Challenges and emerging directions in single-cell analysis

Guo-Cheng Yuan[1,2,*], Long Cai[3], Michael Elowitz[4], Tariq Enver[5], Guoping Fan[6], Guoji Guo[7], Rafael Irizarry[1,2], Peter Kharchenko[8], Junhyong Kim[9], Stuart Orkin[10,11,12], John Quackenbush[1,2], Assieh Saadatpour[1,2], Timm Schroeder[13], Ramesh Shivdasani[14], Itay Tirosh[15]

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, USA

[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA

[3]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, USA

[4]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, USA

[5]Cancer Institute, University College London, London, United Kingdom

[6]Department of Human Genetics, University of California, Los Angeles, USA

[7]Center for Stem Cell and Regenerative Medicine, Zhejiang University, Hangzhou, China

[8]Department of Biomedical Informatics, Harvard Medical School, Boston, USA

[9]Department of Biology, University of Pennsylvania, Philadelphia, USA

[10]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, USA

[11]Division of Hematology/Oncology, Boston Children's Hospital, Boston, USA

[12]Howard Hughes Medical Institute, USA

[13]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

[14]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, USA

[15]Broad Institute of MIT and Harvard, Cambridge, USA

[*]Corresponding author: Email: gcyuan@jimmy.harvard.edu

Abstract

Single-cell analysis is a rapidly evolving approach to characterize genome-scale molecular information at the individual cell level. Development of single-cell technologies and computational methods has enabled systematic investigation of cellular heterogeneity in a wide range of tissues and cell populations, yielding fresh insights into the composition, dynamics, and regulatory mechanisms of cell states in development and disease. Despite substantial advances, significant challenges remain in the analysis, integration, and interpretation of single-cell omics data. Here, we discuss the state of the field and recent advances, and look to future opportunities.

Background

Cell-to-cell variation is a universal property of multi-cellular organisms, which contain diverse cell types characterized by different functions, morphologies, and gene expression profiles. Even within any single tissue, no matter how apparently homogeneous, there is a diverse population of cells, all of which represent different manifestations of that tissue type. Investigation of tissues or cell populations is inherently limited by the fact that the readout of any pooled assay that uses bulk tissue represents a weighted average of that population's cellular constituents. Intrinsic cellular heterogeneity is obscured in the typical ensemble studies on which the canon of modern biology and medicine are constructed.

Consider, for example, the diverse repertoire of cells present in the three most rapidly self-renewing tissues in mammals: blood, skin, and the intestinal epithelium. Although the trajectory from stem to terminally differentiated cell is almost certainly a continuum

of highly variable states, our limited understanding forces us to regard known stem and progenitor cell populations as discrete and stable entities. Even in post-mitotic tissues such as the adult brain, the differentiated cell states resulting from complex bifurcating developmental trajectories may also appear as a continuum. The diversity of cellular states is not only caused by their own inherent cell-to-cell variability, but also influenced by interactions among tens or even hundreds of distinct cells. These considerations question the precise boundary of a cell type and point to the need for single cell analysis to dissect the underlying complexity and the empirical reality of stable and distinct cell states.

The past few years have seen the introduction of technologies that provide genome-scale molecular information at the resolution of single cells, providing unprecedented power for systematic investigation of cellular heterogeneity in DNA [1, 2], RNA[3], proteins [4], and metabolites[5]. These technologies have been applied to identify previously unknown cell types and associated markers [6-8] and to predict developmental trajectories [9-13].

Beyond expanding the catalog of mammalian cell states and identities, single cell analyses have challenged prevailing ideas of cell-fate determination [14-19] and opened new ways of studying the mechanisms associated with disease development and progression. For example, single-cell DNA sequencing (scDNA-seq) has revealed remarkable cellular heterogeneity inside each tumor, significantly revising models of clonal evolution [20-22], whereas single-cell RNA sequencing (scRNA-seq) has shed new lights on the role of tumor microenvironments in disease progression and drug resistance [23].

The ambitious goal of understanding the full complexity of cells in a multi-cellular organism collectively requires not only experimental methods that are considerably better than existing platforms, but synchronous development of computational methods that can be used to derive useful insights from complex and dense data on large numbers of diverse single cells. Several recent papers have discussed various

challenges critical to advance the incipient field of single cell analysis [24-27]; here we expand on these discussions with a focus on looking to the future.

Current challenges in analyzing single-cell data:

While many methods have been successfully used for the analysis of genomic data from bulk samples, the relatively small number of sequencing reads, the sparsity of data and cell population heterogeneity present significant analytical challenges in effective data analysis. Recent advances in computational biology have greatly enhanced the quality of data analyses and provided important new biological insights [24-27].

Data preprocessing:

The goal of data preprocessing is to convert the raw measurements to bias-corrected and biologically meaningful signals. Here we focus on scRNA-seq, which has become the primary tool for single cell analysis. Gene expression profiling by scRNA-seq is inherently noisier than bulk RNA-seq, as vast amplification of small amounts of starting material combined with sparse sampling introduce significant distortions. A typical single-cell gene expression matrix contains excessive zero entries. The limited efficiency of RNA capture and conversion rate combined with DNA amplification bias may lead to significant distortion of the gene expression profiles. On one hand, even transcripts that are expressed at a highly level may occasionally evade detection altogether, resulting false-negative errors. On the other hand, transcripts that are expressed at a low level may appear abundant due to amplification biases. These errors artificially inflate the estimate of the cell-to-cell variability. While a number of methods have been developed to address this issue [28-30], managing dropout events continues to be a challenge. Another source of technical variation is the batch effect, which can be introduced when cells from one biological group are cultured, captured, and sequenced separately from cells in a second condition. If a scRNA-seq experiment is designed improperly, the results can be significantly affected by batch effects [31]. Furthermore,

high throughput technologies typically involve multiplexing of thousands or more barcode sequences. Error in demultiplexing may be caused by barcode impurities or external background, which has become increasingly challenging as thousands or more cells are multiplexed by recent technologies. Finally, the cell-to-cell variation may also be attributed to cell size, cell cycle state, and other factors that are irrelevant for cell type identification. Statistical models have been developed to remove such confounding factors [27]. Together, these technical artifacts pose important challenges for data calibration and interpretation.

The entanglement of technical and biological variations poses a significant challenge for evaluating data reproducibility. One approach to directly measure technical variability is to use dilute bulk RNA to approximately single cell level (~10-50 pg of total RNA) [32, 33]. However, this approach has at least two significant limitations. First, RNA purification leaves out cellular factors that may impede RNA isolation and amplification. Second, accurate dilution up to single cell levels is technical challenging. Another approach is to use external spike-ins, such as ERCC [34]. However, this approach also has a number of limitations [35]. First, the spike-in probes typically have different molecular properties than the RNA molecules of interest. Second, the spike-in probes interact differently with respect to different molecular biology protocols. Furthermore, the dynamic range of spike-in sets like ERCC is often not optimized for the dynamic range of a typical single cell transcriptome (~10^3 to 10^4). As such, there is a great need to develop better-controlled methods for separating technical and biological variations. Considering these limitations, targeted approaches aimed at precise quantification of key pathways may provide more biological insights in some applications.

Lack of spatial-temporal context: Single cell DNA and RNA based assays often contain the following steps: cell isolation, cell sorting, and library preparation and sequencing. During this process, cells are isolated from their local environment and destroyed prior to profiling. These "snapshots" lose important contextual information regarding both, the cells' spatial environment, and the cells' position within a trajectory of dynamic behavior

[25]. Both sources of information are crucial to interpret the precise state of a cell at the time point of its isolation (and usually destruction).

Future directions

in situ transcriptomic analysis: To preserve spatial information, transcriptome can be profiled in situ in fixed cells and tissues, using either *in situ* hybridization (ISH) or sequencing. Single molecule florescence *in situ* hybridization (smFISH) provides a powerful tool for detecting individual transcripts [36, 37]. Using super-resolution microscopy [38, 39], this was extended to image over a dozen mRNA *in situ* regardless of transcript density [40]. More recently, a temporal barcoding scheme was developed that scales exponentially with the number of hybridization, called sequential FISH (seqFISH) [41]. In parallel, *in situ* sequencing methods were developed to directly sequence transcripts in tissue sections [42, 43], which has broad coverage but lower efficiency compared to FISH based methods. More recently, a Hamming distance 2 based error correcting barcode system called merFISH [44] was developed and can be applied to long transcripts (>3 kb). This technology has recently been extended to detect 130 mRNA species [45]. Fundamentally, because of high background in tissues, smFISH based methods are difficult to apply directly for detection of mRNAs in tissues.

An amplified version of seqFISH [46], based on hybridization chain reaction (HCR) [47], allows robust detection of mRNAs in tissues and thick cleared brain samples. Combining amplification and a simple one-drop tolerant error correction scheme, this technology was applied to profile up to 249 genes, with each mRNA detected at ~80% efficiency, in over 15,000 cells in the mouse brain to resolve the structural organization of the hippocampus with single-cell resolution [48]. The authors identified distinct layers in the dentate gyrus corresponding to the granule cell layer and the subgranular zone. They also found that the dorsal CA1 is relatively homogeneous at the single-cell level, while ventral CA1 is highly heterogeneous. For imaging large samples, such as the brain, imaging speed is rate limiting, rather than the switching time between hybridizations. This is because one can toggle between two samples on the

microscope, one that is being imaged and another that is being hybridized. Faster imaging modality such as lattice lightsheet [49] and faster cameras can enable higher throughput in the number of imaged cells.

Future work in spatial genomics will take several directions. First, to combine spatial transcriptome data with scRNA-seq data, one can take an approach where cell states are defined by RNA-seq, and then mapped onto the spatial images and transcription profiles determined by spatial transcriptome data [50]. Second, to increase the optical space available in each cell and allow more mRNAs to be resolved spatially, expansion microscopy [51] can physically enlarge the tissue sample. An alternative image correlation approach [52] can also allow dense transcripts to be decoded. Lastly, analysis of *in situ* transcriptomic data requires development of new computational methods, for example, to automatically detect spatial patterns from combination of multiple genes.

Live imaging transcriptomic analysis:

Cellular and molecular behaviors are highly dynamic and constantly changing. These dynamic behaviors greatly complicate the interpretation of snapshot single-cell analyses because individual cells will differ not only in their molecular state from other cells, but even from themselves if analyzed at a different time point [25]. Importantly, these dynamics may not represent noise, but rather a basis for important regulatory mechanisms controlling cell identity, so it is important to quantify dynamic changes and to understand their relevance [53]. Unfortunately, it is also much more difficult than static snapshot analyses. Cells must be kept alive and unchanged during the continuous – and sometimes very long non-invasive analysis of their behaviors. The acquisition, handling and analysis of time-resolved single-cell data then require specialized technical and theoretical approaches. Not only are the requirements for robustness of data acquisition technologies such as live imaging much higher than for snapshot analyses, but the resulting large and complex volumes of data require specialized solutions. These differ from tools available to analyze snapshot data, and

often require self-made custom developments. This holds true for the required theoretical algorithms and for user-friendly implementation [25].

Lineage tracing:

The objective of lineage tracing is to label the progeny of individual cells using molecular markers and use such information to reconstruct the developmental trajectories. Recently, high-throughput lineage-tracing methods have been developed by CRSIPR/Cas9-based multiplexing DNA barcodes synthesis [54-58]. These barcodes are stably registered in the genome and inherited during cell division and differentiation. Additional mutations are cumulated in time, through either combinatorial editing at multiple gRNA target loci [54, 55, 57] or by sequential editing at a single locus [56, 58]. In the latter approach, the investigators introduced genetic mutations at the *S. pyogenes* gRNA-encoding sequence to circumvent the requirement of PAM motif in gRNA recognition, enabling the resulting gRNA to repeatedly target its own locus. In addition, the DNA barcodes can be sequenced *in situ*, thereby preserving the spatial information [58]. Some of the aforementioned technologies have been applied to study developmental loci [54, 55, 57] and immune response [56]. In one study [54], the investigators traced the cell lineages in zebrafish, and found that the majority of cells in each organ are derived from a small number of progenitor cells, whereas different progenitors are biased toward different germ layers and organs. Similar results are reported in an independent study [55]. These lineage tracing technologies will likely have wide-range applications in mapping developmental and disease progression trajectories.

Single-cell multi-omics: While significant effort has been dedicated to improving the quality and throughput of various omic assays, work is also ongoing to develop methods to profile multiple sources of information in the same cells. Multi-omics profiling is valuable for accurate mapping of cell states and can provide insights into the regulatory mechanisms. For example, genomic DNA and mRNA transcripts from the same cells

can be quantified by either physical separation [59] or pre-amplification [60], followed by high throughput sequencing. In the former, extracted genomic DNA can be further processed to bisulfite conversion, leading to simultaneous quantification of the methylome and transcriptome [61, 62]. Bioinformatic analysis of the bisulfite sequencing data can further detect genetic information [63, 64]. Protein and transcriptome have also been measured in the same cells [65]. Multi-omic methods applied to single cells have revealed some surprises. For example, profiling DNA and RNA variability in single acute lymphoblastic leukemia cells suggests that genetic heterogeneity is not responsible for the diverse response of drug treatment (Enver, unpublished).

Recent technologies have moved even beyond single cells to investigate sub-cellular localization of biologically active molecules. For example, nanoliter-scale cell fractionation or micro-manipulation has been applied to measure subcellular information within single cells [66]. On a different front, super-resolution imaging has been applied to map the nuclear compartmentalization of chromatin domains [67]. These subcellular data provide new insights into the precise mechanisms of various cellular processes. Ultimately, we may be able to understand phenotypic differences between genetically identical cells in terms of such variations in subcellular organization.

Modeling and predictions: Different cell types usually arise from a linear hierarchy of differentiation stages, and one goal of single-cell analysis is to identify previously unknown cell types and lineage relationships. Numerous methods have been developed to isolate similar cell types from single-cell gene expression data [7, 50, 68, 69]. Furthermore, additional methods have been developed to specifically detect rare cell subpopulations [70, 71]. To compensate the dropout effect, methods have also been developed to impute gene expression based on similar cell types [72].

Single-cell analysis has helped refine traditional views of cell differentiation. For example, A number of studies [14-16, 73] report evidence to suggest that megakaryocytes emerge at a "high" level, approximating that of the hematopoietic stem

cell (HSC); this insight challenges the prevailing model that the megakaryocytic lineage emerges late in the differentiation cascade. The cell states defined by transcriptomic patterns are surprisingly continuous instead of forming distinct, transcriptionally defined groups [15, 74]. This apparent continuity of cell states poses practical challenges for cell annotation, and conceptually, implies a need for significant revisions to current models of cell lineage hierarchy.

Data from single-cell studies have enabled the development of mathematical models that represent the distribution of cell states as one sampled from a dynamical system [11, 17, 19, 73]. In this view, cell types are modeled as "attractors" [75], stable states that are determined by the underlying gene regulatory networks and sometimes referred to as the energy landscape. In some models, stochastic fluctuation, either due to intrinsic or extrinsic noise, may facilitate dispersion and transition between attractors [11]. Although complex, these mathematical models can be used not only to explain the continuity of cell states but, in some cases, predict the initiation events during cell differentiation, thereby providing mechanistic insights [11]. In a similar way, the hierarchy of cell states can be measured by entropy, which has been applied to inform cell differentiation directions [76, 77]. These new methods have opened up new ways to think about cell states, not as discrete entities, but as a continuum. To connect these two viewpoints, it is critical to determine with high precision the level of natural variation that defines the same cell type and distinguish this from the changes linked to functional state transitions. A major obstacle for achieving this goal is that the resolution of cell-state identification is limited by the quality of the underlying scRNA-seq data, which varies greatly depending on sequencing depth and other factors. Such differences have contributed to the debate over the organizing structure of hematopoietic lineage hierarchy [15, 16].

Functional validation. As single cell data continue to grow in quality and quantity, new cell states, lineages, and associated markers are being identified at an accelerated rate. It is important to recognize that such findings are typically based on correlative analyses

and that their functional relevance needs to be carefully evaluated through further experimental validation.

A first level of validation is to utilize the identified markers to label the predicted cell-type and visualize in their original tissue. For example, unsupervised clustering of 25,000 single-cell transcriptomes identified 15 types of bipolar neurons [8]. The authors identified cell-type specific markers and fluorescently labeled the predicted cell-types by DNA FISH. They found that the spatial organization of these predictive cell-types is restricted to definitive layers and that different cell-types display distinct morphology, thereby supporting their functional identity.

A deeper level of validation requires design of functional assays to demonstrate that a predicted cell type has unique properties. For example, single-cell analysis showed that common myeloid progenitors (CMP) occur in two varieties associated with differential expression of CD55 [14]. Using an *in vitro* colony forming assay, the investigators found that CD55+ CMP produce predominantly erythroid and megakaryocytic (MegE) colonies, whereas few MegE colonies are formed from CD55- CMP, indicating these two subpopulations are functionally different. A similar strategy has been applied to compare the functional difference between HSC subpopulations, termed MolO and NoMO respectively [78]. These investigators found that MolO cells were enriched for higher than average CD150 and Sca-1 surface marker expression and lower than average CD48 expression.

In the same vein, engineered animal models can allow isolation of cell populations and functional testing. For example, comparative scRNA-seq analysis between HSCs from young and old mice identified a gene signature associated with the MegE lineage [79]. By using a transgenic mouse strain carrying a VWF-EGFP reporter, the authors verified an increased bias toward platelet-priming HSCs in old mice.

Combining scRNA-seq and CRISPR/cas9 based perturbations

CRISPR/Cas9 based genetic screens have been widely used to systematically characterize gene functions [80, 81]. Recently, this technique has been combined with scRNA-seq analysis [82-85], thereby greatly increasing the throughput of functional readouts. In these studies, gene activities are disrupted by either genetic mutations [82, 83, 85] or epigenetic inhibition [84]. gRNA-specific reporter transcripts are synthesized, which can be detected along with the mRNAs by scRNA-seq sequencing. By varying the concentration of the gRNA-containing vectors, the technique can be used to study the gene function either in isolation or in combination. In one study [82], the investigators applied this technology to analyze the effects of 24 TFs in mediating the immune response of dendritic cells. They found that the TFs form distinct modules each targeting a common set of gene program. Further analysis detected significant genetic interactions among a subset of TFs. In another study [83], the engineered hematopoietic progenitor cells were injected into wild-type recipient mice to evaluate their effect in hematopoiesis. This allowed them to identify a previously unknown role of Cebpb in regulating the balance between dendritic cells and monocytes during development. The combination of genome editing and scRNA-seq profiling provides a powerful tool for high-throughput dissection of gene functions and will have a wide range of applications in biomedical research.

Disease applications . Genomic profiling has been widely used to identify markers, mechanisms, and therapeutic targets of diseases. Most studies to date identify disease related alterations by comparing genomic profiles obtained from bulk disease samples and their normal counterparts. However, these average profiles provide a distorted view of the disease sample if it contains significant cellular heterogeneity, as in cancer. Single-cell technologies have provided a set of powerful tools to dissect the cellular heterogeneity and led to important discoveries in cancer [23, 86-89] and other diseases [90, 91]. For example, multiplexing qPCR analysis has identified subtypes of leukemia cells with distinct capacity of proliferation [87, 92]. Application of scRNA-seq to cancer also led to identification of rare subpopulations associated with drug resistance [23] or self-renewal [89], whereas scDNA-seq can be used to reconstruct paths of clonal

evolution [21]. Single cell profiling will provide new opportunities for mechanistic understanding of the initiation and progression of human diseases and to develop novel treatment methods targeting specific cell types.

Interdisciplinary research: We recognize that to overcome each challenge requires significant resources of lab and computational infrastructure. To move forward, the field needs groups of people with diverse expertise to work together. Interdisciplinary approaches are recognized to be important, if not a crucial prerequisite, for addressing many open questions, but also come with numerous challenges. First, the lack of expertise for parts of an interdisciplinary collaboration requires increased effort and tie for communication. The importance of a common language is well known, but remains a significant problem in almost every new project. Ideally, this hurdle will be overcome by a new generation of students and postdocs who are educated in multiple disciplines like biology/medicine, engineering and theoretical sciences. Interdisciplinary science, while leading to higher long-term impact, tends to be slower and published in journal of lesser impact [93], and is hard to organize and fund . Thus, it needs more patience, in particular in environments with funding cycles that require fast short-term output. Finally, not only language, but also career paths, scientific and publication cultures, hiring procedures and age, and scientific talent and academic motivation vary widely across disciplines. While many of these differences pose managerial challenges and should not impact scientific merit, in reality they often are the reason for failures of interdisciplinary endeavors. Overcoming these problems will require changes in teaching, funding, publication and hiring procedures, which would benefit most areas of science, but will only have a measurable effect after a few years.

Conclusions

Single-cell analysis is an exciting and rapidly expanding field that holds tremendous potential to improve our understanding of fundamental biological problems and to better understand the nature and complexity of human disease in order to develop more

effective therapies. To achieve these ambitious goals, proper control needs to be taken to warrant the detection of genuine heterogeneity existing in cell population and tissue samples. In addition, we need to invest in development of new methods. Single-cell data presents a number of intrinsic challenges, including systematic noise, the features of biological systems, and the sparsity and complexity of the data. The past few years have witnessed remarkable growth in the field, a trend we believe will continue, enabling more rigorous development of methods and deeper understanding of biological complexity.

Acknowledgement

Competing Interests

The authors declare that they have no competing interests.

Funding

Authors' Contributions

GCY conceived the study. All authors participated in the discussions and writing of the manuscript.

Availability of Data and Materials

Not applicable.

Ethics Approval

Not applicable.

References

1. Eberwine J, Sul JY, Bartfai T, Kim J: **The promise of single-cell sequencing.** *Nat Methods* 2014, **11:**25-27.
2. Blainey PC, Quake SR: **Dissecting genomic diversity, one cell at a time.** *Nat Methods* 2014, **11:**19-21.
3. Sandberg R: **Entering the era of single-cell transcriptomics in biology and medicine.** *Nat Methods* 2014, **11:**22-24.
4. Spitzer MH, Nolan GP: **Mass Cytometry: Single Cells, Many Features.** *Cell* 2016, **165:**780-791.
5. Zenobi R: **Single-cell metabolomics: analytical and biological perspectives.** *Science* 2013, **342:**1243259.
6. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR: **Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq.** *Nature* 2014, **509:**371-375.
7. Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C, et al: **Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.** *Science* 2015, **347:**1138-1142.
8. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, et al: **Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics.** *Cell* 2016, **166:**1308-1323 e1330.
9. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.** *Nat Biotechnol* 2014, **32:**381-386.
10. Bendall SC, Davis KL, Amir el AD, Tadmor MD, Simonds EF, Chen TJ, Shenfeld DK, Nolan GP, Pe'er D: **Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.** *Cell* 2014, **157:**714-725.
11. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, Yuan GC: **Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape.** *Proc Natl Acad Sci U S A* 2014, **111:**E5643-5650.
12. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, Choi K, Bendall S, Friedman N, Pe'er D: **Wishbone identifies bifurcating developmental trajectories from single-cell data.** *Nat Biotechnol* 2016, **34:**637-645.
13. Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ: **Diffusion pseudotime robustly reconstructs lineage branching.** *Nat Methods* 2016, **13:**845-848.
14. Guo G, Luc S, Marco E, Lin TW, Peng C, Kerenyi MA, Beyaz S, Kim W, Xu J, Das PP, et al: **Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire.** *Cell Stem Cell* 2013, **13:**492-505.
15. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, et al: **Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.** *Cell* 2015, **163:**1663-1677.

16. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, Grimes HL: **Single-cell analysis of mixed-lineage states leading to a binary cell fate choice.** *Nature* 2016, **537:**698-702.

17. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, Endele M, Filipczyk A, Gambardella A, Ahmed N, et al: **Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios.** *Nature* 2016, **535:**299-302.

18. Kim TK, Sul JY, Peternko NB, Lee JH, Lee M, Patel VV, Kim J, Eberwine JH: **Transcriptome transfer provides a model for understanding the phenotype of cardiomyocytes.** *Proc Natl Acad Sci U S A* 2011, **108:**11918-11923.

19. Kim J, Eberwine J: **RNA: state memory and mediator of cellular phenotype.** *Trends Cell Biol* 2010, **20:**311-318.

20. Ni X, Zhuo M, Su Z, Duan J, Gao Y, Wang Z, Zong C, Bai H, Chapman AR, Zhao J, et al: **Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients.** *Proc Natl Acad Sci U S A* 2013, **110:**21083-21088.

21. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, et al: **Clonal evolution in breast cancer revealed by single nucleus genome sequencing.** *Nature* 2014, **512:**155-160.

22. Gawad C, Koh W, Quake SR: **Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics.** *Proc Natl Acad Sci U S A* 2014, **111:**17947-17952.

23. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al: **Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.** *Science* 2016, **352:**189-196.

24. Bacher R, Kendziorski C: **Design and computational analysis of single-cell RNA-sequencing experiments.** *Genome Biol* 2016, **17:**63.

25. Skylaki S, Hilsenbeck O, Schroeder T: **Challenges in long-term imaging and quantification of single-cell dynamics.** *Nat Biotechnol* 2016, **34:**1137-1144.

26. Wagner A, Regev A, Yosef N: **Revealing the vectors of cellular identity with single-cell genomics.** *Nat Biotechnol* 2016, **34:**1145-1160.

27. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells.** *Nat Biotechnol* 2015, **33:**155-160.

28. Kharchenko PV, Silberstein L, Scadden DT: **Bayesian approach to single-cell differential expression analysis.** *Nat Methods* 2014, **11:**740-742.

29. Lun AT, Bach K, Marioni JC: **Pooling across cells to normalize single-cell RNA sequencing data with many zero counts.** *Genome Biol* 2016, **17:**75.

30. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al: **MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data.** *Genome Biol* 2015, **16:**278.

31. Hicks SC, Teng M, Irizarry RA: **On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data.** *bioRxiv* 2015.

32.   Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM, Wold BJ: **From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing.** *Genome Res* 2014, **24:**496-510.
33.   Dueck H, Khaladkar M, Kim TK, Spaethling JM, Francis C, Suresh S, Fisher SA, Seale P, Beck SG, Bartfai T, et al: **Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation.** *Genome Biol* 2015, **16:**122.
34.   Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, et al: **The External RNA Controls Consortium: a progress report.** *Nat Methods* 2005, **2:**731-734.
35.   Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA: **Power analysis of single-cell RNA-sequencing experiments.** *Nat Methods* 2017.
36.   Femino AM, Fay FS, Fogarty K, Singer RH: **Visualization of single RNA transcripts in situ.** *Science* 1998, **280:**585-590.
37.   Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S: **Stochastic mRNA synthesis in mammalian cells.** *PLoS Biol* 2006, **4:**e309.
38.   Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF: **Imaging intracellular fluorescent proteins at nanometer resolution.** *Science* 2006, **313:**1642-1645.
39.   Rust MJ, Bates M, Zhuang X: **Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM).** *Nat Methods* 2006, **3:**793-795.
40.   Lubeck E, Cai L: **Single-cell systems biology by super-resolution imaging and combinatorial labeling.** *Nat Methods* 2012, **9:**743-748.
41.   Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L: **Single-cell in situ RNA profiling by sequential hybridization.** *Nat Methods* 2014, **11:**360-361.
42.   Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wahlby C, Nilsson M: **In situ sequencing for RNA analysis in preserved tissue and cells.** *Nat Methods* 2013, **10:**857-860.
43.   Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R, et al: **Highly multiplexed subcellular RNA sequencing in situ.** *Science* 2014, **343:**1360-1363.
44.   Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X: **RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.** *Science* 2015, **348:**aaa6090.
45.   Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X: **High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization.** *Proc Natl Acad Sci U S A* 2016, **113:**11046-11051.
46.   Shah S, Lubeck E, Schwarzkopf M, He TF, Greenbaum A, Sohn CH, Lignell A, Choi HM, Gradinaru V, Pierce NA, Cai L: **Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing.** *Development* 2016, **143:**2862-2867.
47.   Choi HM, Chang JY, Trinh le A, Padilla JE, Fraser SE, Pierce NA: **Programmable in situ amplification for multiplexed imaging of mRNA expression.** *Nat Biotechnol* 2010, **28:**1208-1212.

48. Shah S, Lubeck E, Zhou W, Cai L: **In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus.** *Neuron* 2016, **92:**342-357.
49. Legant WR, Shao L, Grimm JB, Brown TA, Milkie DE, Avants BB, Lavis LD, Betzig E: **High-density three-dimensional localization microscopy across large volumes.** *Nat Methods* 2016, **13:**359-365.
50. Satija R, Farrell JA, Gennert D, Schier AF, Regev A: **Spatial reconstruction of single-cell gene expression data.** *Nat Biotechnol* 2015, **33:**495-502.
51. Chen F, Tillberg PW, Boyden ES: **Optical imaging. Expansion microscopy.** *Science* 2015, **347:**543-548.
52. Coskun AF, Cai L: **Dense transcript profiling in single cells by image correlation decoding.** *Nat Methods* 2016, **13:**657-660.
53. Etzrodt M, Endele M, Schroeder T: **Quantitative single-cell approaches to stem cell research.** *Cell Stem Cell* 2014, **15:**546-558.
54. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J: **Whole-organism lineage tracing by combinatorial and cumulative genome editing.** *Science* 2016, **353:**aaf7907.
55. Junker JP, Spanjaard B, Peterson-Maduro J, Alemany A, Hu B, Florescu M, van Oudenaarden A: **Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars.** *bioRxiv* 2016.
56. Perli SD, Cui CH, Lu TK: **Continuous Genetic Recording with Self-1 Targeting CRISPR-Cas in Human Cells.** *bioRxiv* 2016.
57. Schmidt ST, Zimmerman SM, Wang J, Kim SK, Quake SR: **Quantitative Analysis of Synthetic Cell Lineage Tracing Using Nuclease Barcoding.** *ACS Synth Biol* 2017.
58. Kalhor R, Mali P, Church GM: **Rapidly evolving homing CRISPR barcodes.** *Nat Methods* 2017, **14:**195-200.
59. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM, et al: **G&T-seq: parallel sequencing of single-cell genomes and transcriptomes.** *Nat Methods* 2015, **12:**519-522.
60. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A: **Integrated genome and transcriptome sequencing of the same cell.** *Nat Biotechnol* 2015, **33:**285-289.
61. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood SA, Ponting CP, Voet T, et al: **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity.** *Nat Methods* 2016, **13:**229-232.
62. Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang CY, Xue Z, Fan G: **Simultaneous profiling of transcriptome and DNA methylome from a single cell.** *Genome Biol* 2016, **17:**88.
63. Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, Wen L, Tang F, Huang Y, Peng J: **Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas.** *Cell Res* 2016, **26:**304-319.
64. Cheow LF, Courtois ET, Tan Y, Viswanathan R, Xing Q, Tan RZ, Tan DS, Robson P, Loh YH, Quake SR, Burkholder WF: **Single-cell multimodal profiling reveals cellular epigenetic heterogeneity.** *Nat Methods* 2016, **13:**833-836.

65. Genshaft AS, Li S, Gallant CJ, Darmanis S, Prakadan SM, Ziegler CG, Lundberg M, Fredriksson S, Hong J, Regev A, et al: **Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction.** *Genome Biol* 2016, **17:**188.

66. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A: **Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution.** *Genome Biol* 2015, **16:**20.

67. Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X: **Spatial organization of chromatin domains and compartments in single chromosomes.** *Science* 2016, **353:**598-602.

68. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir el AD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, et al: **Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis.** *Cell* 2015, **162:**184-197.

69. Xu C, Su Z: **Identification of cell types from single-cell transcriptomes using a novel clustering method.** *Bioinformatics* 2015, **31:**1974-1980.

70. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A: **Single-cell messenger RNA sequencing reveals rare intestinal cell types.** *Nature* 2015, **525:**251-255.

71. Jiang L, Chen H, Pinello L, Yuan GC: **GiniClust: detecting rare cell types from single-cell gene expression data with Gini index.** *Genome Biol* 2016, **17:**144.

72. Prabhakaran S, Azizi E, Carr A, Pe'er D: **Dirichlet Process Mixture Model for Correcting Technical Variation in Single-Cell Gene Expression Data.** *Journal of Machine Learning Research* 2016, **48:**1070-1079.

73. Haas S, Hansson J, Klimmeck D, Loeffler D, Velten L, Uckelmann H, Wurzer S, Prendergast AM, Schnell A, Hexel K, et al: **Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors.** *Cell Stem Cell* 2015, **17:**422-434.

74. Macaulay IC, Svensson V, Labalette C, Ferreira L, Hamey F, Voet T, Teichmann SA, Cvejic A: **Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells.** *Cell Rep* 2016, **14:**966-977.

75. Kauffman SA: **Metabolic stability and epigenesis in randomly constructed genetic nets.** *J Theor Biol* 1969, **22:**437-467.

76. Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H, et al: **De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data.** *Cell Stem Cell* 2016, **19:**266-277.

77. Teschendorff AE: **Single-cell entropy for quantification of differentiation potency from a cell's transcriptome.** *bioRxiv* 2016.

78. Wilson NK, Kent DG, Buettner F, Shehata M, Macaulay IC, Calero-Nieto FJ, Sanchez Castillo M, Oedekoven CA, Diamanti E, Schulte R, et al: **Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations.** *Cell Stem Cell* 2015, **16:**712-724.

79. Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, Macaulay I, Mancini E, Luis TC, Mead A, et al: **Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells.** *Nat Commun* 2016, **7:**11075.

80. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F: **Genome-scale CRISPR-Cas9 knockout screening in human cells.** *Science* 2014, **343:**84-87.

81. Wang T, Wei JJ, Sabatini DM, Lander ES: **Genetic screens in human cells using the CRISPR-Cas9 system.** *Science* 2014, **343:**80-84.

82. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al: **Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens.** *Cell* 2016, **167:**1853-1866 e1817.

83. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I: **Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq.** *Cell* 2016, **167:**1883-1896 e1815.

84. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al: **A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response.** *Cell* 2016, **167:**1867-1882 e1821.

85. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C: **Pooled CRISPR screening with single-cell transcriptome readout.** *Nat Methods* 2017, **14:**297-301.

86. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al: **Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.** *Science* 2014, **344:**1396-1401.

87. Saadatpour A, Lai S, Guo G, Yuan GC: **Single-Cell Analysis in Cancer Genomics.** *Trends Genet* 2015, **31:**576-586.

88. Tsoucas D, Yuan GC: **Recent Progress in Single-Cell Cancer Genomics.** *Current Opinion in Genetics & Development* 2017.

89. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, et al: **Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma.** *Nature* 2016, **539:**309-313.

90. van den Bos H, Spierings DC, Taudt AS, Bakker B, Porubsky D, Falconer E, Novoa C, Halsema N, Kazemier HG, Hoekstra-Wakker K, et al: **Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons.** *Genome Biol* 2016, **17:**116.

91. Gaudilliere B, Fragiadakis GK, Bruggner RV, Nicolau M, Finck R, Tingle M, Silva J, Ganio EA, Yeh CG, Maloney WJ, et al: **Clinical recovery from surgery correlates with single-cell immune signatures.** *Sci Transl Med* 2014, **6:**255ra131.

92. Saadatpour A, Guo G, Orkin SH, Yuan GC: **Characterizing heterogeneity in leukemic cells using single-cell gene expression analysis.** *Genome Biol* 2014, **15:**525.

93. Van Noorden R: **Interdisciplinary research by the numbers.** *Nature* 2015, **525:**306-307.