Review

Unmasking tumor heterogeneity and clonal evolution by single-cell analysis

Xiaoshan Shi, Papia Chakraborty, Amitabha Chaudhuri

MedGenome Inc., Foster City, California 94404, USA.

Correspondence to: Dr. Amitabha Chaudhuri, MedGenome Inc., Foster City, California 94404, USA.
E-mail: amitc@medgenome.com

How to cite this article: Shi X, Chakraborty P, Chaudhuri A. Unmasking tumor heterogeneity and clonal evolution by single-cell analysis. J Cancer Metastasis Treat 2018;4:47. http://dx.doi.org/10.20517/2394-4722.2018.32

Received: 24 May 2018  First Decision: 17 Jul 2018  Revised: 24 Jul 2018  Accepted: 9 Aug 2018  Published: 31 Aug 2018

Abstract

The intratumoral heterogeneity orchestrated by the tumor intrinsic and extrinsic mechanisms enable cancers to persist and spread notwithstanding the use of aggressive interventional therapies. The heterogeneity is revealed at multiple levels - at the level of individual tumor cells, in the cellular composition of tumor infiltrates and in the chemical microenvironment in which the cells reside. Deconvoluting the complex nature of the cell types present in the tumor, along with the homo and heterotypic interactions between different cell types can produce novel insights of biological and clinical relevance. However, most techniques analyze tumors at a gross level missing key inter-cell-type genotypic and phenotypic differences. The advent of single-cell sequencing has given an unprecedented opportunity to analyze the tumor at a resolution that not only captures the diversity of the cellular composition of a tumor but also provides information on the genetic, epigenetic and functional states of different cell types. In this review, we summarize the genesis of tumor heterogeneity, its impact on tumor growth and progression and their clinical consequences. We present an overview of the currently available platforms for isolation and sequencing of single tumor cells and provide evidence of its utility in precision medicine and personalized therapy.

Keywords: Intratumoral heterogeneity, single-cell sequencing, clonal evolution, circulating tumor cells, drug resistance

INTRODUCTION

A single cell is the ultimate denominator of a multicellular organism. In the progression of cancer, a single cell begins its journey to evolve into a malignant tumor cell and forms distinct subpopulations leading to intratumoral heterogeneity (ITH). Clonal diversity, the source of ITH, is the characteristics of all cancers and plays a critical role in cancer invasion, metastasis and development of resistance to targeted and non-
targeted therapies\cite{1-4}. Next-generation sequencing of bulk tumor tissues from many cancers has generated an unprecedented amount of multidimensional data bringing in novel insights into mechanisms of tumor initiation, progression and metastasis [Figure 1]. It has also unmasked the underlying deeper genotypic and phenotypic heterogeneity that exists between tumors belonging to the same cancer type. The ITH originating in the cancer genome can be revealed by deep exome and whole genome sequencing. However, transcriptome data from a complex mixture of cells derived from bulk tumor tissues fail to accurately elucidate the ITH, requiring technologies to study tumors at a single-cell resolution. Over the past ten years, there has been extraordinary progress in the development and application of single-cell analysis in cancer research as evidenced by the rise in publications describing different aspects of single-cell sequencing to characterize tumors at a deeper level [Figure 1]. In this review, we first introduce the concept of ITH and its clinical implications. Next, we outline new technologies enabling single-cell analysis with high sensitivity and finally provide examples of their applications in uncovering new perspectives in cancer diagnostics and treatment.

Figure 1. Application of whole exome sequencing (WES) and single-cell sequencing (sc-sequencing) to cancer research. A: Overview of patient cases to which WES and sc-sequencing were applied to characterize different types of human cancers to understand ITH and tumor microenvironment. The various types of cancers include liver cancer, lung cancer, renal cell cancer, blood cancer, brain cancer, breast cancer, pancreatic cancer, colorectal cancer and ovarian cancer compiled from public databases; B: the number of publications reporting applications of either whole exome or single-cell sequencing to cancer patients within the recent ten years. The key words “exome/single-cell sequencing” and “cancer patients” were used for searching articles from NCBI.
ORIGIN OF ITH

ITH was first described by Fidler et al.\cite{5} more than 30 years ago in murine models as a single tumor consisting of many cell subpopulations. However, this concept of heterogeneity in the composition of a tumor has now been expanded to include the genetic and molecular heterogeneity present within individual tumor cells and cells comprising the tumor microenvironment\cite{6-9}.

Genetic and epigenetic alterations

ITH arises as a result of both genetic and non-genetic changes in the tumor cells and the surrounding environment respectively [Figure 2]\cite{10}. Increased genetic instability as a result of mutations in DNA damage checkpoint control genes and DNA repair genes is one of the hallmarks of cancer and generates divergent clonal population of cells as the tumor grows over time\cite{11,12}. With the significantly high rate of cancer cell divisions, events of random mutagenesis increase, leading to local and global genetic alterations, that influence the future course of tumor development and progression\cite{13}. In addition, these genetic alterations create a hotbed for competition between clones driven by selection processes imposed by changes in the tumor microenvironment and by the use of therapies\cite{14,15}.

A vast majority of established driver mutations are clonal and arise early in the development of the tumor, however, subclonal de novo driver mutations may also arise in the later stages of tumorigenesis - to escape drug sensitivity and successful metastasis, for example\cite{16}. In a recent UK-wide multi-center prospective longitudinal cohort study, “Tracking Renal Cell Cancer Evolution through therapy (TRACERx Renal)”, clonal phylogeny and evolutionary subtypes were elucidated by multi-region sampling on matched primary and metastasis biopsies from 100 renal cell carcinoma patients\cite{17}. Subclonal driver mutations in the VHL and PBRM1 genes that were identified in the original tumor were absent in the widely disseminated metastatic tumor sites. Instead, these metastatic sites acquired loss of 9p and 14q mutations, suggesting that metastatic competence may not be driven by the founder driver mutations that established the primary tumor\cite{17}.

Tumor heterogeneity can also arise from epigenetic variations through DNA methylation that can profoundly modulate the open and closed conformation of chromatin in tumor cells, leading to gene expression alterations and phenotypic changes\cite{18}. For example, the methylation status of the tumor suppressor gene CDKN2B can be used as a biomarker of response to treatment in multiple diseases\cite{19}. However, heterogeneous methylation was observed in individual patients with acute myeloid leukemia, posing a challenge in using CDKN2B methylation as a biomarker\cite{18}. Similarly, differential microRNA expression is known to affect the diversity of cellular phenotype within a single tumor by modulating the expression of target genes\cite{21}. Subclonal expression of microRNAs (miRNA-21, miRNA-34a, miRNA-125, and miRNA-126) in prostate cancer is associated with diverse patient outcomes\cite{22}.

![Figure 2. Origin of ITH. Upon certain oncogenic hits, some cells in the normal tissues undergo genetic alterations to generate cancer cells. ITH arises through clonal evolution in which cells are dictated by transcriptomic and epigenetic factors and the tumor microenvironment. Cancer clones (yellow) propagate and generate successive clones (green) which outcompete the ancestral ones.](http://dx.doi.org/10.20517/2394-4722.2018.32)
Cellular composition of tumors

Cell types present in the tumor stroma, such as immune cells, fibroblasts, vascular cells play a critical role in shaping the composition of tumors by secreting cytokines growth factors and extracellular matrix that changes the stiffness of the tumor tissue. In a tumor microenvironment infiltrated by CD8 T cells at the tumor site is associated with increased overall survival, whereas myeloid-derived suppressor cells (MDSCs) possessing strong immune suppressive activity decreases overall survival. The diversity of these functionally different immune cell types creates a heterogeneous tumor microenvironment and regulate tumor growth, metastasis and treatment response. In addition, the distribution and density of the vasculature impact the supply of nutrients and oxygen selecting for tumor cells with specific metabolic phenotypes further contributing to tumor heterogeneity. Tumor heterogeneity has a significant bearing on the management of disease as summarized in the next section.

CLINICAL IMPACT OF THE ITH

Resistance to therapy

The resistance of tumors to therapies is often attributed to the presence of rare drug-resistant clones in the tumor before therapy or appears after treatment. An example of clonal resistance was observed in patients with anaplastic lymphoma kinase (ALK gene) rearranged non-small cell lung cancer (NSCLC) post treatment with ALK inhibitors. Patients that developed drug resistance displayed a distinct spectrum of ALK resistance mutations in response to different generations of ALK inhibitors. Particularly, ALK G1202R mutation is highly enriched in resistant tumors after treatment with second-generation ALK inhibitors, highlighting the significance of repeat biopsies and genotyping during the course of targeted therapy treatment.

In addition, studies investigating the mechanism of resistance of NSCLC tumors to EGFR tyrosine kinase inhibitors have revealed a variety of drug resistance mechanisms, including gatekeeper mutation T790M detected in > 50% of the EGFR TKI resistant tumors, amplification of MET receptor tyrosine kinase, activating mutation in PI3K pathway, and other uncharacterized mechanisms involving changes in the cellular phenotype. The appearance of a rare clonal population of tumor cells harboring drug resistance mutations or drug resistance phenotype can be captured by single-cell sequencing of the tumor and may not be discernible from whole tumor analysis, especially when present at a very low frequency. In an alternative model of drug resistance, resistant clones can be pre-existing in the tumor as a rare cell population and emerge post clearance of the drug-susceptible clones. In fact, in a study involving a cohort of 20 breast cancer patients, 8 out of 10 patients that did not show complete clearance of the tumor displayed unique somatic mutations in chemoresistant clones by single-cell sequencing. These mutations were pre-existing and were adaptively selected by the chemotherapy treatment. It is possible to detect de novo or drug-induced resistant clones present at low frequency by ultra-deep exome sequencing, however, two critical pieces of information - number of cells harboring the mutation and the zygosity of the mutation - cannot be accurately assessed from the bulk sequencing.

Challenges in diagnostic and prognostic biomarker identification

Identifying clinically relevant diagnostic biomarkers are challenging given that the tumor is heterogeneous and diagnostic or prognostic biomarkers are not expressed uniformly in all cells and across longitudinal assessment periods. For example, the divergent genetic landscape of metastatic cells can render biomarkers identified from primary tumors irrelevant.

In prostate cancer, ITH represents a major challenge for diagnostic and prognostic biomarker identification. Enhanced DNA ploidy and loss of PTEN, a tumor-suppressor gene, are critical prognostic markers of prostate cancer. In a clinical study of 304 patients who underwent radical prostatectomy, a significant difference in DNA ploidy classification and loss of PTEN expression was observed by analyzing all tumor areas in comparison to a single biopsy sample, suggesting that the heterogeneous chromosomal alterations com-
promise the accuracy of histopathology analysis and confound disease prognosis. Prognostic markers in ovarian cancer such as unique CpG methylation patterns have been suggested for progression-free survival as well as early disease recurrence following chemotherapy. However, DNA methylation patterns are heterogeneous and occurs in both large and poorly defined genomic regions, posing a challenge in using CpG methylation as a biomarker. In a recent study by Rajaram et al., a data-driven framework based on single-cell analysis has been reported that provides an estimate of the depth of sampling that may be minimally required to cover the full range of phenotypic heterogeneity for accurate biomarker discovery. Based on the analysis of 215 single-cell features, three replicates were sufficient to capture the heterogeneity for many features if they were defined by clear biomarkers without background noise. For example, nuclear staining (the number of nuclei staining by DAPI: an easily detectable feature) requires 1-2 cores to capture the heterogeneity in > 90% of the patients, while 10 cores or more are needed to assess the heterogeneity of YAP transcription factor expression (a sparsely detectable feature). Therefore, both the complexity of the feature and the biomarkers that define the feature determine the number of samples required for studying heterogeneity.

UNCOVERING ITH BY SINGLE-CELL ANALYSIS

Single-cell analysis is a powerful tool to resolve ITH of solid tumors and to detect the genetic makeup of rare cancer cells such as circulating tumor cells (CTCs) to ultimately guide personalized treatment strategies. The sensitivity of detecting somatic variants or changes in gene expression at a single-cell level has improved dramatically over the years through the introduction of new technologies. Single-cell analysis workflow includes isolation of single cells, either from the tumor site or circulating tumor cells from the blood. Following tumor dissociation, single-cells can be obtained by serial dilution, flow cytometry or microfluidics technology and then sequenced at sufficient depth to capture the genetic changes.
A major challenge in single-cell analysis is obtaining a viable cell sample from complex tumor tissues. Current methods include mechanical or enzymatic dissociation of tissues followed by isolation of single cells. Once the tissue is processed, multiple techniques to isolate single cells can be implemented. A more labor-intensive technique of laser capture microdissection (LCM) is also a viable approach for single-cell isolation from sectioned tumor samples. One challenge for single-cell transcriptomics is the poor RNA quality extracted from archival tumor samples such as formalin-fixed paraffin-embedded (FFPE) samples. However, with the Smart-3SEQ method, it is now feasible to perform single-cell RNA-seq on FFPE samples. Additionally, recent advances using the SMART seq technology and cDNA synthesis methods using...
random priming (SMART-Seq Stranded Kit, Takara Inc.) have been beneficial in extracting reliable gene expression information from poor quality RNA from FFPE samples.

**Single-cell isolation by mechanical or enzymatic dissociation**

Conventionally, tumor tissues are dissociated into single cells by mechanical dissociation (e.g., meshing, trituration with a pipette/tip)\(^43,44\) or by enzymatic dissociation\(^45,46\) or a combination of both. Enzymes such as collagenase\(^41\), DNase\(^46\), trypsin\(^47\) are commonly used for dissociating the cell-cell contacts and the extracellular matrix to generate single cell suspensions. The various dissociation methods may largely differ in their yield of viable cells\(^48,49\), limiting their downstream applications. Therefore, tumor dissociation protocols optimized for different tumor types is a key gap that needs to be addressed for high-throughput single-cell analysis.

**Single-cell isolation by LCM**

To preserve the native properties of tumor cells shaped by the complex tumor microenvironment, LCM can be used to isolate tumor cells directly from sectioned tissues. It is a method to procure subpopulations of tissue cells under direct microscopic visualization by cutting away unwanted cells and obtain histologically pure cell population [Figure 4A]\(^50\). A variety of downstream applications exist for microdissected cells such as DNA genotyping, RNA transcript profiling or cDNA library generation. Even though the majority of the studies take advantage of approximately 100-1000 dissected cells, LCM can also be used for single-cell isolation directly\(^51-53\).

**Isolation of rare CTCs**

Currently, tumor biopsies are obtained to establish the diagnosis and determine whether the predictive biomarkers are consistent between the primary and the metastatic tumors. However, getting biopsies is invasive, expensive and not always feasible. Additionally, it is difficult to get biopsies of metastatic lesions or get repeat biopsies for difficult to access tumors. Analysis of disseminated tumor cells (DTCs) is a useful alternative to tumor biopsy in clinical setting for patient stratification, therapy selection and monitoring drug resistance during the course of treatment\(^54\). DTCs originate from the primary or metastatic tumors, extravasate into the bloodstream or lymphatics and carry genomic profiles of tumors from which they originate\(^55,56\). Disseminated cancer cells are usually detectable as CTCs in the circulation\(^54\). A small fraction of them that have reached to a secondary organ such as the bone marrow and lymph nodes is termed as DTCs\(^54\). Though for certain cancers, the presence of DTCs in distant organs is a strong predictive marker for cancer metastasis, the challenge with DTC isolation due to the invasive procedure is a deterrent in studying this population by single-cell sequencing. On the contrary, CTCs circulating in patient blood has proved to be a valuable resource for diagnostic and prognostic biomarker discovery\(^57\), although distinguishing a DTC from a pool of CTCs is challenging.

CTCs contain signatures of tumor heterogeneity and carry the spectrum of somatic mutations present in both the primary and metastatic lesions in different cancers\(^55,56,58\). Because conventional molecular analysis of whole tumors provides genotype/phenotype information of the dominant clones or aggregated information of all clones, single-cell analysis of the CTCs is a potential solution to investigate heterogeneity. By isolating and sequencing single CTCs in the blood, it is possible to measure somatic mutations that are present at both the primary and metastatic tumor sites without performing an invasive core biopsy\(^59,60\). Two types of isolation methods - microfluidic-based and immunoaffinity-based are used for capturing CTCs.

**Microfluidic-based cell isolation**

The microfluidic platform can be used for single-step isolation of CTCs from unprocessed blood specimens\(^61,62\). As whole blood flows through the CTC-chip, individual CTCs are captured onto the microposts coated with anti-EpCAM antibody. This type of microfluidic processing enables high yield of pure CTCs\(^63\). Subsequent studies demonstrated the ability and reliability to isolate CTCs from patients with metastatic
lungs cancer using this CTC-chip to perform an EGFR mutational analysis\textsuperscript{[63]}. An improved microfluidic CTC isolation platform, the herringbone (HB)-chip, is also developed by the same group\textsuperscript{[64]}. The HB-chip uses calibrated microfluidic flow patterns to drive cells to come in contact with the antibody-coated walls of the device, thereby reducing cell collisions and improving target cell capture efficiency. A commercial microfluidic circuitry chip DEPArray System (Menarini Silicon Biosystems, Inc.) containing an array of individually controllable electrodes to create a dielectrophoretic (DEP) cage around each cell for single CTC isolation is also available\textsuperscript{[65]}. Besides isolation of CTCs from blood, the microfluidic platform can also be used for single-cell isolation from other tissues\textsuperscript{[66,67]}. For example, an innovative workflow using DEPArray system was established to examine tumor heterogeneity using FFPE samples, providing a solution for genetic analysis using minute archival clinical samples\textsuperscript{[68]}.

**Immunoaffinity-based cell isolation**

The CellSearch Circulating Tumor Cell Kit (Menarini Silicon Biosystems, Inc.) is based on ferrofluid- and fluorochrome-couple antibodies with high binding affinities for the EpCAM antigen of CTCs. After immunomagnetic capture and enrichment, CTCs in peripheral blood are detected and enumerated as measured by fluorescence intensity. ITH has been reported for PIK3CA and TP53 mutations in metastatic breast cancer using a combination of CellSearch and DEPArray technologies\textsuperscript{[69,70]}. CTCs can also be purified and enriched using an immunomagnetic enrichment device termed MagSweeper\textsuperscript{[71]}. Using this technique, high level of heterogeneity among individual CTCs was detected in the blood of metastatic breast cancer patients\textsuperscript{[72]}.

**Isolation of single cells using Fluorescence-activated cell sorting**

Flow cytometry using fluorescence-activated cell sorting is a powerful method of isolating single cells that share the same marker from liquid suspensions. Cells passing through the lasers emit optical signals enabling their separation and capture from other cells that lack the signal\textsuperscript{[73,74]}. Single cells can be sorted individually onto a 96 well plate format [Figure 4B]. Alternatively, a serial dilution can be performed using the sorted cell suspensions into a 96 well plate such that each well contains a single cell. Downstream sequencing can be performed using a 96 well plate format.

Isolated single cells can be interrogated by a variety of genomic technologies for deeper genotype-phenotype characterization. Significant technological advancement summarized in the next section is producing novel insights into the biology of the disease and applications in the clinic.

**Downstream analysis of single cells**

**Single-cell genomics**

The work-flow of single-cell sequencing involves amplification of genomic DNA or RNA transcripts to produce enough material for library construction. The earliest method of sequencing DNA from single-cells combined flow-sorting cells by DNA ploidy followed by single-nucleus sequencing by degenerative-oligonucleotide-PCR technique\textsuperscript{[74,75]}. However, this method failed to generate genome-wide single nucleotide variants due to low coverage of ~6\%\textsuperscript{[74,75]}. A non-PCR-based multiple-displacement DNA amplification method using Phi29 enzyme and random hexamers [Table 1] produced good genome coverage with high sequence fidelity in multiple single-cell studies\textsuperscript{[76-79]}. Another amplification method - multiple annealing and looping-based amplification cycles (MALBAC) reduced whole-genome amplification bias and improved genome coverage [Table 1]. In the MALBAC method, limited isothermal amplification using degenerate primers, followed by PCR amplification produced 93% genome coverage for a single cell and both copy-number variations and single nucleotide variations were detected\textsuperscript{[80]}. Amplification bias is a serious limitation in single-cell sequencing, which can reduce the accuracy of genomic information from single-cell genomes\textsuperscript{[81]}. Statistical models have been developed to calibrate allelic bias in single-cell whole-genome amplification to reduce the sequencing artifacts\textsuperscript{[81]}.
The first study of single-cell RNA transcriptome of mouse blastomere detected novel splice junctions and expression of more genes than previous microarray studies. However, this method was found to have a strong 3' bias due to the inefficiency of first-strand cDNA synthesis by reverse transcriptase. To overcome this problem, Smart-seq technique was developed using MMLV reverse transcriptase with template-switching activity. This Smart-seq method utilizes an intrinsic property of MMLV to add three to four cytosines specifically to the 3' end of the first cDNA strand, which is subsequently used to anchor a universal PCR primer for amplification. In a single-cell RNA-seq of CTCs from melanoma patients, Smart-seq has improved read coverage across transcripts despite increased noise in gene expression estimates. Moreover, distinct gene expression patterns including candidate biomarkers for melanoma CTCs were reported in this study.

**MDA:** multiple-displacement DNA amplification; **DOP-PCR:** degenerative-oligonucleotide-PCR; **IVT:** *in vitro* transcription

### Table 1. Techniques for single-cell analysis

<table>
<thead>
<tr>
<th>Methods</th>
<th>Example</th>
<th>Advantage</th>
<th>Disadvantage</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genome</strong></td>
<td>DOP-PCR</td>
<td>High-throughput, high coverage</td>
<td>Amplification bias, allelic dropout</td>
<td>[74,75]</td>
</tr>
<tr>
<td>MDA</td>
<td></td>
<td>High-throughput, even coverage</td>
<td>Amplification bias, allelic dropout</td>
<td>[58,76-79]</td>
</tr>
<tr>
<td>MALBAC</td>
<td></td>
<td>High-throughput, even coverage</td>
<td>Amplification bias, allelic dropout</td>
<td>[80]</td>
</tr>
<tr>
<td><strong>Transcriptome</strong></td>
<td>MMLV</td>
<td>Smart-seq</td>
<td>Full-length transcript, amplify quickly</td>
<td>[83,84]</td>
</tr>
<tr>
<td>IVT</td>
<td>CEL-Seq</td>
<td>Full-length transcript, specificity, ratio fidelity</td>
<td>3' bias, low efficiency</td>
<td>[87,88]</td>
</tr>
<tr>
<td>Phi29 DNA polymerase</td>
<td>Full-length transcript, high efficiency, low bias</td>
<td>No strand specificity</td>
<td>[89,90]</td>
<td></td>
</tr>
</tbody>
</table>

**In vitro** transcription (IVT) -based linear RNA amplification uses T7 RNA polymerase to produce transcripts with high specificity and low error rate [Table 1], it has the drawback of lower efficiency and is biased towards the 3’ end of input transcripts. CEL-Seq method of pooling cells and libraries reduced some of the limitations of IVT and was used to capture differential gene expression in two-cell stage embryo of *C. elegans*. The third strategy used Phi29 DNA polymerase for cDNA library generation from single cells [Table 1]. RNA is reverse transcribed, circularized and then amplified using Phi29 polymerase which preserves full-length transcript coverage. Additionally, random primers can be incorporated to generate cDNA, making this method suitable for prokaryotes.

### A combined method of single-cell isolation and single-cell sequencing

Microfluidic devices for single-cell isolation coupled with single-cell RT-qPCR or whole transcriptome has been developed by multiple groups. A good example is a microfluidic device developed by White et al. capable of performing high precision RT-qPCR measurements of gene expression from hundreds of single cells per run. This device combines cell loading, cell lysis, reverse transcription and quantitative PCR in one cell processing unit [Figure 4Ci]. Once cells are loaded, a single cell is trapped in a cell capture chamber. After cell lysis, the transcript target is reverse transcribed before being injected into the PCR chamber. Master mixes for RT and qPCR are loaded onto the common feed channel sequentially to enable each reaction step. A similar device, featuring additional cell processing chambers and sample elution capabilities has been released as a commercial product (Fluidigm C1) in 2012. Since then, an increasing number of studies investigated ITH using Fluidigm’s microfluidic device.

Efforts to reduce amplification bias by incorporating unique molecular identifiers before transcriptome am-
plification are ongoing. A novel technique termed Drop-seq uses the microfluidic chamber to isolate single cells followed by labeling RNA of individual cells with a different barcode, allowing pooling of cDNA during sequencing thereby greatly improving the multiplexing efficiency. Applying Drop-seq to mouse retinal bipolar cells resulted in the identification of different types of neurons by matching molecular expression to cell morphology. A similar technique was commercialized by 10x Genomics Inc [Figure 4Cii] in 2016.

The 10x platform applies unique barcodes to separately index each cell by partitioning thousands of cells into Gel Bead-in-Emulsions. Libraries are generated and sequenced and the 10x barcodes are used to associate individual reads back to the individual cells. The platform can profile up to 10,000 cells from a complex mixture of different cell types.

**APPLICATIONS OF SINGLE-CELL SEQUENCING**

Recent technical advances have enabled generation of unprecedented amount of information on genomics and transcriptomics at the single-cell level [Table 2]. Compared to bulk transcriptomics data obtained from tumor tissues, single-cell RNA-seq allows capturing of the gene expression profile from individual cells of heterogenous origin, which is a significant advantage over bulk sequencing that captures the average gene expression of a sample. Secondly, for the samples with limited amount of material, single-cell analysis is a good alternative to characterize the genotype. Taking CTCs for an example, mutations identified in CTCs

<table>
<thead>
<tr>
<th>Tumor type</th>
<th>Sample type</th>
<th>Method</th>
<th>Description</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td>CTC</td>
<td>DNA-seq</td>
<td>Mutation profiling, clonal evolution</td>
<td>[55]</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>CTC</td>
<td>DNA-seq</td>
<td>Genetic lineage</td>
<td>[58]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>CTC</td>
<td>RNA-seq</td>
<td>Transcriptome profiling</td>
<td>[72]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Clonal diversity</td>
<td>[75]</td>
</tr>
<tr>
<td>Melanoma</td>
<td>CTC</td>
<td>RNA-seq</td>
<td>Transcriptome profiling</td>
<td>[83]</td>
</tr>
<tr>
<td>Leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Mutation profiling, clonal evolution</td>
<td>[97]</td>
</tr>
<tr>
<td>Glioblastoma multiforme</td>
<td>Primary tumor</td>
<td>RNA-seq</td>
<td>Clonal evolution</td>
<td>[106]</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Mutation profiling, clonal evolution</td>
<td>[105]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Copy number evolution, clonal evolution</td>
<td>[74]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Copy number evolution, clonal evolution</td>
<td>[77]</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Clonal evolution</td>
<td>[109]</td>
</tr>
<tr>
<td>Kidney cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Mutation profiling</td>
<td>[76]</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Mutation profiling, clonal evolution</td>
<td>[110]</td>
</tr>
<tr>
<td>Colon cancer</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Clonal evolution</td>
<td>[111]</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Clonal evolution</td>
<td>[112]</td>
</tr>
<tr>
<td>Chronic lymphocytic leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Genotype-phenotype relationship, clonal evolution, mutation profiling</td>
<td>[113]</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>CTC</td>
<td>DNA-seq</td>
<td>Copy number evolution</td>
<td>[56]</td>
</tr>
<tr>
<td>Pancreatic ductal adenocarcinoma</td>
<td>CTC</td>
<td>RNA-seq</td>
<td>Phenotype characterization</td>
<td>[115]</td>
</tr>
<tr>
<td>Glioblastoma</td>
<td>Primary tumor</td>
<td>RNA-seq</td>
<td>Transcriptional profiling, phenotype characterization</td>
<td>[43]</td>
</tr>
<tr>
<td>Glioblastoma</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>EGFR evolution</td>
<td>[116]</td>
</tr>
<tr>
<td>B cell leukemia</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Karyotype heterogeneity</td>
<td>[117]</td>
</tr>
<tr>
<td>Myeloproliferative neoplasm</td>
<td>Primary tumor</td>
<td>DNA-seq</td>
<td>Mutation profiling, clonal evolution</td>
<td>[78]</td>
</tr>
<tr>
<td>Melanoma</td>
<td>CTC</td>
<td>DNA-seq</td>
<td>Mutation profiling, copy number evolution</td>
<td>[118]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>CTC</td>
<td>RNA-seq</td>
<td>Transcriptome profiling</td>
<td>[120]</td>
</tr>
<tr>
<td>Various cancers</td>
<td>Primary tumor</td>
<td>RNA-seq</td>
<td>TCR repertoire analysis</td>
<td>[124,126]</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>Primary tumor</td>
<td>RNA-seq</td>
<td>Characterization of T cell functional states</td>
<td>[130]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>RNA-seq</td>
<td>Tumor microenvironment characterization</td>
<td>[132]</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>CTC</td>
<td>RNA-seq</td>
<td>Heterogeneity in signaling pathways</td>
<td>[136]</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>CTC</td>
<td>DNA-seq</td>
<td>Copy number evolution</td>
<td>[137]</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>DNA-seq, RNA-seq</td>
<td>Clonal evolution, transcriptome profiling</td>
<td>[32]</td>
</tr>
</tbody>
</table>

ITH: intratumoral heterogeneity; CTC: circulating tumor cell
are also present in the primary tumor and may be found in the metastatic lesions\textsuperscript{[15]}, suggesting that single-cell analysis on CTCs is an effective option to non-invasively monitor cancer progression and predict metastatic risk. Last but not the least, single-cell analysis facilitates researchers to dissect tumor heterogeneity at a much higher resolution than before. For example, the degree of karyotypic anomalies in human cancer is associated with tumor progression and therapeutic response to cancer treatment\textsuperscript{[102]}. However, current karyotypic analysis methods rely on a small fraction of dividing mitotic subpopulations in the sample and do not provide in-depth information on copy number variations (CNV)\textsuperscript{[102,103]}. Single-cell whole genome sequencing offers a significant advantage over traditional methods in analyzing karyotypic anomalies and CNVs at a much higher resolution.

**Understanding tumor evolution**

Tumor evolution is a dynamic process and describes the emergence of cancer cell subpopulations under environmental pressure. As the tumor grows, each generation of cells acquire novel somatic mutations that provide cells with survival advantages thereby determining the overall fitness of the clonal population\textsuperscript{[104]}. Waves of clonal expansion and contraction driven by changes in the tumor microenvironment govern the life cycle of a tumor. Single-cell sequencing can potentially identify low abundance clones carrying driver mutations, which can be further leveraged to refine therapeutic strategies. Although low abundance driver mutations are possible to detect by deep exome sequencing, the fraction of cells carrying the mutation, or the zygosity of the change (relevant for loss of function mutations in tumor suppressor genes) are hard to estimate without single cell sequencing. A computational approach to map single-cell mutational profile from exome sequencing was successfully used to chart the chronological acquisition of mutations and create a phylogenetic map of tumor evolution in both glioblastoma multiforme and secondary acute myeloid leukemia (AML)\textsuperscript{[105,106]}. A similar analysis in breast cancer identified three clonal populations in the primary tumor of which only one clone was present in the metastatic lesion\textsuperscript{[74]}. This observation supports the hypothesis that rare clones present in the primary tumor harbor genetic signatures of metastasis even before they have spread and colonized distant sites\textsuperscript{[74,107,108]}. In a follow-up breast cancer study, aneuploidy rearrangements were shown to occur early in tumor evolution, which remained highly stable as the tumor grew, whereas, point mutations generated clonal diversity\textsuperscript{[77]}. A similar pattern is observed in lymphoblastic leukemia patients where recurrent translocations appear earlier than structural nucleotide variants\textsuperscript{[109]}. This suggests that large structural alterations offer selective advantage early during tumor growth followed by accumulation of mutations producing clonal diversity. This is supported by the finding that subclonal populations arise more frequently in tumors with high mutational burdens such as bladder and colon cancer, but not in tumors with low mutational burden such as renal cell carcinoma\textsuperscript{[76,110,111]}. A clonal progression of multiple mutations was mapped in hematopoietic stem cells of AML patients, suggesting the clonal evolution of AML genomes from founder mutations\textsuperscript{[112]}. An interesting finding from single-cell analysis is that phenotypic diversity fails to recapitulate genotypic diversity detected in subclones strongly implicating that a large proportion of genotypic variation may lack functional consequences, appearing and disappearing without contributing to tumor evolution\textsuperscript{[113]}.

**Disease diagnosis and therapeutic stratification of patients**

Modern cancer treatment relies heavily on accurate molecular and immuno/histopathological tissue analysis of needle biopsies or surgically resected tissues for diagnosis. Tumor heterogeneity often confounds accuracy of disease diagnostics by subsampling a subset of tumor cells that may not represent the whole tumor. This calls for obtaining multiregional and longitudinal samples to guide therapeutic intervention, which is often not routine. High-resolution single-cell analysis of tumor samples or CTCs can aid in refining diagnostic parameters and patient stratification.

In a single-cell sequencing study of CTCs from metastatic lung cancer, patients who share the same subtype of lung cancer displayed similar patterns of copy number variations in their CTCs, providing a potential
biomarker of CTC-based cancer diagnostics\textsuperscript{[56]}. In pancreatic cancer, pancreas epithelial cells can be present in the blood at pre-cancerous stages in pancreatic ductal adenocarcinoma patients\textsuperscript{[114]}. In another study, single-cell sequencing analysis on CTCs obtained from pancreatic ductal adenocarcinoma patients identified a macrophage-pancreatic tumor cell fusion product with high proliferative and metastatic potential\textsuperscript{[115]}. These studies suggest that early detection of these pancreatic epithelial cells in the bloodstream can serve as an important diagnostic tool for pancreatic cancer detection\textsuperscript{[114,115]}. The treatment of glioblastomas, an aggressive type of brain tumor has benefited from single-cell sequencing because of a high degree of tumor heterogeneity harboring a diverse population of cells with a large spectrum of stemness, differentiation states, and variable proliferative capacity\textsuperscript{[43]}. By applying single-cell sequencing to EGFR-amplified glioblastomas, novel EGFR truncation variants were identified\textsuperscript{[116]}. \textit{In vitro} and \textit{in vivo} functional studies revealed that a specific EGFR variant (EGFRvII, deletion of exons 14 and 15) was sensitive to EGFR inhibitors, which are currently in clinical trials\textsuperscript{[116]}. In chromosomally unstable B cell leukemia patients, different degrees of karyotypic abnormalities were detected by single-cell whole genome sequencing, which bulk sequencing failed to detect. Because karyotypic abnormalities associate with poor clinical outcome in multiple cancers\textsuperscript{[102]}, the degree of karyotypic anomalies assessed by single-cell sequencing can be utilized as an important readout for stratifying patient risk\textsuperscript{[117]}. Single-cell analysis has identified novel mutations in JAK2-negative myeloproliferative neoplasm such as SESN2 and \textit{NTRK1}, chronic lymphocytic leukemia such as \textit{LCP1} and \textit{WNK1} and chromosomal abnormalities in melanoma such as chromosomal 12 amplification\textsuperscript{[78,113,118]}, opening up opportunities to target these neoplasms. For example, \textit{NTRK1} encodes a tyrosine kinase receptor and inhibitors are available to target its \textit{NTRK1} gene fusions that results in constitutive activation of the kinase\textsuperscript{[116]}. For patients who are JAK2 mutation negative but harbor \textit{NTRK1} mutation, it is tempting to speculate that \textit{NTRK1} can be a target for the treatment of myeloproliferative neoplasm.

\textbf{Disease monitoring and prognostic biomarkers}

Cancer heterogeneity in part is driven by selection pressure that arises during drug treatment. Capturing this dynamic heterogeneity at the genetic and cellular composition level prior to, during, or post-treatment is crucial in assessing drug efficacy and predicting patient survival. Single-cell analysis is an extremely powerful tool to capture the dynamic events at a molecular level for disease monitoring and in predicting prognostic biomarkers. Below are few examples of the application of single-cell sequencing in developing prognostic and predictive biomarkers.

\textit{CTC analysis}

Single-cell analysis of CTCs can provide prognostic markers in several cancers. Microfluidics-based RNA sequencing has aided identification of CTC clusters held together by the cell junction component plakoglobin that mediate intercellular adhesion. Presence of high levels of CTC clusters over single CTCs correlated with poor prognosis indicating their role in the metastatic spread of cancer\textsuperscript{[120]}. Indeed, heterogenous expression of plakoglobin in the primary tumor supports the evidence that tightly adhered groups of cells from the primary tumors serve as the precursors to CTC clusters in circulation. Thus, single-cell identification of plakoglobin-positive clonal cell populations of tumor cells in conjunction with the presence of CTC clusters in the patient blood is a potent prognostic marker of breast cancer metastasis\textsuperscript{[120]}.

\textit{TCR repertoire analysis}

Anti-tumor immunity is largely driven by antigen-specific CD8 T cells, which recognize tumor-derived neoantigenic peptides complexed with human leukocyte antigen also referred to as major histocompatibility complex (MHC) in mouse, to mount an anti-tumor immune response\textsuperscript{[121]}. Adoptive cell therapy using autologous tumor infiltrating lymphocytes (TILs) has been shown to be effective for the treatment of multiple cancers\textsuperscript{[122,123]}. The anti-tumor effects observed post T cell therapy are associated with the activation of
To improve the efficacy of the T cell therapy, engineering TILs to express the neoantigen-specific TCR can be a promising next-generation immunotherapy drug. However, to develop these engineered T cells, identifying paired sequences of both TCR α and β chains from the vast repertoire of TCRs is a challenge. One way to overcome this challenge is to perform single-cell TCR profiling to obtain paired TCR α/β sequence information. Using patient samples, neoantigen specific CD8 T cells were clonally expanded in vitro and multiple paired TCR sequences were identified by single-cell analysis. Importantly, the transduced T cells expressing TCRs recognized the neoantigen presented by autologous antigen-presenting cells. Another study using single-cell TCR repertoire analysis revealed that clonally expanded CD8 T cells were antigen-specific and showed cytotoxic activity against tumors in mouse models. Intriguingly, the combination of 10x Genomics’ single cell TCR sequencing platform coupled to gene expression holds enormous potential for assessing and monitoring patient response to cancer vaccines and immunotherapy drugs.

**Monitoring the functional state of CD8 T cells**

In the tumor microenvironment, the ability of CD8 T cells to secrete pro-inflammatory cytokines and exert cytotoxic function can be compromised during persistent immune activation. Such exhausted CD8 T cells differ profoundly from memory CD8 T cells and co-express multiple co-inhibitory immune checkpoint regulators such as PD-1, LAG-3, and TIM-3 and lack successful anti-tumor immune response. Even though various checkpoint inhibitors show clinical efficacy by unleashing cytotoxic T cells activity, a large fraction of patients fails to respond to these immunotherapies. Therefore, a detailed understanding of the mechanisms of CD8 T cell exhaustion is required. Further, since the transcriptional signatures of T cell exhaustion are closely intertwined with their activated T cell state, single-cell analysis is an optimal approach to identify biomarkers specific to T cell dysfunction. In a single-cell RNA-seq analysis of T cells from hepatocellular carcinoma patients, 11 unique T cell subsets were identified based on their molecular and functional properties. Exhaustion signature gene LAYN was identified and associated with inhibition of IFN-γ production. A single-cell RNA-seq of CD8 tumor-infiltrating lymphocytes from murine tumor models has also aided identification of novel molecular pathways of T cell exhaustion that is uncoupled from T cell activation.

**Profiling of immune suppressive cell types present in the tumor microenvironment**

Single-cell transcriptome profiling enables characterization of the complex tumor microenvironment with its heterogeneous mixture of tumor cells along with stromal and immune cells. Targeting of immunosuppressive cell types in the tumor microenvironment can sometimes be key to the efficacy of checkpoint inhibitors such as anti-CTLA-4 therapy. A variety of cell types including T regulatory cells (Tregs), tumor-associated macrophages, type 2 NKT cells, M2 macrophages and MDSCs enforce immune suppression in the tumor helping tumor cells to survive anti-tumor immune attack. Identifying MDSCs has been challenging from bulk sequencing data due to the absence of unique MDSC markers. In addition, the presence of over 10 different myeloid subsets further complicates bioinformatics analysis. Tregs are potent immune modulators and assessing their frequency, phenotype, and function at tissue sites has been profoundly challenging due to the fact that majority of the defining markers like CD25, FOXP3 and CTLA4 are also present in effector T cells. Single-cell analysis of tumor infiltrated immune cells can help circumvent some of these hurdles in tumor characterization. In a recent single-cell analysis study tumor cells from 11 breast cancer patients, cancer cells were separated from immune cells based on their copy number variations. Analysis of the immune cell fraction revealed the presence of immunosuppressive macrophages of M2 phenotype and activated T effector cells. Interestingly, the T cells also expressed markers of T cell exhaustion such LAG3 and TIGIT suggesting that they could be targeted by immune checkpoint inhibitors.

**Understanding mechanisms of disease resistance**

Resistance to chemotherapy and molecularly targeted therapies is a major barrier to achieving long-term
benefit to treatment. ITH arising from diverse cell subpopulations with distinct molecular features produce varying levels of drug sensitivity and resistance. Retrospective analysis of CTCs from patients who had developed resistance to inhibitors of the androgen receptor (AR) showed higher activation of non-canonical Wnt pathway beside altered expression and mutations in AR compared to untreated patients. In castrate-resistant prostate cancers high content single-cell longitudinal profiling of CTCs from a patient undergoing chemotherapy and targeted therapy revealed a selective clonal expansion of cells with AR amplification supporting the adaptive model of therapy resistance evolution. Similar observation of selective clonal persistence was seen in breast cancer patients treated with chemotherapy. In this study, single-cell sequencing post-chemotherapy revealed transcriptional reprogramming of resistant signatures, elucidating the mechanism of therapy resistance.

Based on aforementioned studies, an accurate assessment of ITH by single-cell sequencing using multiregional, longitudinal sampling is essential to understand the mechanism of drug resistance and facilitate the development of more effective therapies.

FUTURE DIRECTIONS

With the development of precision microfluidic devices and sequencing technologies, single-cell analysis has transformed our understanding of ITH and clonal evolution. Single-cell genomics promises to deconvolute complex biological processes in cancer, reveal epigenetic alterations and monitor the evolution of metastatic and treatment resistance clones. By applying single-cell sequencing to different experimental systems, such as cells in culture, patient-derived xenografts, murine models and analysis of human tumors, novel diagnostics and therapies can be developed. A major hurdle in single-cell sequencing is the high cost of the technology. Moreover, the volume and complexity of single-cell sequencing datasets exceed that of the traditional bulk sequencing, calling for better statistical algorithms to deconvolute the data. Additional caution should be given on the transcriptome coverage and number of cells taken for single-cell analysis to ensure the accuracy of gene expression distribution estimates. Future breakthroughs in developing cost-effective sequencing methods and powerful data analysis pipeline for single-cell sequencing are likely to expand the scope of this technology beyond cancer to other diseases.

DECLARATIONS

Authors’ contributions
Manuscript drafting: Shi X, Chakraborty P, Chaudhuri A

Availability of data and materials
Not applicable.

Financial support and sponsorship
This work was supported by MedGenome Inc.

Conflicts of interest
All authors are full-time employees of MedGenome Inc.

Ethical approval and consent to participate
Not applicable

Consent for publication
Not applicable.

Copyright
© The Author(s) 2018.
REFERENCES


