve >90% coverage of a single-cell

**Table 9.1** Comparison of single-cell sequencing modalities

Single-cell sequencing modality	Technique	Description	Strengths	Weaknesses	Reference
DNA	Nuc-Seq Single nucleus exome sequencing (SNES)	Combines flow sorting of single nuclei	High coverage data High detection efficiencies for single-nucleotide variants and indels Suitable for analyzing highly proliferative cells Reduced cost	Not suitable for analyzing normal or slow cell division Less genomic coverage than whole genome sequencing Time limited	Leung et al. [16]
	DNA seq Whole genome amplification approaches	Multiple genome amplification (MDA) Multiple-annealing, looping-based amplification-based cycle method (MALBAC)	High coverage Can generate genomic changes based on copy number alterations, structural variants, or single-nucleotide mutations	Artifacts Poor physical coverage High error rates	Van Loo and Voet [17] Trapnell [15] Navin [2]
RNA	RNA seq	MDA based or in vitro transcription based amplification of reverse transcribed single-cell mRNA (see above)	Most widely used and therefore many single-cell genomic assays available Inter- and intra-subclonal differences in expressed genes and expression level	Non-reliable amplification and deletion of transcripts expressed <10 copies/cell Only selectively amplify the polyA RNAs of a cell—bias in transcript coverage (3' UTR)	Van Loo and Voet [17] Trapnell [15] Navin [2]

(continued)

**Table 9.1** (continued)

Single-cell sequencing modality	Technique	Description	Strengths	Weaknesses	Reference
DNA/ RNA	Microfluidics: DropSeq (RNA) Compartmented droplet multiple displacement amplification (cd-MDA) (16)	Droplets containing individual cells with a barcoded RNA capture bead Bead contains cell identification and unique molecular identifier (UMI) Reverse transcription the bead oligonucleotides as templates Pool cDNA for library construction and subsequent sequencing	Can integrate labor-intensive experimental WGA processes in a single, closed device Improved reaction efficiency and detection sensitivity at the single-molecule level	Doublets can increase sequencing depth leads to plateau Expensive	Hosokawa et al. [18] Islam et al. [19]
Epigenetic	ATAC-seq	Uses engineered transposase to construct sequencing library from accessible DNA elements	Can take place within intact nuclei	Integration w/mutational and transcriptome analysis is challenging	Cusanovic et al. [20]
	Single-cell Hi-C	Active chromatin domains in cell nuclei localize to the surface of spatial chromosome territories Measures physical proximity between sites in the genome in three dimensions	Can take place within intact nuclei	Integration with mutational and transcriptome analysis is challenging	Cusanovic et al. [20] Buenrostro et al. [21]

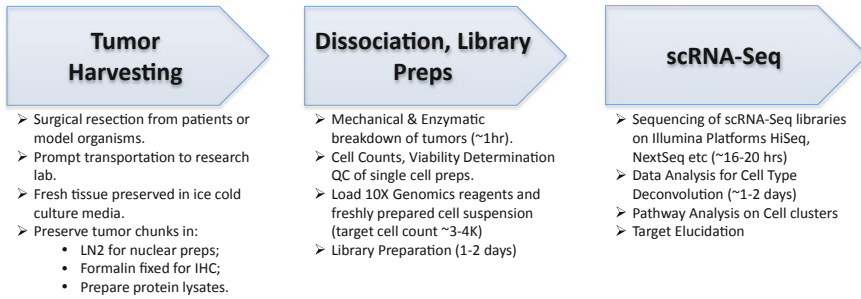
genome, and mutations can be detected at a single base-pair resolution [17, 2]. Technical challenges remain, however, including effective isolation and lysis of single cells, uniform amplification of whole genome, quality assessment of single-cell amplified genomes, sequencing library preparation, and data analysis [18].

Detection of point mutations or base substitutions in single cells has to be discriminated from polymerase base infidelities and sequencing errors. Such errors can occur during the WGA process, including allelic dropout (one allele is not amplified), transcripts can be over or under amplified, false-positive errors due to the infidelity of the DNA polymerase, and uneven amplification [2, 12]. By nature of the amplification process using DNA polymerases, errors that occur in the initial rounds of amplification are then inherited by all subsequent molecules [17]. Single-cell genomewide DNA sequencing is more challenging than single-cell transcriptomics due to the fact that there is simply less template available for single-cell genomics with DNA sequencing. Whereas single cells contain thousands of copies of each mRNA molecule, there are only two copies of each chromosome (or gene for that matter), and therefore only two template DNA molecules for WGA reactions.

### 9.2.3 Single-Cell RNA Sequencing

Single-cell RNA sequencing essentially reveals the transcriptional status at the single-cell level. Whole transcriptome sequencing, or RNA seq, where exclusively messenger RNA is assayed from single cells, is the most widely used method of single-cell analysis [24, 17, 25, 26]. It measures global gene expression by reverse transcription of mRNA into cDNA, and downstream sequencing libraries are made of hundreds to thousands of individual cells (Fig. 9.2). Gene expression is measured directly by counting the number of reads or the unique molecular index (UMI) that originate from each gene in a single cell. RNA seq of single cells achieves greater sequencing resolution than cell populations at the cost of less coverage [14]. The challenge with single-cell transcriptomics, as with most of the sequencing methods, is the “noise” generated from such experiments [27, 28, 12]. For example, biological variation is derived from genetic, epigenetic, environmental, and cellular factors. Technical noise can be introduced in the course of processing from sample handling, cell isolation, reverse transcription, cDNA amplification, sequencing, and analysis.

New techniques and methods are continuously been developed and reformed to limit the aforementioned challenges [29]. A crucial step of single-cell RNA seq is the unbiased amplification of cDNA before sequencing [27, 30, 28]. The use of unique molecular identifiers (UMIs), which bar codes each molecule, allows a robust quantification by intercepting amplification. Cell throughput is high, and the use of unique molecular identifiers to barcode individual transcripts also helps distinguish heterogeneous gene expression differences [19]. Microfluidics uses a process of capturing cells within nanofluidic chambers and has considerably



**Fig. 9.2** Workflow of the single-cell RNA sequencing process. Starting from harvesting tumors from patients or model organisms, fresh tissues are rapidly broken down into single cells, which are the input material for the 10× Genomics pipeline to generate sequencing libraries from single cells. Finally, the libraries are sequenced and data is analyzed to elucidate the cellular heterogeneity and biology of individual tumors

improved sensitivity for mutation detection by minimizing allelic dropouts [19, 12]. Drawbacks include reduced sensitivity such that only the 10–20% of the most abundant transcripts can be quantitated. In addition, reliable amplification and deletion of transcripts expressed at less than 10 copies per cell is a challenge and can lead to inaccurate quantification of low abundance transcripts [27, 28, 2]. Doublets, or cells that share the same UMI and bar codes, can also occur during sample processing; therefore, it is important to validate true single-cell capture before subsequent analysis.

### 9.2.4 Single-Cell Epigenetic Analysis

Whereas transcriptomics can be described as tracking “output” signals of a given genetic locus, such as a protein encoding gene, the analysis of chromatin, and epigenetic changes may be thought of as tracking the “input” signals of the locus in question (Wills 2015). With advances in technology, it is now easier to probe epigenetic phenomenon at the single-cell level including single-cell analysis of DNA accessibility, methylation status, histone modifications, and chromosome conformation by bisulfite sequencing, DNAase I hypersensitivity sequencing, ATAC-seq, and single-cell Hi-C, with the latter methods being the most developed for single-cell sequencing approaches [21, 31, 32, 33, 34, 35]. Challenges associated with these techniques are very similar to those of single-cell DNA sequencing.

Whole genome bisulfite sequencing assays identify DNA methylation (CpG islands). Chromatin immunoprecipitation sequencing (ChIP-seq) has yet to be adapted for single-cell sequencing; however, a method termed single-cell Hi-C analyzes active chromatin domains in cell nuclei [35]. This method measures proximity between sites in the genome in three dimensions, producing a “contact map” that can be used to identify looping interactions between regulatory elements



and gene loci [33, 35, 15]. Single cell assays for transposase-accessible chromatin with sequencing (ATAC-seq) is another method used to identify open chromatin regions which, in cancer cells, are often associated with oncogene expression and represent sites with increased vulnerability to mutagenic assault [20].

### 9.2.5 Other Single-Cell Analysis Techniques

One weakness of current single-cell techniques is that it requires analysis of single cells in suspension, which does not capture information pertaining to cell-to-cell interactions [2, 12]. Moreover, tumor clones evolve dynamically in space and time, and single-cell samples from an individual tumor may reveal mutations that are clonally dominant but may not be apparent in other regions of the tumor [2]. This has been partially overcome by spatiotemporal dynamics (discussed further below) which can be obtained by serial sampling of the same patient and provides information about evolution of a tumor through time (this is easier for liquid vs. solid tumors), multiple anatomically distinct biopsies for intratumor heterogeneity, in situ sequencing and imaging techniques for spatial resolution, or laser capture microdissection, which aids with information lost with cell-to-cell interactions [12].

Finally, new methods that integrate different single-cell genomic approaches and functional assays have recently been developed and include simultaneously measuring two or more modalities, whether it be genome and transcriptome, transcriptome and methylome, or RNA and protein [36, 17]. The ultimate goal of linking phenotypes of cells and their genotypes includes validation of gene expression and further development of precision medicine [36]. There are many different approaches available, some of which include gDNA-mRNA sequencing (DR-seq), genome and transcriptome sequencing (G&T-seq), single-cell DNA methylation analysis through bisulfite sequencing (scBS-seq) and reduced representation bisulfite sequencing (scRRBS-seq), as well as single-cell methylome and transcriptome sequencing (scTrio-seq) [37]. When extracting multiple “omic” datasets from individual cells, there are similar quality compromises as discussed previously [12].

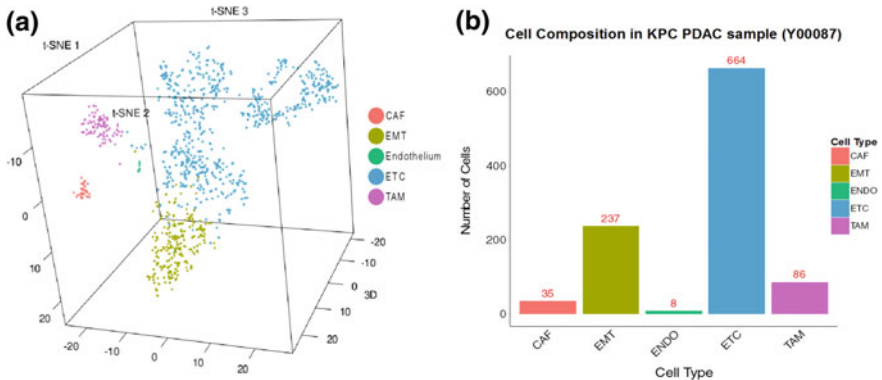
---

## 9.3 Data Analysis

Single-cell measurements preserve crucial information that is lost in bulk assays. Statistical and computational methods are critical to extract meaningful information from the data [36]. Single-cell analysis is based on the analysis of a cell modality (genetic variations, cell expression profile, changes in chromatin conformation, etc.), compared with some critical threshold. This threshold depends on the variability that exists in the assay as well as biological variability [15]. Statistical models can account for this variability.

Whole genome and whole transcriptome amplification as well as sequencing data are more difficult to analyze at the single-cell level than in bulk experiments [38, 39, 40, 1]. Bulk experiments have dozens of samples, and genome measurements cannot distinguish between fluctuations due to changes in gene regulation versus shifts in the ratio of different cell types. For single-cell genome amplification experiments, DNA is extracted from millions of cells, with intermixed sequences from different tumor clones, as well as normal cells [17]. Single-cell expression measurements can be variable; therefore, separating technical variability from biological variability is essential. Computational methods can help determine which mutations are clonal (present in all tumor cells) and which are subclonal. Point mutations and copy number data can be further analyzed with bioinformatics algorithms, and phylogenetic trees of different tumor subclones can be inferred. Analyzing expression levels in properly grouped subpopulations of cells allows a more accurate measurement of expression among different cell subpopulations [41, 15].

One of the major tasks of single-cell RNA seq data analysis is the resolution of cellular heterogeneity (Fig. 9.3). Most pathological samples such as tumor tissues consist of multiple cell types. Tumor tissues usually contain primary tumor cells, stromal cells, endothelial cells, and immune cells recruited from the peripheral blood and lymphatic organs. They are derived from different cellular lineages and play different roles in the tumor initiation, progression, and metastasis, which makes a tumor sample very complex and difficult to investigate. Before looking deeper into the tumor clonal variation, the tumor and its microenvironment should be resolved first. By identifying signature genes expressed in each cell type, one can further delineate the states of different cell types.



**Fig. 9.3** Single-cell RNA sequencing data analysis. **a** A representative example of a three-dimensional tSNE rendering of scRNA seq data for a mouse pancreatic tumor is shown. Each dot depicts a single cell, and colored clusters represent distinct cell types identified in the tumor. **b** Quantification of each individual cell type is shown with five different most commonly observed cell types identified in this tumor. CAF: cancer-associated fibroblasts; EMT: tumor cells undergoing epithelial to mesenchymal transition; ENDO: endothelial cells; ETC: epithelial tumor cells; TAM: tumor-associated macrophages

Understanding the cellular composition and tumor evolution can facilitate the evaluation of the tumor state and aid treatment decisions. Given the assumption that biological variability is larger than the technical variability, it is generally believed that similar cell types will cluster together by the cell type-specific feature expressions. The dynamic change of a specific cell type from one state to another sometimes could also be captured by such metric, though it is very challenging.

The calculation of the similarity between cells is usually based on multiple gene expression features. The relative position of these cells in a high-dimensional space is not easy to comprehend. Principal components analysis (PCA) is one of the most commonly used algorithms for reducing the dimensionality of data. Nonetheless, for single-cell RNA seq data, the first two to three components of PCA analysis cover a very small proportion of variance and therefore make the clustering effect less representative than for those other gene expression datasets [42]. A more advanced, nonlinear dimensionality reduction algorithm called t-Distributed Stochastic Neighbor Embedding, or t-SNE [42], has been proposed to explore high-dimensional single-cell RNA seq data. t-SNE itself does not necessarily define the similarity metric between cells but makes it visualizable in low-dimensional space. This algorithm is being widely used in part because it can be adapted to many visualization tools that are easily understood and interpreted by non-computational biologists (Fig. 9.3). The relationships across cell types can be inferred by the high-dimensional calculation and explored by the low-dimensional visualization.

The single-cell clustering and trajectory method of analysis can be used to define cell types and stable cell states. For example, clustering of three different cell types from one tumor by gene expression can group single cells into those of invasive tumor cells, noninvasive tumor cells, and stromal cells. When clustering, or grouping cells by gene expression, each cell represents a point in space based on expression of genes (approximately 30,000 genes). Clustering is based on the measured distances between points, and cells are grouped based on mutual proximity.

Besides resolving the cellular composition of heterogeneous tissues, single-cell studies also aim to characterize genes in such a way so that presumptions can be made of where a cell is in time and what drives a cell to transition from one state to another [15]. Using current methods, cells are analyzed at a particular state and time. The intact sample is destroyed by, for example, cell lysis (in the case of RNA seq and qRT-PCR) or cell fixation (in the case of FISH). Time series experiments extrapolate cell transition states through time; however, they misconstrue results by averaging cell expression as they proceed through a biological process in unsynchronized manner [43, 15]. Any particular sample at a given time-point contains cells of varying stages of cell growth and transition, reflecting the underlying dynamics of transitional state relevant genes. Single-cell analysis can define genes that are differentially expressed during these transitional states. In order to recover true signal of relevant expression, cells are re-ordered in something called “pseudotime” according to biological progress (for example, percent cells differentiated instead of time). Two algorithms are currently available based on pseudotime,

Wanderlust and Monocele, both of which attempt to define those genes responsible for cell transition [43, 15].

---

## 9.4 Application of Single-Cell Analysis in Precision Cancer Therapy

The translational application of single-cell sequencing in precision cancer therapy has the potential to improve cancer diagnostics, prognostics, targeted therapy, early detection, and noninvasive monitoring [2]. It is now technically and economically feasible to sequence single-cell DNA and RNA. Single-cell sequencing allows highly sensitive detection of rare mutations and cell-specific gene expression profiles. This method can identify rare tumor tissue variants that have the potential to drive drug resistance or serve as biomarkers of therapeutic success and ultimately advancing cancer genomics [14].

The importance of single-cell techniques in the clinical setting can be illustrated in tumor sampling. A single sample from a tumor does not represent the tumor as a whole. Spatially separate samples from a single tumor (or elsewhere in the body from metastasis) is composed of varying proportions of cell types and/or diverse underlying genetic and epigenetic makeup, otherwise known as tumor heterogeneity. Greater tumor heterogeneity may predict poorer response to therapy, higher probability of metastasis, or poor overall survival [2]. Identification of founder mutations, constructed from tumor phylogenetic trees, may aid in prediction of response to treatment.

Sequencing at the single-cell level can detect low abundance mutations, facilitating the identification of drivers of drug resistance. Drug resistance dynamics have been previously modeled in metastatic breast cancer cell line using RNA seq technology [14]. When metastatic breast cancer cells were treated with paclitaxel, stressed cells arrest and die, whereas those rare drug-tolerant cells resume proliferation and their clones expand. The ability to profile both the genome and transcriptome of the same cells has potential to elucidate heterogeneity at the genome, epigenome, transcriptome level.

Drug development is a lengthy and expensive endeavor with a high failure rate [10]. Drug development includes many steps: identification of drug targets, candidates, assessing drug resistance, drug toxicity, and pharmacokinetics. Many drugs emerge from preclinical studies only to fail in clinical trials. NGS has identified new target candidates for drug development. Single-cell sequencing in drug development expands on bulk genomic data by offering a more thorough and comprehensive picture on the underlying genetics, epigenetics, and transcriptomics of responders versus non-responders at an individual cell level. This ultimately allows for improved efficiency, accuracy, and identification. Applications of single-cell sequencing in drug development include identification of drug candidates and drug targets, drug resistance, and drug responses and toxicities [10].

Single-cell sequencing has shown potential to advance early detection and noninvasive monitoring. This concept is being elucidated by studies on circulating tumor cells (CTC), ultimately providing insight on metastatic dissemination (Navin 2015). CTC studies also aid in understanding evolution of the genome in early stages of cancer by identifying clones (and their underlying genetics, transcriptomics, etc.) that invade surrounding tissue.

Algorithms like Wanderlust and Monocle allows one to reconstruct transcriptional dynamics of development, differentiation, and/or clonal evolution from single-cell transcriptome data. Given such insight, we look forward to single-cell sequencing's ability to identify signature transcriptions of tumor states, which will strongly facilitate treatment decisions and healthcare strategies for patients.

---

## 9.5 Perspectives

Single-cell sequencing will transform cancer research over the coming years as even initial experiments have revolutionized our current understanding of gene regulation and disease. Indeed, the data available with single-cell techniques has never been possible before. Since the initial single-cell sequencing experiments, there have been many technical and experimental advances and the field continues to advance at a remarkable speed.

Drawbacks to single-cell sequencing include loss of tumor characteristics including spatial information, intratumor heterogeneity, and important cell-to-cell interactions. This issue stems from the fact that single-cell preparation and isolation capture techniques require intact single cells to be dissociated from fresh tissues. Most single cells are derived from a biopsy or small piece of tissue; therefore, single-cell sequencing may not accurately represent the underlying genome/transcriptome of the original tumor. Even the process of dissociating single cells from tissues may alter the cells and their underlying gene expression. In addition, microfluidic devices lose entire cell populations and may have bias for certain cells sizes, which, along with inherent weaknesses of selective amplification, can skew results.

With increasing amounts of complex data generated by single-cell sequencing techniques, there exists the dilemma of accurate interpretation and what to do with the sheer quantity of data generated. While many tools for analysis have been developed, there is a need for further analytic improvement in filtering noise and scalability [43]. Some of the above issues, particularly spatial information, can be overcome by single-cell analysis techniques, but this highlights the need for skilled bioinformaticians to accurately analyze the data. In addition, there is currently no universal analysis technique available, allowing for potentially more bias. Another challenge that limits the wide application of whole genome single-cell sequencing in the research and clinical settings is its relatively high cost compared to other bulk sequencing techniques. For example, a single-cell RNA seq experiment using the Chromium System from 10× Genomics currently costs at least 10 times higher than

a typical bulk whole genome RNA seq experiment. However, new and cheaper techniques are being developed which will greatly improve the accessibility of the technology so as time goes on.

Emerging new technologies that combine single-cell sequencing with other techniques acquire even deeper and richer genomic/biological information of cells and tissues. Spatial transcriptomics, which integrates single-cell RNA seq with the in situ hybridization (ISH), is one of such new technologies. It analyzes intact tissue sections on slides and does not require the need for cell isolation from tissue. The process involves the placement of histological sections on slides that contain reverse transcription primers with unique positional bar codes and subsequent placement of millions of oligonucleotides in micrometer subsections. This is followed by reverse transcription [23]. This method has the potential to overcome the loss of spatial information, intratumor heterogeneity, and the potential alteration of cells during the process of dissociation. Application of such technologies in clinical samples could potentially revolutionize patient care.

**Acknowledgements** We would like to thank the members of the Pancreatic Cancer Research Laboratory at the Translational Genomics Research Institute for their technical support and insightful discussions during the preparation of this manuscript. The work was supported in part by the Stand Up to Cancer—Cancer Research UK—Lustgarten Foundation (SU2C-CRUK-Lustgarten) Dream Team Translational Research Grant, a Program of the Entertainment Industry Foundation, and the National Foundation for Cancer Research (NFCR).

---

## References

1. Trapnell C, Cacchiarelli D, Grimsby J et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32:381. <https://doi.org/10.1038/nbt.2859>
2. Navin NE (2015) The first five years of single-cell cancer genomics and beyond. *Genome Res* 25(10):1499–507. <https://doi.org/10.1101/gr.191098.115>. PubMed MID: 26430160
3. Tang F, Barbacioru C, Wang Y et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6:377. <https://doi.org/10.1038/nmeth.1315>
4. Aceto N, Bardia A, Miyamoto DT et al (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158(5):1110–1122. <https://doi.org/10.1016/j.cell.2014.07.01>
5. Heitzer E, Auer M, Gasch C et al (2013) Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Res* 73(10):2965 LP–2975. <http://cancerres.aacrjournals.org/content/73/10/2965.abstract>
6. Lohr JG, Adalsteinsson VA, Cibulskis K et al (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* 32:479. <https://doi.org/10.1038/nbt.2892>
7. Ni X, Zhuo M, Su Z et al (2013) Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc Natl Acad Sci* 110(52):21083 LP–21088. <http://www.pnas.org/content/110/52/21083.abstract>
8. Patel AP, Tirosh I, Trombetta JJ et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* (80-), 344(6190):1396–1401
9. Wang Y, Waters J, Leung ML et al (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512:155. <https://doi.org/10.1038/nature13600>

10. Xu X, Hou Y, Yin X et al (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148(5):886–895. <https://doi.org/10.1016/j.cell.2012.02.025>
11. 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* (80-), 338(6114):1622 LP–1626. <http://science.sciencemag.org/content/338/6114/1622.abstract>
12. Wills QF, Mead AJ (2015) Application of single-cell genomics in cancer: promise and challenges. *Hum Mol Genet* 15:24(R1):R74–84. <https://doi.org/10.1093/hmg/ddv235>. PubMed PMID: 26113645
13. Kan Z, Jaiswal BS, Stinson J et al (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466:869. <https://doi.org/10.1038/nature09208>
14. Lee M-CW, Lopez-Diaz FJ, Khan SY et al (2014) Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc Natl Acad Sci* 111(44):E4726–E4735. <https://doi.org/10.1073/pnas.1404656111>
15. 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>
16. Leung ML, Wang Y, Waters J, Navin NE (2015) SNES: single nucleus exome sequencing. *Genome Biol* 16(1):55. <https://doi.org/10.1186/s13059-015-0616-2>
17. Van Loo P, Voet T (2014) Single cell analysis of cancer genomes. *Curr Opin Genet Dev* 24:82–91. <https://doi.org/10.1016/j.gde.2013.12.004>
18. Hosokawa M, Nishikawa Y, Kogawa M, Takeyama H (2017) Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics. *Sci Rep* 7(1):1–12. <https://doi.org/10.1038/s41598-017-05436-4>
19. Islam S, Zeisel A, Joost S et al (2013) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11:163. <https://doi.org/10.1038/nmeth.2772>
20. Cusanovich DA, Daza R, Adey A et al (2015) Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* (80-), 348(6237):910 LP–914. <http://science.sciencemag.org/content/348/6237/910.abstract>
21. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10:1213. <https://doi.org/10.1038/nmeth.2688>
22. 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>
23. Hedlund E, Deng Q (2018) Single-cell RNA sequencing: technical advancements and biological applications. *Mol Aspects Med* 59:36–46
24. Cloonan N, Forrest ARR, Kolle G et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5:613. <https://doi.org/10.1038/nmeth.1223>
25. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621. <https://doi.org/10.1038/nmeth.1226>
26. Nagalakshmi U, Wang Z, Waern K et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* (80-). 320(5881):1344 LP–1349. <http://science.sciencemag.org/content/320/5881/1344.abstract>
27. 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>
28. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16:133. <https://doi.org/10.1038/nrg3833>
29. 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>
30. Marr C, Zhou JX, Huang S (2016) Single-cell gene expression profiling and cell state dynamics: collecting data, correlating data points and connecting the dots. *bioRxiv*, 44743. <https://doi.org/10.1101/044743>

31. Buenrostro JD, Wu B, Litzenburger UM et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523:486. <https://doi.org/10.1038/nature14590>
32. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res* 23:2126–2135. <https://doi.org/10.1101/gr.161679.113>
33. Lieberman-Aiden E, van Berkum NL, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (80-), 326 (5950):289 LP–293. <http://science.sciencemag.org/content/326/5950/289.abstract>
34. Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133(3):523–536. <https://doi.org/10.1016/j.cell.2008.03.029>
35. Nagano T, Lubling Y, Stevens TJ et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502:59. <https://doi.org/10.1038/nature12593>
36. Linnarsson S, Teichmann SA (2016) Single-cell genomics: coming of age. *Genome Biol* 17 (1):16–18. <https://doi.org/10.1186/s13059-016-0960-x>
37. Macaulay IC, Ponting CP, Voet T (2017) Single-cell multiomics: multiple measurements from single cells. *Trends Genet* 33(2):155–168
38. Klein AM, Mazutis L, Akartuna I et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161(5):1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>
39. Macosko EZ, Basu A, Satija R 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>
40. Shalek AK, Satija R, Shuga J et al (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510:363. <https://doi.org/10.1038/nature13437>
41. Simpson EH (1951) The interpretation of interaction I contingency tables. *J R Stat Soc Series B Stat Methodol* 13:238–241
42. Hinton G (2008) Visualizing data using t-SNE visualizing data using t-SNE. *J Mach Learn Res* 9(2579–2605):85. <https://doi.org/10.1007/s10479-011-0841-3>
43. Svensson RR, Teichmann SA, Kar G (2017) Computational approaches for interpreting scRNA-seq data. *FEBS Lett* 591(15):2213–2225. <https://doi.org/10.1002/1873-3468.12684>