Biophysical *Journal*
Biophysical Perspective

Biophysical Society

# Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics

Sabyasachi Dasgupta,[1,*] Gary D. Bader,[2,3,*] and Sidhartha Goyal[1,4,*]
[1]Department of Physics, [2]The Donnelly Centre, [3]Department of Molecular Genetics, and [4]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada

ABSTRACT Single-cell genomics has recently emerged as a powerful tool for observing multicellular systems at a much higher level of resolution and depth than previously possible. High-throughput single-cell RNA sequencing techniques are able to simultaneously quantify expression levels of several thousands of genes within individual cells for tens of thousands of cells within a complex tissue. This has led to development of novel computational methods to analyze this high-dimensional data, investigating longstanding and fundamental questions regarding the granularity of cell types, the definition of cell states, and transitions from one cell type to another along developmental trajectories. In this perspective, we outline this emerging field starting from the "input data" (e.g., quantifying transcription levels in single cells), which are analyzed to define "identities" (e.g., cell types, states, and key genes) and to build "interactions" using models that can infer relations and transitions between cells.

Single-cell-level experiments aim to characterize the rich heterogeneity in cell populations within complex tissues. The cell-level information observed in such studies provides valuable insights into the dynamics of important biological processes, such as cellular differentiation during development or tissue regeneration. Ideally, a wide range of molecules could be measured with high sensitivity and coverage at high cellular resolution, but this has not been possible. Recently, single-cell genomics technologies have been developed that represent a major advance toward this ideal (1). How will these new technologies change the field of biophysics?

Traditionally, fluorescence-imaging methods provided static snapshots of the complexity of cells at the molecular scale, which, when coupled with fast super-resolution imaging techniques and live imaging in the past decade, exposed the rich dynamics at play within cells. Light-based or mass-cytometry methods are increasingly used to simultaneously quantify expression levels for multiple proteins within many cells. However, these methods typically remain limited to a maximum of a few dozen distinct molecules being studied simultaneously (2). On the other end of the spectrum, traditional transcriptomics technologies, such as RNA-seq, provide quantitative measures of expression for thousands of genes, thereby providing a rich characterization of complex phenotypes. Although these transcriptome technologies provide a "high-dimensional" characterization, they measure gene expression that is averaged over a large population of cells and hence do not provide direct information about individual cells or subpopulations that are present in a complex tissue. New single-cell genomic methods combine the best of both of these worlds and enable simultaneous study of thousands of cells and expression levels (mRNA counts) of thousands of genes or tens to hundreds of proteins, as well as chromatin state and genome sequence, within each cell (3–6). These single-cell genomic technologies challenge us to understand how gene expression in the high-dimensional space of thousands of genes maps distinct cell "identities" (cell types or states) and, more generally, phenotypes. Ideally, we can use single-cell genomics information to identify all cell types and states in a tissue and their genomic properties and infer cell lineages, cell-cell communication networks, and spatial tissue maps. This exciting technology advance has already led to major international scientific efforts, such as the Human Cell Atlas, which seeks to map all cell types in the human body at unprecedented resolution (https://www.humancellatlas.org/).

This perspective helps navigate readers from a range of biophysical fields embarking on the relatively new and promising field of single cell genomics, analytics, and modeling. We organize experimental and analysis methods into three major categories: input, identity, and interactions. Gene expression counts measured by the experiments are the "input" followed by statistical analysis to define

*Submitted February 1, 2018, and accepted for publication July 3, 2018.*
*Correspondence: sabyadg.softbio@gmail.com or gary.bader@utoronto.ca or goyal@physics.utoronto.ca
Editor: Brian Salzberg.
https://doi.org/10.1016/j.bpj.2018.07.003

Please cite this article in press as: Dasgupta et al., Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics, Biophysical Journal (2018), https://doi.org/10.1016/j.bpj.2018.07.003



"identities," such as distinct cell types or states defined by their characteristic gene expression patterns. Finally, "interactions" among genes and cells, such as cell state transitions, can be computationally modeled to predict system dynamics over time.

## Input

The technological race to measure the transcriptome (the set of all expressed transcripts) of a large number of cells in a single experiment has led to both in-house laboratory scale platforms (7) (e.g., drop-seq (3)) and commercially available platforms (e.g., chromium from 10× Genomics, Pleasanton, CA; InDrop from 1CellBio, Cambridge, MA; and μEncapsulator from Dolomite Bio, Royston, UK/Blacktrace Holdings, Royston, UK). Before using these technologies, single cells must be experimentally extracted from a sample, such as a tissue. Once the dissected tissue is obtained, single cells are isolated using mechanical or other solution-suspension methods and may also be subject to sorting, e.g., based on fluorescence-activated cell sorting. Cell extraction methods must be designed for each tissue type and ideally must limit the number of experimental steps to reduce the chance of damaging cells. Some cells are more sensitive to dissociation conditions than others, and cells can be transcriptionally affected by extraction conditions, including temperature, time exposed to reagents, and stress of a sorting technique (8). Single cells are then prepared for genomics measurements. Current popular methods for single-cell RNA-seq (scRNA-seq) use microfluidic technology to encapsulate cells within droplets or microwells (9) along with DNA-barcoded beads displaying library primers. Once encapsulated with a cell-specific barcoded sequencing library, cells are lysed, and transcripts hybridize to the bead and are amplified and sequenced. Each transcript is also barcoded using a unique molecular identifier to normalize amplification artifacts. Single-cell genomic technology development is progressing extremely rapidly, and new experimental methods are published regularly. In the final step, a digital expression matrix (transcript counts by cell) is obtained by using bioinformatics approaches to align the unique-molecular-identifier-associated sequences to known genes and to count the transcripts (Fig. 1, *inset*). Standard computational workflows are provided by commercial vendors to complete this step, and standard analysis workflows are available for data normalization (10) and analysis (11). Databases of single-cell sequencing matrices are available online with data sets ranging from a few hundred to over 1 million cells (e.g., https://portals.broadinstitute.org/single_cell).

The gene-by-cell expression count matrix can be used to answer a variety of questions, including clustering to identify cell types and states. Although valuable information is often identified, multiple technical challenges are still open, and

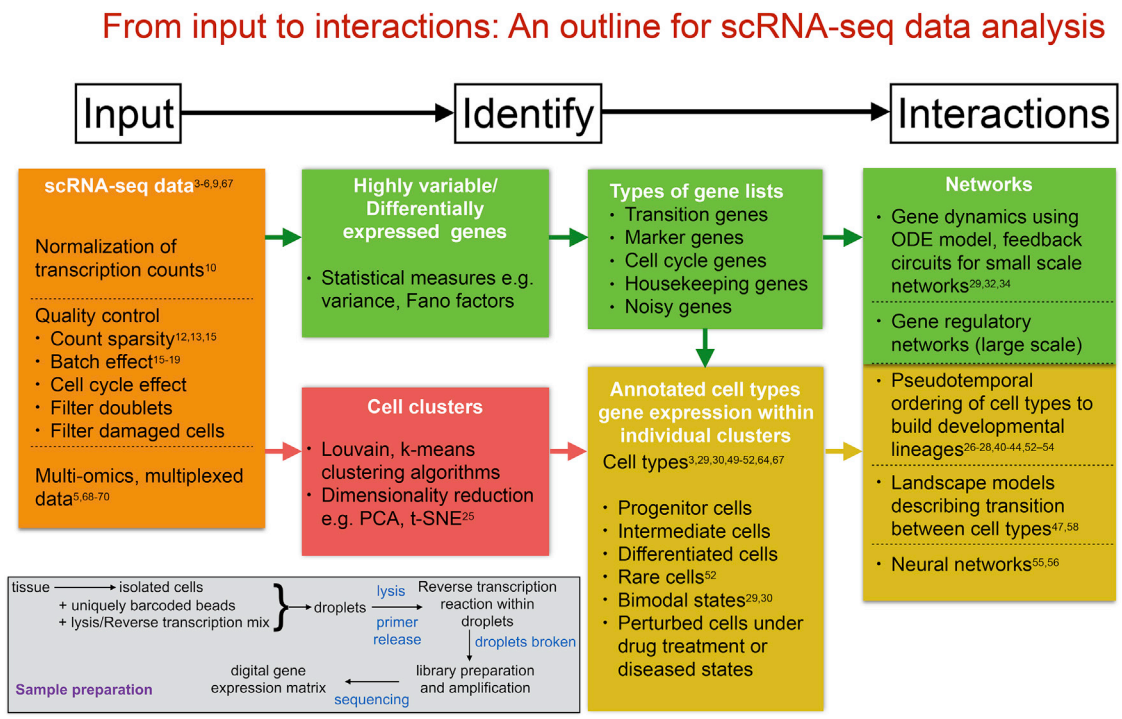


FIGURE 1 Overview of single-cell-genomic data generation and evaluation, showing examples of experimental and computational analysis steps and specific methods along with citations. Sample generation uses a specific tissue to obtain single cells for analysis and ultimately a digital gene-expression matrix of transcript counts per gene per cell. This matrix is analyzed to identify relevant genes and cell types and to describe the complex relationships between these identities. These models can be used, for instance, to select genes to alter and control cellular transitions. To see this figure in color, go online.

Please cite this article in press as: Dasgupta et al., Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics, Biophysical Journal (2018), https://doi.org/10.1016/j.bpj.2018.07.003



analysis methods are actively being developed to improve performance and extract more knowledge from current data sets. Current single-cell transcriptomics methods cannot identify all transcripts within each cell due to biological or technical factors, and the resulting expression matrix is typically sparse, containing mostly zeros, reducing sensitivity for cell type and state identification. Biological factors include temporal effects, e.g., cell cycle progression, spatiotemporal differences in developing tissues, or stochastic transcription fluctuations such that the expression is below the sensitivity of the experimental approach. Technical factors include shallow sequencing of the sample and sequencing errors (called dropouts) arising from the tagging effects, allelic dropout due to low amplification by polymerase chain reaction, experimental design (e.g., doublet contamination), instrument errors, and errors in genome alignment. Statistical methods to model these processes and improve the signal-to-noise ratio are being developed (12,13). Transcript counts per gene and library size (the total number of reads per cell) vary from one cell to another, causing challenges for normalization methods. Popular normalization tools exist (10), but there is no consensus of approach because of the large variability in reads obtained across different cells in typical experiments. Ideally, we would know the expected transcriptional profile of each cell type under each state and, as a result of different sample preparation conditions, enable appropriate correction of confounding factors. However, correlation studies between data or cell types collected under different conditions and from different labs remain few, though this is likely to change in the next few years. Thus, another important challenge is to compare different single-cell sequencing data sets or combine them in an unbiased way postnormalization. Data originating from similar cells and tissues can vary because of experimental conditions such as administration of external drugs or growth factors, morphogens present within a developing tissue, and natural progression toward disease and aging. We may address this variability by integrating our knowledge of how cells progress along a differentiation axis or transition to different identities as a response to perturbations in gene-expression patterns at the input level or postidentification of genes and cell types (identities) from each data set or during the final stage of inference and model building (14).

Even in homogenous populations and in a priori identical experimental conditions, cell-to-cell variability in scRNA-seq count measurements is widespread. Fortunately, scRNA-seq provides fast and accurate measurements over a large sample of cells, thereby providing an ensemble of measurements ideally over the same candidate cell type. This provides a reasonable number of independent measurements to measure statistical quantities, such as mean or variance of transcription levels across different cells. This also addresses the sparsity problem, as combining sparse, stochastic measurements over many cells results in a rich set of measurements for a cluster of similar cells. Thus, on the one hand, technology is evolving toward better statistics, but on the other hand, experimental data have "inherent biological noise" that may be assigned to hidden confounding variables that may also vary from cell to cell. A direct statistical approach used traditionally to address such noise is to compare technical replicates, which is so far challenging for scRNA-seq, as the experiments are costly and destroy the cells upon measurement. In principle, scRNA-seq provides many samples of an individual cell type, and these could be considered replicates but in the traditional sense are biological replicates. Considering approaches from biophysics, such noise may have physical signatures that can be identified as systematic errors or may simply emerge from randomness among biological samples collected at different times. Statistical methods can be used to remove systematic noise, such as batch effects, and this is an open area of investigation (15–19). Biophysical phenomena, such as cell cycle variability or stochastic gene expression, may also lead to variable expression across cells. Quantifying fluctuations or noise in experimental data has previously proven useful in biophysics (20,21). One may, however, treat this as an "uncertainty of measurements" and generate a model-dependent estimate for a number of expected "zeroes" compared with actual experimental data sets (22). An example of this approach was recently used to estimate the total number of contributing blood stem cells from an undersampled data set (23). Such approaches may shed light on how much undersampling affects our current knowledge of expression matrices.

Once technical noise can be averaged or removed, fluctuating variables and quantifying the level of fluctuations may lead to insights about the dynamics of gene expression. Several recent attempts have been made using nonequilibrium physics to quantify the fluctuations of stochastic variables. Thus, "stochastic thermodynamics," which is a relatively new field in physics, may become useful to quantifying stochastic expression patterns in genes (24).

## Identity

The typical scRNA-seq workflow applies a clustering algorithm to the transcript count matrix to identify stable cell clusters (groups of similar cells) and then visualizes the data with the aid of exploratory linear or nonlinear dimensionality reduction tools, such as t-distributed stochastic neighbor embedding (25), phage annotation toolkit and evaluation (26), principal component analysis, and diffusion maps (27) (Fig. 2). Stable cell clusters represent cell types and are interpreted using known cell-type-specific gene expression markers to catalog the cell types identified in the experiment. Clusters of cells that do not match known markers are candidates for new cell types.

Cell clusters may have different phenotypes, which involve different subcellular actors (genes) for controlling their identity (*marker genes*) and transitions (*transition genes*). Genes that are "highly variable" across different

Please cite this article in press as: Dasgupta et al., Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics, Biophysical Journal (2018), https://doi.org/10.1016/j.bpj.2018.07.003



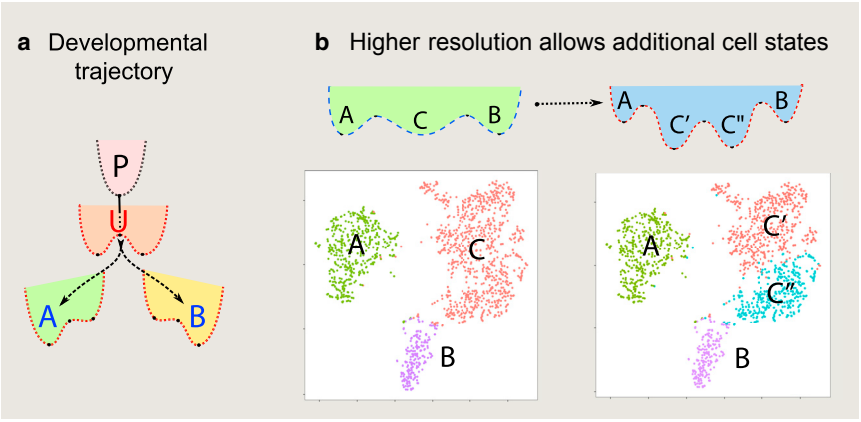


FIGURE 2 Landscape picture of states and transitions: cell types remain embedded within a high-dimensional gene expression landscape and are available as fixed points or stable states of a cost function accounting for gene expression variability and mutual interactions between genes. Development initiates from a stable precursor (or progenitor) state (P) and as development proceeds intermediate unstable states (U) enable novel stable cell states (A and B) to emerge. Transitions between these stable states occur via "tilting" the landscape (*a*) where cell type B transitions to cell type A. Such transitions can occur because of the presence of "fields" such as growth factors and morphogens or because of genetic mutations. (*b*) Depending on the resolution used to describe the model, one may be able to resolve further types of cells that appear as one group at lower resolutions. For example, cell type C might have a bimodal population C' and C", and each subpopulation might have a preferential transition to neighboring minima (stable cell types) such as C' to A and C" to B. Genetic and epigenetic changes may also be interpreted to create such changes in the landscape by creating additional cell types such as the splitting of one wide basin (C) to two subpopulations C' and C", thereby changing the developmental program of cells and lineage evolutions. t-SNE plots are shown for different resolutions, where points represent individual cell transcriptomes embedded in two-dimensional space, and point color corresponds to landscape regions. To see this figure in color, go online.

cells or clusters may be important for both cell identity and transitions between cell identities, and their identification is a key challenge. Ideally, comparing gene expression patterns across the clusters enables one to infer a set of transition and marker genes (28). Similarly, protein interaction and pathway data can be used to identify functions active in each cell or gene group and aid cluster interpretation, for example, whether two nearby clusters are functionally different because they express different genes and pathways. This approach is useful but suffers from the assumption that genes that do not express differentially and that are constitutively expressed (often referred to as housekeeping genes) are not crucial for describing cell types. Differential gene expression may also be attributed to cyclical expression patterns arising because of gene oscillations (e.g., over the cell cycle), which create bi- or multimodal populations for an assumed distinct cell type (29,30). Thus, multiple subclusters that manifest under different time-varying developmental fields (e.g., induced by morphogen gradients) may be present in a single cell cluster that is thought to be homogenous. Furthermore, cell-to-cell variability within a cluster may be an indicator of an early signal of other differentiated types in a Waddington-like landscape or suggest a heterogeneous biological context for cells of the same type.

Many cells from a typical scRNA-seq experiment do not easily fit into tight clusters. Single-cell RNA-seq measures the transcriptional state of each cell at a single time point from the very large theoretical space of possible states that could exist. An important debate to consider while analyzing single-cell genomics data is whether cells are best represented as stable, distinct types or as a fluid and dynamic space of states. We view this either as a continuum for dynamic (e.g., blood, gut) and developing tissues or as being more discrete for very stable postmitotic systems (e.g., neurons and rod photoreceptors). Both views are useful, and models must consider both to be generally applicable to natural systems.

## Interactions

Developing predictive mathematical and mechanistic models of dynamic biological processes is a major driver of new discoveries in biology and medicine. A key goal of this research is to predict cellular perturbations that can control the cell to perform engineered tasks, such as differentiating into useful cells for regenerative therapy, and to identify sensitive points in a cell that can be used to kill cancer cells using drugs. Although many successful examples of such work exist, major challenges include the following: 1) model construction requires substantial experimental effort; 2) most parameters are not available because they are difficult to measure; 3) and scaling up to whole cell models is difficult. The availability of large amounts of data is now known to enable the inference of scalable, predictive models, such as deep learning, that can overcome these challenges (31), though often without providing mechanistic insight. It would be useful to be able to accurately infer mechanistic and predictive models from large data, achieving the best of both approaches. Single-cell genomics technologies, such as scRNA-seq, may provide enough information about cellular processes to make this possible.

Mathematical models incorporating gene expression measurements have, until now, focused on small gene-regulatory networks and their effect on a restricted selection of cell identities based on empirical biological data by using ODE (ordinary differential equation) formalisms to capture the on/off and oscillating nature of transcription levels of a few candidate transcription factors (32–34). These methods inform how genetic circuits implement function via local bifurcations or oscillations between a few stable cell fates.

Please cite this article in press as: Dasgupta et al., Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics, Biophysical Journal (2018), https://doi.org/10.1016/j.bpj.2018.07.003



This has led to insights about modules of a few genes that control cellular behavior and to the recent revolution in synthetic biology that aims to design cells for particular functions (35,36). Although such models may be useful in the context of single-cell data, they do not realistically tackle the complexity and scale of the typical high-dimensional data that single-cell studies provide. Typically, such ODE-based techniques quickly become intractable for large dimensional data in which $10^3$ genes are sequenced within each cell for $10^3$–$10^4$ cells. Statistical tools can be used to simplify the problem. For example, key "identities" like cell types and gene sets (typically markers, transition genes, and pathways) found using the clustering and dimensionality reduction techniques discussed above can be used to build mathematical models similar to what has been done in past with ODE-based formulations.

Transitions between cell identities are important and can be used for modeling. Inferring cell type transitions from static single-cell expression data is now possible. Initial attempts inferred a cell lineage based on a minimal spanning tree across an undirected graph with weights defined by Euclidian distance in gene expression space (37). For single-cell-level data, the first attempt to describe a temporal ordering of expression pattern from a heterogeneous mixture of cells using mass-cytometry data was presented using the Spanning Tree Progression of Density-Normalized Events (38) and Wanderlust (39) methods. Wishbone (40), as an updated version of Wanderlust, produces developmental trajectories and identifies bifurcation patterns from cytometry data. Monocle (41) and Census, or Monocle 2.0 (42), use a method based on a nearest-neighbor graph to obtain lineage relationships between cells. Assuming a random-walk-like approach, one can obtain diffusion maps to embed the heterogeneous single-cell population in a connected landscape with temporal ordering in pseudotime (27). Although these methods rely on using the metric between clustered data (or cells) and then embedding in a graph, other distance measures based on ideas from physics and information theory, such as calculating single-cell entropy and then inferring an ordering using a Waddington-like landscape, were applied in single-cell lineage inference using cell expression similarity and entropy (43). Using tools of nonlinear dynamics, singe-cell clustering using bifurcation analysis (44) tries to obtain lineage trees by bifurcation analysis of the expression patterns (applicable to temporal data in which bifurcation points are defined by splitting one cluster into two). Several other tree-based algorithms have also been recently proposed (45–48). Linear combination of a few principal component analysis or diffusion components may help to distinguish cell subpopulations within these systems (49,50). Bayesian approaches or probabilistic methods (27,51–53) have an advantage over tree-based methods, as they can assign an uncertainty measure to the connectivity link between two distinct cell clusters. The minima or maxima expression pattern observed for genes within a triad of cell types is one such example that has been used to infer lineages (28). Recently, a novel developmental trajectory reconstruction algorithm was proposed called RNA Velocity (54), which uses the ratio of spliced and unspliced reads of each transcript within a cell to extrapolate the expression profile in the near future (up to a few hours), thereby quantifying cell identity dynamics within a single cell.

Both gene regulatory network models based on prior knowledge and inferred network models based on Bayesian methods usually rely on the assumption that high-dimensional data can be effectively projected to a low-dimensional space captured by a few genes or pathways. Although this might be a valid assumption given experimental evidence that only a handful of transcription factors can control important transitions between cell states (35), it may not always apply. And if indeed the cell cluster cannot be explained in a low-dimensional space and instead remains embedded in a moderate- to high-dimensional space ($10^2$–$10^3$ genes), then one needs frameworks that can support predictions about cellular transitions and temporal dynamics of cell clusters. Such frameworks may invoke ideas from dynamical systems in high-dimensional space, in which cell clusters are dynamic but can be defined locally in time. This approach has been successfully implemented on sequencing data sets using machine learning to build an algorithm that can learn and classify (55,56) and to describe states of memory in the context of neuroscience (57).

Using an analogy from thermodynamics, transitions across states occur because of barrier energy crossing, in which the "height of the barrier" modulates the probability of escape. In a "genetic landscape," the presence of inducing time-varying fields like morphogen gradients, which push toward a particular developmental fate, makes these barriers disappear, and possible transitions become facilitated under specific directed field vectors along certain landscape directions (58,59). The transition of one cell type to another may occur by use of existing basins or minimas, which appear as stable states accessed by tilting the landscape (Fig. 2 *a*) or by the splitting of available basins in additional subpopulations of cell states (Fig. 2 *b*). The notion of stable states and transitions among them conjures a Waddington-like picture of cell identities appearing as basins or valleys of a complex surface topography (60). For example, in tissue development, the position of an individual cell along this landscape may indicate its differentiation state (Fig. 2). The knowledge of the "barrier heights" and the relevant directions of change helps us to engineer cell fates (61). Certain machine-learning approaches have been used to make similar predictions by comparing a known target (e.g., endogenous cell transcriptome) and new data to quantify the closeness of the cell identity found in experiments with the desired engineering target (62,63). Such approaches may predict possible control points in time, such as time-varying fields (e.g. growth factors), that help fulfill the engineering goals of regenerative medicine to

Please cite this article in press as: Dasgupta et al., Single-Cell RNA Sequencing: A New Window into Cell Scale Dynamics, Biophysical Journal (2018), https://doi.org/10.1016/j.bpj.2018.07.003



grow a specific cell population. Landscape models constructed based on sequencing data from a single developmental time point already readily model multiple cell states and predict specific predictions for accessing individual states under certain conditions; however, it is far from obvious whether developmental dynamics over longer times can be accurately captured or if we need models in which additional temporal data can be incorporated.

## Outlook

Single-cell genomics is a powerful technique to measure cellular dynamics and gene expression variation across thousands to millions of cells. This technology represents a new kind of microscope, as it opens a new window into cellular complexity and can even be used in situ to map cell diversity across an entire tissue (64). Ideally, this enables us to find a simpler meaning in this complexity by looking for key genes, such as transcription or growth factors, that are responsible for cellular heterogeneity and to decode the time-resolved dynamics of cellular processes, such as differentiation or development (65). Another relevant dimension to consider is the spatial localization of cells in the context of a tissue or organ. With our current knowledge of the gene expression differences across cell types and temporal data providing time-ordered maps of these cells as they develop and function, it will be instrumental to build three-dimensional atlases in which individual cell types can be visualized in situ (64,66). This will help us to investigate how gene expression patterns manifest spatially in an organism and how local coordinates and geometry interplay with these patterns.

Important questions remain to be addressed, including the definition of cell types and cell states as well as how cells change among these configurations and the relevant timescales for such processes. The explosion of single-cell measurements at multiple omic levels coupled with novel technological advancements such as single-nuclei RNA sequencing from complex tissue (scNuc-seq) (67) and multimodal measurements (5,68–70) to elucidate the genotype-phenotype interplay in the context of cell function provides a major opportunity to answer these important biological questions. Answering these questions will require an interdisciplinary approach, including biophysics, computer science, biology, engineering, and other fields.

## ACKNOWLEDGMENTS

S.D., G.D.B., and S.G. acknowledge the support of grants from the Medicine by Design program funded by the Canada First Research Excellence Fund. S.G. also acknowledges the support of a Natural Sciences and Engineering Research Council discovery grant.

## REFERENCES

1. Tanay, A., and A. Regev. 2017. Scaling single-cell genomics from phenomenology to mechanism. *Nature.* 541:331–338.
2. Spitzer, M. H., and G. P. Nolan. 2016. Mass cytometry: single cells, many features. *Cell.* 165:780–791.
3. Macosko, E. Z., A. Basu, …, S. A. McCarroll. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell.* 161:1202–1214.
4. Klein, A. M., L. Mazutis, …, M. W. Kirschner. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell.* 161:1187–1201.
5. Stoeckius, M., C. Hafemeister, …, P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods.* 14:865–868.
6. Cusanovich, D. A., R. Daza, …, J. Shendure. 2015. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science.* 348:910–914.
7. Kalisky, T., S. Oriel, …, S. Pyne. 2018. A brief review of single-cell transcriptomic technologies. *Brief. Funct. Genomics.* 17:64–76.
8. van den Brink, S. C., F. Sage, …, A. van Oudenaarden. 2017. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods.* 14:935–936.
9. Gierahn, T. M., M. H. Wadsworth, II, …, A. K. Shalek. 2017. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods.* 14:395–398.
10. Lun, A. T., K. Bach, and J. C. Marioni. 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17:75.
11. Butler, A., P. Hoffman, …, R. Satija. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36:411–420.
12. Mohammadi, S., J. Davila-Velderrain, …, A. Grama. 2018. DECODEing sparsity patterns in single-cell RNA-seq. *bioRxiv* https://doi.org/10.1101/241646.
13. van Dijk, D., J. Nainys, …, D. Pe'er. 2017. MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. *bioRxiv* https://doi.org/10.1101/111591.
14. Karaiskos, N., P. Wahle, …, R. P. Zinzen. 2017. The Drosophila embryo at single-cell transcriptome resolution. *Science.* 358:194–199.
15. Hicks, S. C., F. W. Townes, …, R. A. Irizarry. 2017. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics*, Published online November 6, 2017 http://doi.org/10.1093/biostatistics/kxx053.
16. Haghverdi, L., A. T. L. Lun, …, J. C. Marioni. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36:421–427.
17. Shaham, U., K. P. Stanton, …, Y. Kluger. 2017. Removal of batch effects using distribution-matching residual networks. *Bioinformatics.* 33:2539–2546.
18. Buettner, F., N. Pratanwanich, …, O. Stegle. 2017. f-scLVM: scalable and versatile factor analysis for single-cell RNA-seq. *Genome Biol.* 18:212.
19. Tung, P. Y., J. D. Blischak, …, Y. Gilad. 2017. Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7:39921.
20. Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell.* 135:216–226.
21. Eldar, A., and M. B. Elowitz. 2010. Functional roles for noise in genetic circuits. *Nature.* 467:167–173.
22. Risso, D., F. Perraudeau, …, J.-P. Vert. 2017. ZINB-WaVE: a general and flexible method for signal extraction from single-cell RNA-seq data. *bioRxiv* https://doi.org/10.1101/125112.
23. Goyal, S., S. Kim, …, T. Chou. 2015. Mechanisms of blood homeostasis: lineage tracking and a neutral model of cell populations in rhesus macaques. *BMC Biol.* 13:85.
24. Seifert, U. 2012. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* 75:126001.

25. Van Der Maaten, L. J. P., and G. E. Hinton. 2008. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* 9:2579–2605.
26. Moon, K. R., D. van Dijk, …, S. Krishnaswamy. 2017. Visualizing transitions and structure for high dimensional data exploration. *bioRxiv* https://doi.org/10.1101/120378.
27. Haghverdi, L., F. Buettner, and F. J. Theis. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics.* 31:2989–2998.
28. Furchtgott, L. A., S. Melton, …, S. Ramanathan. 2017. Discovering sparse transcription factor codes for cell states and state transitions during development. *eLife.* 6:1–33.
29. Ochab-Marcinek, A., and M. Tabaka. 2010. Bimodal gene expression in noncooperative regulatory systems. *Proc. Natl. Acad. Sci. USA.* 107:22096–22101.
30. Shalek, A. K., R. Satija, …, A. Regev. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 498:236–240.
31. LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature.* 521:436–444.
32. Mitra, M. K., P. R. Taylor, …, B. Chakrabarti. 2014. Delayed self-regulation and time-dependent chemical drive leads to novel states in epigenetic landscapes. *J. R. Soc. Interface.* 11:20140706.
33. Rosenfeld, N., J. W. Young, …, M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science.* 307:1962–1965.
34. Elowitz, M. B., and S. Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature.* 403:335–338.
35. Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 126:663–676.
36. Mertens, J., M. C. Marchetto, …, F. H. Gage. 2016. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat. Rev. Neurosci.* 17:424–437.
37. Magwene, P. M., P. Lizardi, and J. Kim. 2003. Reconstructing the temporal ordering of biological samples using microarray data. *Bioinformatics.* 19:842–850.
38. Qiu, P., E. F. Simonds, …, S. K. Plevritis. 2011. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* 29:886–891.
39. Bendall, S. C., K. L. Davis, …, D. Pe'er. 2014. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell.* 157:714–725.
40. Setty, M., M. D. Tadmor, …, D. Pe'er. 2016. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 34:637–645.
41. Trapnell, C., D. Cacchiarelli, …, J. L. Rinn. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381–386.
42. Qiu, X., A. Hill, …, C. Trapnell. 2017. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods.* 14:309–315.
43. Guo, M., E. L. Bao, …, Y. Xu. 2017. SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Res.* 45:e54.
44. Marco, E., R. L. Karp, …, G. C. Yuan. 2014. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. USA.* 111:E5643–E5650.
45. Grün, D., M. J. Muraro, …, A. van Oudenaarden. 2016. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell.* 19:266–277.
46. Ji, Z., and H. Ji. 2016. TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44:e117.
47. Chen, J., A. Schlitzer, …, M. Poidinger. 2016. Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nat. Commun.* 7:11988.
48. Rizvi, A. H., P. G. Camara, …, R. Rabadan. 2017. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* 35:551–560.
49. Tirosh, I., A. S. Venteicher, …, M. L. Suvà. 2016. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature.* 539:309–313.
50. Scialdone, A., Y. Tanaka, …, B. Göttgens. 2016. Resolving early mesoderm diversification through single-cell expression profiling. *Nature.* 535:289–293.
51. Lönnberg, T., V. Svensson, …, S. A. Teichmann. 2017. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves $TH_1$/TFH fate bifurcation in malaria. *Sci. Immunol.* 2:1–10.
52. Buettner, F., K. N. Natarajan, …, O. Stegle. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33:155–160.
53. Reid, J. E., and L. Wernisch. 2016. Pseudotime estimation: deconfounding single cell time series. *Bioinformatics.* 32:2973–2980.
54. La Manno, G., R. Soldatov, …, P. Kharchenko. 2017. RNA velocity in single cells. *bioRxiv* https://doi.org/10.1101/206052.
55. Ghahramani, A., F. M. Watt, and N. M. Luscombe. 2018. Generative adversarial networks uncover epidermal regulators and predict single cell perturbations. *bioRxiv* https://doi.org/10.1101/262501.
56. Lin, C., S. Jain, …, Z. Bar-Joseph. 2017. Using neural networks for reducing the dimensions of single-cell RNA-seq data. *Nucleic Acids Res.* 45:e156.
57. Hopfield, J. J., and D. W. Tank. 1985. "Neural" computation of decisions in optimization problems. *Biol. Cybern.* 52:141–152.
58. Schiebinger, G.…, 2017. Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. *bioRxiv* https://doi.org/10.1101/191056.
59. Lang, A. H., H. Li, …, P. Mehta. 2014. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLoS Comput. Biol.* 10:e1003734.
60. Waddington, C. H. 1957. The Strategy of the Genes. Routledge, Abingdon, UK.
61. Briggs, J. A., V. C. Li, …, M. W. Kirschner. 2017. Mouse embryonic stem cells can differentiate via multiple paths to the same state. *eLife.* 6:1–23.
62. Bian, Q., and P. Cahan. 2016. Computational tools for stem cell biology. *Trends Biotechnol.* 34:993–1009.
63. Radley, A. H., R. M. Schwab, …, P. Cahan. 2017. Assessment of engineered cells using CellNet and RNA-seq. *Nat. Protoc.* 12:1089–1102.
64. Shah, S., E. Lubeck, …, L. Cai. 2016. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron.* 92:342–357.
65. Alemany, A., M. Florescu, …, A. van Oudenaarden. 2018. Whole-organism clone tracing using single-cell sequencing. *Nature.* 556:108–112.
66. Belle, M., D. Godefroy, …, A. Chédotal. 2017. Tridimensional visualization and analysis of early human development. *Cell.* 169:161–173.e12.
67. Habib, N., I. Avraham-Davidi, …, A. Regev. 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods.* 14:955–958.
68. Peterson, V. M., K. X. Zhang, …, J. A. Klappenbach. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–939.
69. Macaulay, I. C., W. Haerty, …, T. Voet. 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods.* 12:519–522.
70. Cadwell, C. R., F. Scala, …, A. S. Tolias. 2017. Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using Patch-seq. *Nat. Protoc.* 12:2531–2553.