Trends in Genetics

Special Issue: Human Genetics

Review

Single-Cell Analysis in Cancer Genomics

Assieh Saadatpour,[1,2] Shujing Lai,[3] Guoji Guo,[3,*] and Guo-Cheng Yuan[1,2,4,*]

Genetic changes and environmental differences result in cellular heterogeneity among cancer cells within the same tumor, thereby complicating treatment outcomes. Recent advances in single-cell technologies have opened new avenues to characterize the intra-tumor cellular heterogeneity, identify rare cell types, measure mutation rates, and, ultimately, guide diagnosis and treatment. In this paper we review the recent single-cell technological and computational advances at the genomic, transcriptomic, and proteomic levels, and discuss their applications in cancer research.

Cancer is a Disease of Multitudes

Cellular heterogeneity, which results from mutation, differences in gene regulation, stochastic variation, or environmental perturbations, is reflected at the genomic, transcriptomic, and proteomic levels. Such heterogeneity is increasingly appreciated as a factor of cancer treatment failure and disease recurrence, because a treatment that targets one tumor cell population may not be effective against another [1]. Not only is cancer itself a complex disease made up of a collection of individually distinct pathologies, but also within each tumor there is significant heterogeneity among different cells. Current theories propose that cancer development involves both a process of clonal evolution from mutated cells of origin and a differentiation hierarchy from cancer stem cells [2]. It is increasingly clear that traditional bulk experiments, which only measure the average profile of the population, have limitations in characterizing complex diseases such as cancer.

Single cells have been studied since the invention of the microscope, but it is not until recently that genome-scale approaches have been applied to single-cell biology [3–7]. For example, microfluidic-based single-cell sorting methods [8,9], high-throughput multiplexed quantitative PCR (qPCR) [10–14] or sequencing approaches [15–23], mass cytometry-based proteomic strategies [24–26], and data analysis methods [27–30] provided an unprecedented opportunity to identify rare cell types, such as cancer stem cells, and to investigate the dynamic processes of cell fate transitions.

One of the important application areas of single-cell analysis is in cancer genomics (Figure 1, Key Figure). Recently, several studies have applied single-cell analysis to characterize the cellular heterogeneity in different cancers [13,23,31–33]. The comprehensive knowledge about cellular heterogeneity will not only provide fundamental insights into development and other biological processes but also have important applications in therapy because drug resistance is often caused by heterogeneous response at the cellular level.

In this paper we review the recent technological and computational advances in single-cell analysis, and discuss their applications in cancer genomics. We conclude by offering a personal view of the potential challenges and future prospects for this field.

Trends

Recent advances in single-cell technologies have enabled researchers to profile mutations and expression levels of a large number of genes and proteins at individual cells.

Developments of new computational methods have greatly aided the calibration, quantification, and interpretation of single-cell data.

Single-cell analysis has been applied to study cancer initiation, variation, and evolution and will have potentially high clinical impact.

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[3]Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou 310058, China
[4]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

*Correspondence:
ggj@zju.edu.cn (Guo, G.),
gcyuan@jimmy.harvard.edu (Yuan, G-C.)

 http://dx.doi.org/10.1016/j.tig.2015.07.003

Key Figure

An Overview of Single-Cell Cancer Genomics

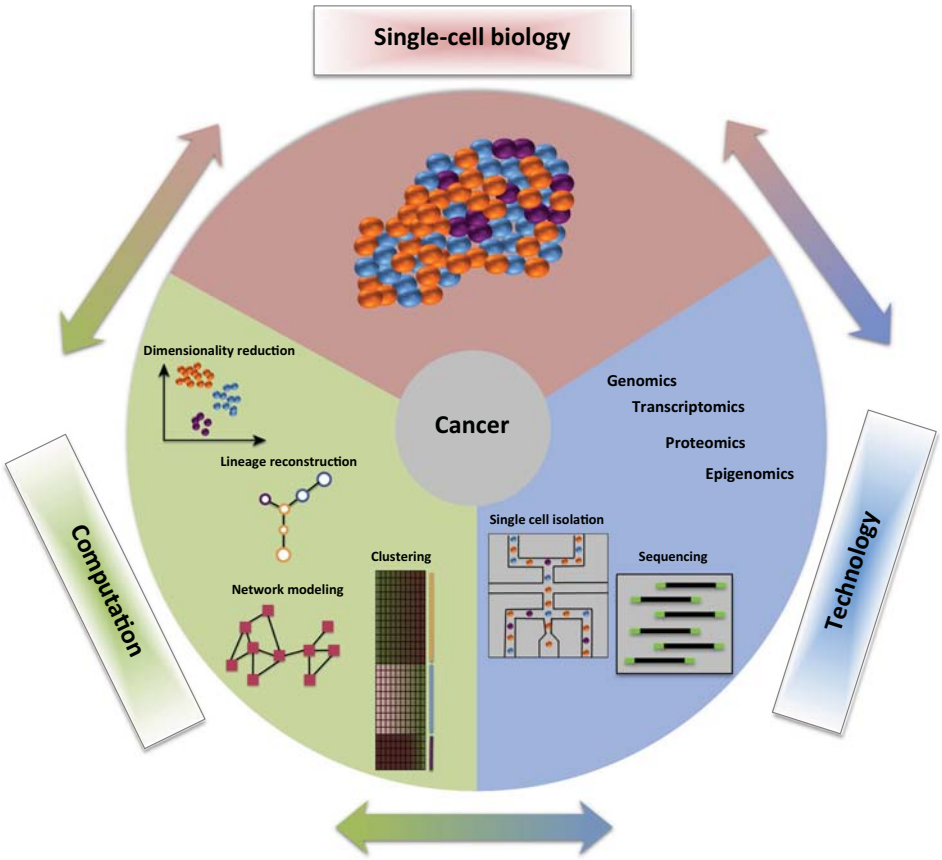

Figure 1. Single-cell technologies are used to generate genomic, transcriptomic, and proteomic data from cancer cells. These data are analyzed by computational methods to identify clusters, lineages, and networks, which in turn generate new biological hypotheses. Biological discoveries in turn guide the development of new technologies and computational approaches. The figure also shows a schematic example with a heterogeneous cancer sample containing three cell types (orange, blue, and purple). An integrated single-cell analysis is used to identify the cell types, lineages, and network profiles.

Technological Developments in Single-Cell Analysis

Methods for single-cell measurement, such as flow cytometry [34], RNA fluorescence *in situ* hybridization (FISH) [35,36], and dynamic profiling of fluorescent fusion proteins [37], were developed years ago and are routinely used in modern labs. However, these traditional methods provide limited information from single-cell samples because only a few genes or proteins can be profiled at the same time. In the past few years a new wave of technologies has emerged in the areas of single-cell isolation, nucleic acid amplification, and genomic/transcriptomic/proteomic profiling (Table 1). These new methods have significantly increased the throughput and scale of single-cell analysis.

Table 1. Advanced Single-Cell Technologies for Genomic, Transcriptomic, and Proteomic Analysis

Method	Amplification	Application	Coverage	Refs
Genomic Analysis				
GenomePlex PCR	Multiplexed PCR	Copy number	Low coverage	[32]
MDA	MDA	Genome/exome	High coverage	[38,39]
MALBAC	MALBAC	Copy number/genome	High coverage and uniform amplification	[41,42]
Transcriptomic Analysis				
Single-cell qPCR	Multiplexed PCR	Transcriptome	Targeted regions	[11,13]
Tang *et al.* method	PolyA tailing	Transcriptome	3′ Bias	[17,18]
Smart-seq	Template-switching	Transcriptome	Full-length	[16]
CEL-seq	IVT	Transcriptome	3′ Bias	[45]
CytoSeq	Multiplexed PCR	High-throughput transcriptome	Targeted regions	[47]
inDrop	IVT	High-throughput transcriptome	3′ Bias	[48]
Drop-seq	Template-switching	High-throughput transcriptome	3′ Bias	[49]
Proteomic Analysis				
Mass Cytometry	N/A[a]	Proteomic analysis	Targeted proteins	[24]
MIBI	N/A	Proteomic analysis with spatial information	Targeted proteins	[58]

[a]NA, not applicable.

One of the fundamental challenges in single-cell analysis is the amplification of a small amount of initial nucleic acid material to reach the detection threshold level. Recently, significant technical advances in whole-genome amplification (WGA) have been achieved to overcome this challenge for single-cell genome analysis. Based on the protocols used for WGA, there are three main categories of single-cell techniques. GenomePlex PCR [32] uses degenerate-oligonucleotide PCR to amplify DNA from single cells. The method achieves low physical coverage, but the amplification is uniform across the genome. It is therefore suitable for copy-number profiling in single cancer cells [32]. Another popular single-cell WGA method, multiple displacement amplification (MDA), uses bacteriophage Φ29 polymerase and random primers to amplify DNA in a linear process through multiple displacement mechanisms [38,39]. This approach generates long DNA products and achieves high-coverage amplification, and therefore is suitable for the detection of point mutations at base-pair resolution. The MDA protocol was first used in single-cell exome-sequencing studies to uncover the genetic landscape of cancer cells [38,39], and was subsequently coupled with a microfluidic system to amplify genomes from single human spermatozoa [40]. Multiple annealing and looping-based amplification cycles (MALBAC) [41,42] is a new WGA method that uses quasi-linear pre-amplification to reduce the bias associated with nonlinear amplification. In MALBAC, single-stranded amplicons generated through strand-displacement are used as templates to produce full amplicons, and then the full amplicons form looped DNA to avoid exponential amplification. This approach achieves high coverage and uniform amplification, and enables genome-wide detection of both single-nucleotide polymorphisms (SNPs) and copy-number variations (CNVs) of a single cell. The method has been applied to single SW480 cancer cells [41] as well as to human oocytes [42].

Another frontier with significant progress is single-cell transcriptomic analysis. Although there are more copies of mRNA than DNA in single cells, this application faces its own difficulties in quantification of different RNA species. To amplify the limited amount of mRNA in single cells,

several approaches have emerged. The poly-A tailing method uses terminal transferase to add anchoring sequences to the 3′ ends of the synthesized cDNA, so that each cDNA has two primer binding sites for PCR amplification. The method was used in the first single-cell microarray study [43] and in the first single-cell high-throughput mRNA sequencing (mRNA-seq) study [17]. Sequence-specific amplification (SSA) uses multiplexed reverse transcription and PCR (RT-PCR) to amplify hundreds of specific targets in single cells. This method has a simple one-step protocol but is limited to analyzing only a small number of genes [11,13]. The Smart-seq amplification method is a widely used approach for full-length mRNA analysis of single cells [16,21,44]. The method uses a protocol based on template-switching to anchor a primer binding site at the 3′ end of the cDNA. The cDNA is then amplified by PCR and sequenced by Illumina sequencers. Smart-seq has high coverage across transcripts, and enables identification of SNPs as well as different transcript isoforms. One limitation of Smart-seq is that the efficiency for template-switching is low and thus it has difficulty in profiling poorly expressed mRNAs. CEL-seq (single-cell RNA-seq using multiplexed linear amplification) adds bacteriophage T7 promoters to the cDNA and utilizes *in vitro* transcription (IVT) to amplify mRNA. The method also shows robust efficiency and sensitivity for single-cell transcriptomic profiling [15,45]. By coupling IVT with a degenerate PCR-based approach, the recently published DR-seq method was able to achieve integrated genome and transcriptome sequencing at the same time from the same cell [46].

For all the aforementioned single-cell transcriptomic methods, a common drawback is the need to handle each cell sample independently, which limits the throughput of the analysis and also may inadvertently introduce human error. Very recent breakthroughs solve these problems by high-throughput molecular barcoding of single cells in microwells or microdroplets before sequencing-library generation [47–49]. The CytoSeq platform randomly deposits single cells and transcript barcoding probes into an array of picoliter wells before cell lysis and reverse transcription; any selection of genes can be amplified and analyzed from the barcoded cDNAs [47]. The inDrop and Drop-seq strategies, however, separate thousands of single cells into aqueous droplets, associate a different barcode to the RNAs from each cell, and sequence them all simultaneously [48,49]. These massively-parallel barcoding strategies have significantly increased the throughput of single-cell transcriptomic analysis.

The broad applications of single-cell genomic/transcriptomic analysis in the biomedical field have also been supported by the rapid development of microfluidic devices. Microfluidic devices help to automate the distribution, processing, and analysis of biological materials, and have significantly increased the measurement throughput. Microfluidic devices have been used as the basis for various single-cell technologies, such as the single-cell capture and amplification platforms [44,49], as well as high-throughput single-cell qPCR analysis [13]. Because single-cell analysis protocols are highly sensitive to technical errors induced by manual processing, the accurate control provided by the microfluidic devices is a significant advantage. Microfluidic devices also improve the sensitivity of single-cell assays by confining the reaction volume and increasing the local concentration.

In comparison to the progress made in assaying nucleic acids, single-cell proteomic analysis is much more challenging because, unlike DNA or RNA sequences, it is not possible to amplify protein sequences using current technologies. Standard immunofluorescence methods have been routinely used to analyze four markers at single-cell level. Highly multiplexed fluorescence microscopic now allows analysis of up to 60 proteins in tissue specimens [50]. Notably, the development of mass cytometry has dramatically increased the multiplexity of cytometry-based analysis by labeling antibodies with isotopes [24]. This innovation resolves the problem of spectral overlap that is common in normal flow cytometry. It is now possible to measure more than 40 parameters in a large number of single cells in a short period of time.

The methods discussed above require the isolation of cells from their *in situ* environment. Recently *in situ* methods have been developed to preserve spatial information [51]. By computational integration of single-cell RNA-seq data with *in situ* RNA patterns, one can accurately infer cellular localization within complex patterned tissues [52–56]. Similarly, mass cytometry can be coupled with immunohistochemical data to obtain highly multiplexed proteomic information at subcellular resolution [57]. Another method, termed multiplexed ion beam imaging (MIBI), uses secondary-ion mass spectrometry to image antibodies tagged with isotopically-pure elemental metal reporters [58]. Taken together, these technologies have greatly facilitated the systematic analysis of gene and protein expression variability at single-cell resolution.

Computational Methods for Analyzing Single-Cell Data

With the technological breakthroughs that have generated large amounts of high-throughput single-cell data, the development of novel computational tools has become an integral part of the analysis. Single-cell technologies present a number of challenges that cannot be addressed by traditional computational methods. First, each cell is typically measured only once, and thus there are no technical replicates in the strict sense. Second, the amount of starting material is subject to strong stochastic variation. Still at an early stage, several computational methods have been developed to address these issues (Figure 2).

Preprocessing and quantification are the first steps of any large-scale data analysis. The purpose of these steps is to convert raw data to quantitative biological information. Significant effort is

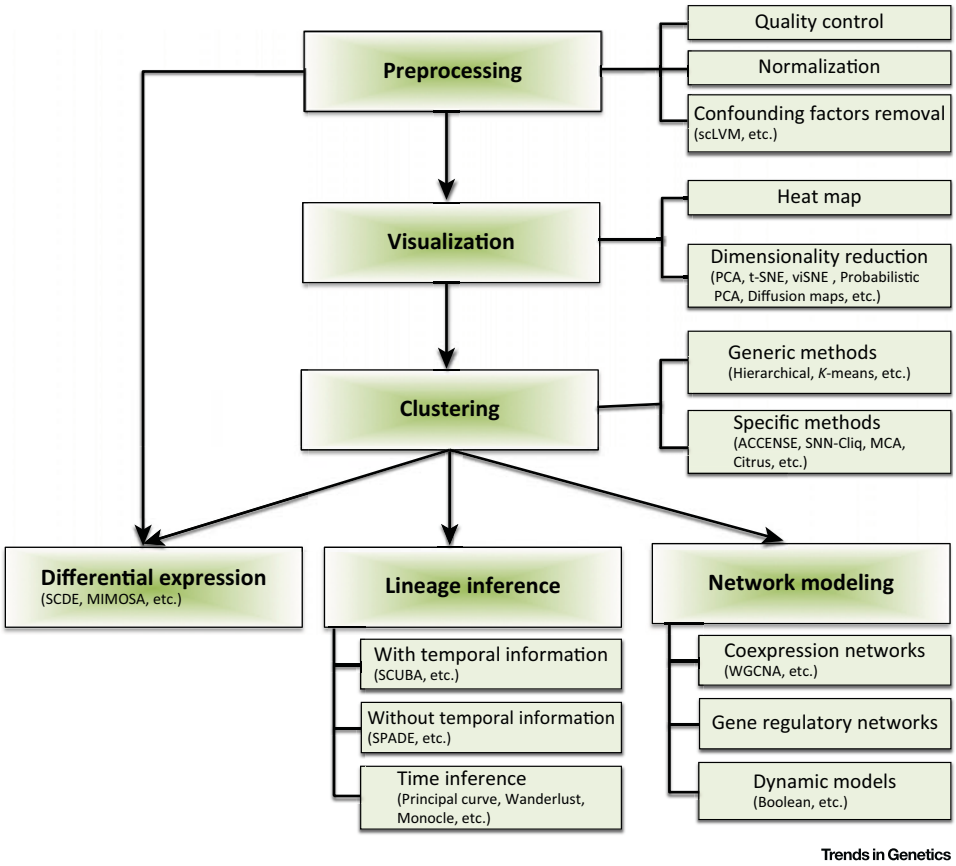

Figure 2. A Typical Flowchart for Single-Cell Data Analysis. Representative methods are mentioned. See the main text for detailed description.

paid to the estimation and removal of systematic biases due to technical variability. A major issue in single-cell analysis is that technical variation is always confounded with biological variation. One simple approach to estimate technical variability compared the pooled sample with bulk RNA-seq experiments [59]. More precise calibration can be achieved by adding spike-in RNA to the library as a control, followed by building an error model based on the variation of the spiked-in RNA [60,61]. Methods that account for single-cell specific noise, such as dropout events and amplification biases, can also help to separate technical and biological variability of individual genes [62]. Recently, scLVM (single-cell latent variable model) [63] was developed to account for the confounding effects of the cell cycle on modulating differentiation and gene expression profiles. Another approach to estimating reproducibility is to divide the RNA material from a single cell into two equal fractions which are then analyzed independently [20]. A recent review article [64] has provided a detailed survey of the computational methods and, in particular, the normalization steps for single-cell RNA-seq data from counts to expression values, with or without unique molecular identifiers. For single-cell qPCR data, normalization to an endogenous control is not usually recommended due to the biological variation and transcriptional noise exhibited by single cells [65]. It was shown that normalizing by the median Ct (threshold cycle) reduces variability in single-cell qPCR data [65]. For mass-cytometry data, technical variations can be corrected with bead standards [66]. Normalization methods for CNV detection based on channel, genome composition, and recurrent genome artifact corrections have also been developed [67]. A pipeline of the computational approaches to correct for biases in the WGA procedure and accurately determine copy-number profiles has been presented before [68].

The high dimensionality of single-cell data provides a challenge for visualization. Several dimensionality-reduction approaches are available to map the datapoints into a lower-dimensional space while maintaining the single-cell resolution. The conventional principal component analysis (PCA) has been used to visualize single-cell data in different contexts [13,24,69]. Despite its success, this method relies on a linear assumption, and thus cannot fully capture the nonlinear relationships inherent in many single-cell datasets. This limitation can potentially be overcome by using a wide variety of non-linear methods [27,30,70–73], although the performance of each method is likely to be context-dependent. The t-distributed stochastic neighbor embedding (t-SNE) method [72,73] preserves both the global layout and local structure of the high-dimensional data by converting the Euclidean distances between each pair of datapoints into heavy-tailed conditional probabilities. A distributed implementation of the t-SNE algorithm, termed viSNE [27], has been employed to visualize single-cell mass cytometry data. Another approach based on the Gaussian process latent variable model generates a smooth mapping from the latent space to the original data space [71]. This method was extended to a probabilistic PCA approach to account for the censoring effect due to undetected transcripts [70]. More recently, a dimensionality-reduction approach based on diffusion maps was adapted to identify and visualize hematopoietic developmental progression in mouse embryo [30]. In this approach, the cells are related to each other through a gradual but stochastic, diffusion-like process. All these methods can help to interrogate the relationships among different cell types in a lower-dimensional space.

Unsupervised clustering is a widely used approach to group samples with similar properties, which can be used for identifying previously unknown subpopulations from single-cell data. In addition to the classical clustering methods, several approaches have recently been developed to analyze single-cell data. For example, ACCENSE (automatic classification of cellular expression by nonlinear stochastic embedding) [74] combines the t-SNE algorithm with density-based partitioning without the need to pre-specify the number of target clusters. Another recent effort in this direction is SNN-Cliq [75], which achieves clustering of single-cell transcriptomic data by a graph theory-based algorithm. For low-dimensional single-cell expression data emerging from qPCR or FACS, multiresolution correlation analysis (MCA) [76] can be useful in identifying

subpopulations based on local pairwise gene correlations. In an effort to improve traditional manual gating in flow cytometry data, Citrus [77] was developed, which identifies stratifying subpopulations of cells whose abundance or behavior correlates with a known endpoint of interest.

Having identified the cell subpopulations, one can determine the sets of genes that best discriminate these subpopulations. In addition to the standard differential expression tools for bulk experiments, new methods have been developed to address the specific challenges in single-cell data analysis. For example, SCDE (single-cell differential expression) uses a Bayesian approach that accounts for the likelihood of dropout events in single-cell RNA-seq data [62]. In another approach, MIMOSA (mixture models for single-cell assays) employs a mixture model where information is shared across subjects through exchangeable priors, allowing an increase in the power to detect true differences [78].

Although clustering approaches reveal the underlying group structure within the data, they cannot provide information on the lineage relationships between different developmental stages. One method along this direction is SCUBA (single-cell clustering using bifurcation analysis) [28], which first infers the cellular hierarchy using dynamic clustering and then models gene expression dynamics using bifurcation theory. However, the application of SCUBA requires temporal information, which is often difficult to obtain experimentally. Computational methods have been developed to infer temporal information from snapshot single-cell data, including using principal curve analysis [79], and graph-model based algorithms such as Monocle [29] and Wanderlust [80]. The inferred temporal information can then be used as the input to identify bifurcation events [28]. However, it is challenging to accurately infer temporal information if the bifurcation structure is complex. A related approach is SPADE (spanning-tree progression analysis of density-normalized events) [24,81], which infers cell lineages without assigning temporal order. In this case, additional biological knowledge is necessary to interpret the resulting tree structure. Similar approaches have been developed to infer clonal structure using single-cell genomic data [82].

Network modeling can provide mechanistic insights into the coordination of gene activities and help in understanding the overall dynamics of the system. Efforts are underway to apply network-modeling approaches to single-cell data. A simple but popular approach is to construct networks based on coexpression data. For example, an approach, termed weighted gene coexpression network analysis (WGCNA) [83,84], uses a soft threshold for modeling coexpression and also identifies network modules (i.e., genes with coordinated activities). Coexpression networks have been applied to single-cell analysis of the mammalian embryonic development [19], hematopoiesis [14], neural stem cells [85], and leukemia [33,86]. Although network analysis provided novel insights in these studies, existing methods are applicable only if the sample size is sufficiently large, and are therefore not directly applicable to studying networks associated with rare cell types. In addition, correcting for latent confounding factors in single-cell data can help to reduce false positive links in these networks [63,87]. Coexpression networks can also be integrated with other types of data, such as chromatin precipitation combined with high-throughput sequencing (ChIP-seq) data, to estimate the underlying gene regulatory network [10]. Most network models are only a static representation of the system and do not explicitly consider the underlying gene expression dynamics. Building self-contained dynamic network models is challenging, although there are some approaches, such as Boolean networks, that have been applied to study stem cell differentiation processes [30,88].

Taken together, these computational methods have greatly enabled researchers to systematically extract quantitative information from the single-cell data, thereby playing an important role in applying single-cell technologies to investigate biomedical problems.

Applications in Cancer Genomics

Genome instability is a hallmark of cancer. Spatial and temporal knowledge of the cancer genome will have a significant impact on identifying cancer subtypes and developing patient-specific treatment strategies. A notable application of single-cell genome sequencing is in inferring tumor evolution paths. For example, single-cell genome sequencing applied to two human breast cancer cases suggested that tumors grow by punctuated clonal expansions with few persistent intermediates [32]. Another study on a thrombocythemia patient uncovered the likely monoclonal origin of this neoplasm [38]. Compared to the widely used bulk-sequencing methods, single-cell cancer genome analysis has the advantage of characterizing intratumor cellular heterogeneity. For example, it has been used to map the intratumoral genetic landscape in kidney cancer [39], colorectal cancer [31], and leukemia [82]. Another intensely researched area is the detection and sequencing of circulating tumor cells (CTCs) either for understanding the metastatic process or for early tumor detection. For example, a study on the reproducibility of CNV patterns in CTCs of lung cancer patients suggested that CNVs at specific genomic loci are selected for during cancer metastasis [89]. A recent whole-exome sequencing of CTCs provided insights into the mutational landscape of metastatic prostate cancer [90]. Recently, a high-coverage, whole-genome/exome single-cell sequencing method (Nuc-seq) was developed and applied to breast cancer data wherein a large number of subclonal and *de novo* mutations were found, suggesting that point mutations evolved gradually over long periods of time [91]. In another study, by using single-cell whole-exome sequencing in multiple myeloma, it was demonstrated that the disease develops through a branching and parallel evolutionary pattern, where two divergent clones independently acquired the same convergent phenotype [92].

Single-cell transcriptomic advances in cancer research are also notable. For example, by using single-cell qPCR analysis in human colon cancer, it was found that multi-lineage differentiation represents a key source of *in vivo* functional and phenotypic cancer cell heterogeneity [13]. The Smart-seq method was used for profiling full-length mRNA from single cells wherein, by analyzing CTCs from melanomas, distinct gene expression signatures as well as alternative-splicing events specific to the disease were identified [16]. A recent single-cell qPCR analysis of a mouse model of acute myeloid leukemia identified two subpopulations of leukemic cells, each characterized by distinct coexpression networks [33]. Another study using single-cell RNA-seq analysis in five primary glioblastomas (GBs) revealed that current GB classifiers are variably expressed across single cells within a tumor, suggesting that single-cell data can capture the true diversity of transcriptional subtypes within a tumor that cannot be detected by population-level data alone [23].

Single-cell proteomic approaches, ranging from flow cytometry to mass cytometry and multiplexed imaging, have also made great contributions to cancer research [27,57,58,93,94]. For example, an application of the viSNE approach to mass cytometry data on healthy and leukemic bone marrow samples showed that, although the maps of healthy samples overlap, the leukemic samples from different patients form distinct populations from healthy bone marrow and from each other [27]. Moreover, integration of mass cytometry with multiplexed imaging techniques on breast cancer samples revealed substantial tumor microenvironment heterogeneity [57,58].

All these examples demonstrate that single-cell technologies provide a powerful approach to study the diversity and evolution of single cancer cells, which can ultimately be applied to the clinic from early detection to identifying therapeutic strategies for cancer patients.

Concluding Remarks and Future Perspectives

Single-cell analysis is still a new field, and several significant challenges lie ahead (see Outstanding Questions). A major goal for technological development is to improve the throughput and accuracy of the assays while reducing the cost. Promising results have been obtained by the

Outstanding Questions

How can single cells be isolated while maintaining the temporal and spatial information?

How can technical variations be distinguished from biological variations in single-cell data?

How can mechanistic studies be integrated with single-cell gene expression data?

How can single-cell data be used in clinical decision-making?

recent development of several approaches such as massive barcoding, microwells, and microdroplets [47–49]. Most technologies for single-cell analysis require the destruction of cells, and thus the temporal information is lost during the process. Along these lines, live-cell imaging technologies have generated exciting results [95]. Similarly, isolating single cells from a tissue results in loss of information about the spatial context, imposing a barrier for investigating the role of microenvironmental factors in gene regulation and cell fate decisions. This issue is especially problematic for studying tumor progression, which is known to depend heavily on its interaction with the microenvironment. In this direction, several promising approaches have been developed as discussed above [51–56]. Similarly, methods for single-cell epigenomic profiling are still underdeveloped, although some promising strides have been made [96–100]. Further developments in this area would help to dissect the role of DNA methylation heterogeneity in cancer cells. Ideally, this would involve the measurement of gene expression, chromatin states, and DNA methylation states in a single cell to elucidate the precise regulatory mechanisms at single-cell resolution. However, such an integrated approach will require applying multiple measurement platforms to the same molecule without alteration of its properties, a task that seems to be daunting if not impossible.

Computational method development is an integral component of every new technology. However, single-cell analysis presents unique challenges that require not only incremental changes but also revolutionary breakthroughs. Each analytical pipeline begins with extracting the signal from raw data, which requires the identification and correction of systematic technical noise to properly calibrate different samples and batches. For single-cell analysis, it remains challenging to distinguish technical variations from biological variations [60–63]. Rare cell types, by definition, consist of a small cellular population and may be detectable only in a relatively large cellular population. It is difficult to distinguish them from technical artifacts because traditional clustering algorithms, which favor robustness, tend to identify large subpopulations. Another challenge is to distinguish transient variations, such as those caused by stochastic noise or regular cell cycle variation, from those that are essential for cell identity. Integration analysis of gene expression data with chromatin states and transcription factor binding information has been very useful for understanding the underlying gene regulatory networks. However because it remains difficult to map chromatin states and transcription binding at single-cell resolution, it is important, but challenging, to develop novel computational methods to integrate single-cell gene expression data with population level datasets. Perhaps most importantly, substantial efforts will be necessary to improve single-cell technologies and computational methods for these to have direct implications in clinical decision-making.

Despite these challenges, single-cell analysis undoubtedly presents tremendous opportunities. Taken together, applications of single-cell analysis will greatly enhance the power of systematic characterization of cancer heterogeneity and lead to mechanistic insights into cancer progression, which ultimately will aid the development of novel therapeutic strategies, help us to better understand the mechanisms of drug resistance, and lead to improvement of clinical outcomes.

Acknowledgments

This work was supported in part by a Claudia Adams Barr Award to G-C.Y.

References

1. Bedard, P.L. *et al.* (2013) Tumour heterogeneity in the clinic. *Nature* 501, 355–364
2. Clevers, H. (2011) The cancer stem cell: premises, promises and challenges. *Nat. Med.* 17, 313–319
3. de Vargas Roditi, L. and Claassen, M. (2014) Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics. *Curr. Opin. Biotechnol.* 34C, 9–15
4. Di Palma, S. and Bodenmiller, B. (2015) Unraveling cell populations in tumors by single-cell mass cytometry. *Curr. Opin. Biotechnol.* 31, 122–129
5. Junker, J.P. and van Oudenaarden, A. (2014) Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* 157, 8–11
6. Navin, N.E. (2014) Cancer genomics: one cell at a time. *Genome Biol.* 15, 452

7. Tsioris, K. *et al.* (2014) A new toolbox for assessing single cells. *Annu. Rev. Chem. Biomol. Eng.* 5, 455–477
8. Nicholas, C.R. *et al.* (2007) A method for single-cell sorting and expansion of genetically modified human embryonic stem cells. *Stem Cells Dev.* 16, 109–117
9. Thorsen, T. *et al.* (2002) Microfluidic large-scale integration. *Science* 298, 580–584
10. Guo, G. *et al.* (2013) Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 13, 492–505
11. Guo, G. *et al.* (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685
12. Buganim, Y. *et al.* (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209–1222
13. Dalerba, P. *et al.* (2011) Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29, 1120–1127
14. Moignard, V. *et al.* (2013) Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* 15, 363–372
15. Jaitin, D.A. *et al.* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779
16. Ramskold, D. *et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782
17. Tang, F. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382
18. Tang, F. *et al.* (2011) Development and applications of single-cell transcriptome analysis. *Nat. Methods* 8, S6–S11
19. Xue, Z. *et al.* (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597
20. Deng, Q. *et al.* (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196
21. Shalek, A.K. *et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240
22. Shalek, A.K. *et al.* (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510, 363–369
23. Patel, A.P. *et al.* (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401
24. Bendall, S.C. *et al.* (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 687–696
25. Behbehani, G.K. *et al.* (2012) Single-cell mass cytometry adapted to measurements of the cell cycle. *Cytometry A* 81, 552–566
26. Bodenmiller, B. *et al.* (2012) Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* 30, 858–867
27. Amir, E.D. *et al.* (2013) viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* 31, 545–552
28. Marco, E. *et al.* (2014) Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5643–E5650
29. Trapnell, C. *et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386
30. Moignard, V. *et al.* (2015) Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33, 269–276
31. Yu, C. *et al.* (2014) Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing. *Cell Res.* 24, 701–712
32. Navin, N. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94
33. Saadatpour, A. *et al.* (2014) Characterizing heterogeneity in leukemic cells using single-cell gene expression analysis. *Genome Biol.* 15, 525
34. Fulwyler, M.J. (1965) Electronic separation of biological cells by volume. *Science* 150, 910–911
35. Maamar, H. *et al.* (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317, 526–529
36. Raj, A. *et al.* (2010) Variability in gene expression underlies incomplete penetrance. *Nature* 463, 913–918
37. Taniguchi, Y. *et al.* (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538
38. Hou, Y. *et al.* (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 148, 873–885
39. Xu, X. *et al.* (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148, 886–895
40. Wang, J. *et al.* (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150, 402–412
41. Zong, C. *et al.* (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626
42. Hou, Y. *et al.* (2013) Genome analyses of single human oocytes. *Cell* 155, 1492–1506
43. Kurimoto, K. *et al.* (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* 34, e42
44. Treutlein, B. *et al.* (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375
45. Hashimshony, T. *et al.* (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673
46. Dey, S.S. *et al.* (2015) Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33, 285–289
47. Fan, H.C. *et al.* (2015) Combinatorial labeling of single cells for gene expression cytometry. *Science* 347, 1258367
48. Klein, A.M. *et al.* (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201
49. Macosko, E.Z. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214
50. Gerdes, M.J. *et al.* (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11982–11987
51. Crosetto, N. *et al.* (2015) Spatially resolved transcriptomics and beyond. *Nat Rev Genet* 16, 57–66
52. Achim, K. *et al.* (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* 33, 503–509
53. Chen, K.H. *et al.* (2015) Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090
54. Lee, J.H. *et al.* (2014) Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363
55. Lubeck, E. *et al.* (2014) Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11, 360–361
56. Satija, R. *et al.* (2015) Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol* 33, 495–502
57. Giesen, C. *et al.* (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 11, 417–422
58. Angelo, M. *et al.* (2014) Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* 20, 436–442
59. Marinov, G.K. *et al.* (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* 24, 496–510
60. Brennecke, P. *et al.* (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10, 1093–1095

61. Grun, D. *et al.* (2014) Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11, 637–640
62. Kharchenko, P.V. *et al.* (2014) Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742
63. Buettner, F. *et al.* (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33, 155–160
64. Stegle, O. *et al.* (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145
65. Livak, K.J. *et al.* (2013) Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. *Methods* 59, 71–79
66. Finck, R. *et al.* (2013) Normalization of mass cytometry data with bead standards. *Cytometry A* 83, 483–494
67. Cheng, J. *et al.* (2011) Single-cell copy number variation detection. *Genome Biol.* 12, R80
68. Baslan, T. *et al.* (2012) Genome-wide copy number analysis of single cells. *Nat. Protoc.* 7, 1024–1041
69. Stahlberg, A. *et al.* (2011) Defining cell populations with single-cell gene expression profiling: correlations and identification of astrocyte subpopulations. *Nucleic Acids Res.* 39, e24
70. Buettner, F. *et al.* (2014) Probabilistic PCA of censored data: accounting for uncertainties in the visualization of high-throughput single-cell qPCR data. *Bioinformatics* 30, 1867–1875
71. Buettner, F. and Theis, F.J. (2012) A novel approach for resolving differences in single-cell gene expression patterns from zygote to blastocyst. *Bioinformatics* 28, i626–i632
72. van der Maaten, L. (2014) Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* 15, 3221–3245
73. van der Maaten, L.J.P. and Hinton, G.E. (2008) Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605
74. Shekhar, K. *et al.* (2014) Automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE). *Proc. Natl. Acad. Sci. U.S.A.* 111, 202–207
75. Xu, C. and Su, Z. (2015) Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* 31, 1974–1980
76. Feigelman, J. *et al.* (2014) MCA: multiresolution correlation analysis, a graphical tool for subpopulation identification in single-cell gene expression data. *BMC Bioinformatics* 15, 240
77. Bruggner, R.V. *et al.* (2014) Automated identification of stratifying signatures in cellular subpopulations. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2770–E2777
78. Finak, G. *et al.* (2014) Mixture models for single-cell assays with applications to vaccine studies. *Biostatistics* 15, 87–101
79. Hastie, T. and Stuetzle, W. (1989) Principal curves. *J. Am. Stat. Assoc.* 84, 502–516
80. Bendall, S.C. *et al.* (2014) Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157, 714–725
81. Qiu, P. *et al.* (2011) Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* 29, 886–891
82. Gawad, C. *et al.* (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17947–17952
83. Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, 17
84. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559
85. Luo, Y. *et al.* (2015) Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. *Cell* 161, 1175–1186
86. Kouno, T. *et al.* (2013) Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome Biol.* 14, R118
87. McDavid, A. *et al.* (2014) Modeling bi-modality improves characterization of cell cycle on gene expression in single cells. *PLoS Comput. Biol.* 10, e1003696
88. Xu, H. *et al.* (2014) Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput. Biol.* 10, e1003777
89. Ni, X. *et al.* (2013) Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc. Natl. Acad. Sci. U.S.A.* 110, 21083–21088
90. Lohr, J.G. *et al.* (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotechnol.* 32, 479–484
91. Wang, Y. *et al.* (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512, 155–160
92. Melchor, L. *et al.* (2014) Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia* 28, 1705–1715
93. Bendall, S.C. and Nolan, G.P. (2012) From single cells to deep phenotypes in cancer. *Nat Biotechnol* 30, 639–647
94. DiGiuseppe, J.A. *et al.* (2015) Detection of minimal residual disease in B lymphoblastic leukemia using viSNE. *Cytometry B Clin. Cytom.* Published online May 14, 2015. http://dx.doi.org/10.1002/cyto.b.21252
95. Etzrodt, M. *et al.* (2014) Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* 15, 546–558
96. Smallwood, S.A. *et al.* (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820
97. Nagano, T. *et al.* (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64
98. Guo, H. *et al.* (2014) The DNA methylation landscape of human early embryos. *Nature* 511, 606–610
99. Kind, J. *et al.* (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178–192
100. Buenrostro, J.D. *et al.* (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490