Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in
Systems Biology

Current progress and potential opportunities to infer single-cell developmental trajectory and cell fate

Lingfei Wang[1,2,3,a], Qian Zhang[1,2,3,a], Qian Qin[1,2,3,a], Nikolaos Trasanidis[1,2,4,a], Michael Vinyard[1,2,3,5,a], Huidong Chen[1,2,3,a] and Luca Pinello[1,2,3]

Abstract
Rapid technological advances in transcriptomics and lineage tracing technologies provide new opportunities to understand organismal development at the single-cell level. Building on these advances, various computational methods have been proposed to infer developmental trajectories and to predict cell fate. These methods have unveiled previously uncharacterized transitional cell types and differentiation processes. Importantly, the ability to recover cell states and trajectories has been evolving hand-in-hand with new technologies and diverse experimental designs; more recent methods can capture complex trajectory topologies and infer short- and long-term cell fate dynamics. Here, we summarize and categorize the most recent and popular computational approaches for trajectory inference based on the information they leverage and describe future challenges and opportunities for the development of new methods for reconstructing differentiation trajectories and inferring cell fates.

Addresses
[1] Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital Research Institute, Charlestown, USA
[2] Department of Pathology, Harvard Medical School, Boston, USA
[3] Broad Institute of Harvard and MIT, Cambridge, MA, USA
[4] Centre for Haematology, Department of Immunology and Inflammation, Imperial College London, UK
[5] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA

Corresponding author: Pinello, Luca (lpinello@mgh.harvard.edu)
[a] These authors contributed equally to this work.

Current Opinion in Systems Biology 2021, **26:**1–11

This review comes from a themed issue on **Theoretical approaches to analyze single-cell data**

Edited by **Olivier Gandrillon** and **Michael Stumpf**

For a complete overview see the Issue and the Editorial

Available online 26 March 2021

https://doi.org/10.1016/j.coisb.2021.03.006

2452-3100/© 2021 Elsevier Ltd. All rights reserved.

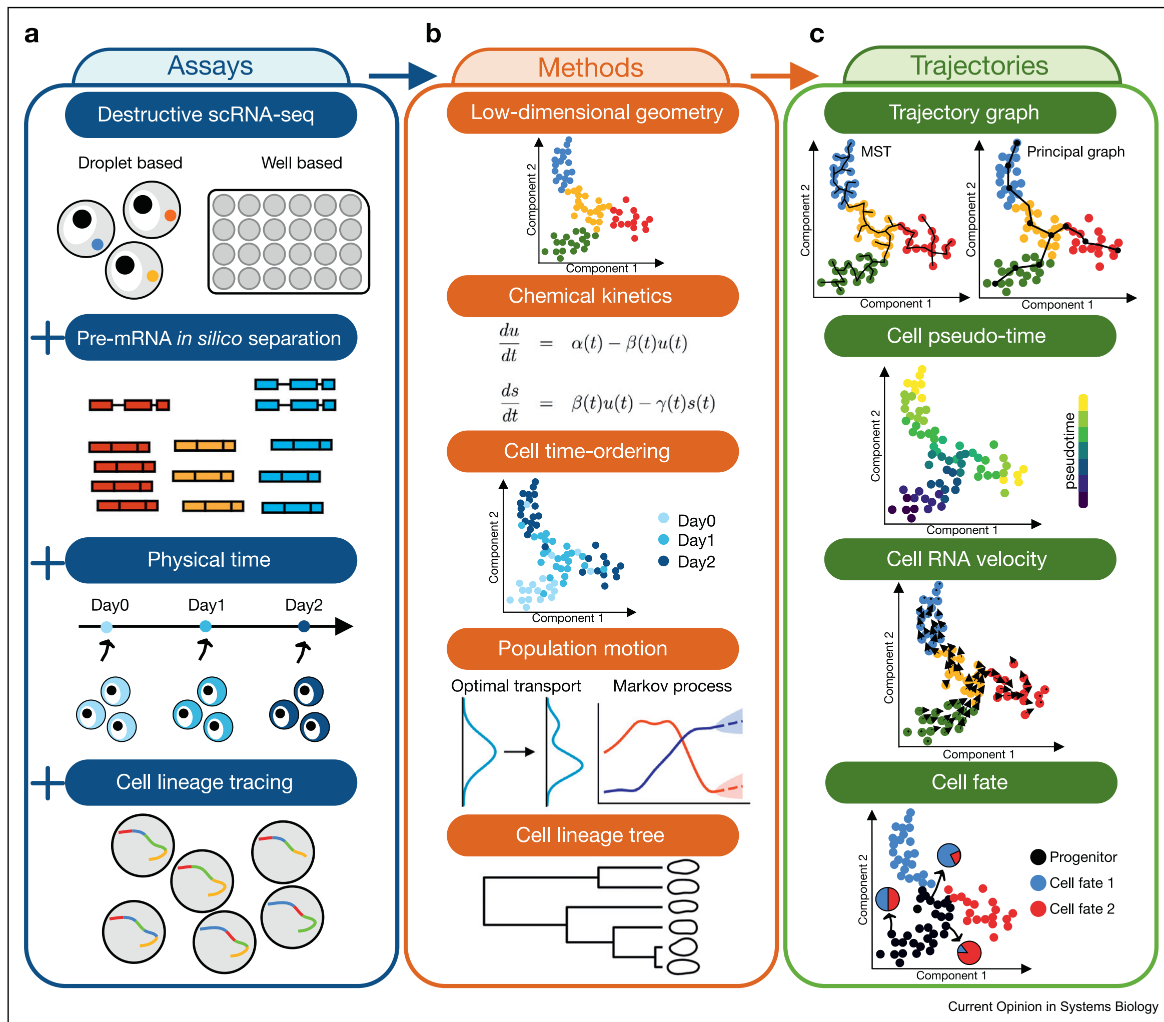
Keywords
Cell fate, Trajectory inference, Single-cell, scRNA-seq, Pseudotime, Lineage tracing, Multiomics.

Introduction

During the last decade, the development of single-cell technologies has revolutionized the way we investigate the dynamics of cellular processes across development and disease. Various cellular trajectory inference methods have been implemented to reconstruct the molecular continuum of cellular profiles and to study cell-cycle phase transitions, cell activation, and cell fate decisions [1]. Cellular trajectory analysis has enabled the discovery of novel cell types and subpopulations during differentiation [2,3], as well as helped to refine established developmental paths and hierarchies in several biological contexts [4,5]. In addition, decomposition of the clonal architecture of cancer cell populations across time has provided unique insights into the evolution of tumor clones, new mechanisms of chemotherapy resistance, and novel targets for drug development [6]. Cell lineage tracing has also pinpointed cell reprogramming events, such as in hematopoiesis [7,8] and epithelial—mesenchymal transition [9].

These discoveries were possible, thanks to the introduction of single-cell assays (Figure 1a). Conventionally destructive single-cell RNA sequencing (scRNA-seq) captures a continuous distribution of cells across differentiation stages. Each cell stage can be thought as a vector in a multidimensional space where each dimension corresponds to a gene, and similar transcriptional states occupy similar regions in this space. To capture the major variation modes that define cell differences and relations, the full transcriptomic profiles are projected onto a lower-dimensional 'state space' with dimension reduction techniques (e.g. Principal Component Analysis or PCA, Uniform Manifold Approximation or UMAP). In this space, trajectory inference methods can recover a scale-free 'pseudo-time' to represent the relative differentiation stage each cell is at, and/or a trajectory graph based on curves, trees, or graphs that summarizes the developmental paths across cell populations. In addition, it is possible to infer the direction and speed of cellular motion in the state space, known as 'RNA velocity,' based on mRNA processing kinetics and by interrogating reads corresponding to pre- (unspliced) and

Figure 1



A brief summary of **(a)** current assays (left, blue), **(b)** computational methods (middle, orange), and **(c)** single-cell trajectory definitions (right, green) typically considered in trajectory inference or cell fate prediction.

mature (spliced) mRNAs. When single-cell measurements incorporate time or clonal information, methods are also available to leverage this information to better predict cell differentiation dynamics and ultimately ‘cell fate,’ that is, the probability of reaching a given differentiated state by a given cell.

Meanwhile, the emergence of multiomics single-cell technologies has increased the complexity and dimensionality of single-cell measurements. These advancements provide new opportunities to better investigate gene regulation and dynamic molecular processes [10–12]. In fact, independent or integrated mappings of transcriptomic, epigenomic, and proteomic measurements provide the potential to understand the mechanisms that contribute to gene regulation and function along developmental trajectories and cell state transitions. For example, these assays can provide insights into the transcriptional regulators and functional genomic regions mediating these processes and can enable the reconstruction of dynamic gene regulatory networks that orchestrate these molecular programs.

Here, we provide an overview of methods to reconstruct single-cell trajectories or to model cell fate, their required inputs, and their outputs based on available single-cell assays and experimental designs (Figure 1). Within the length requirements of this short review, here we focus on discussing methods of high popularity and performance based on overall citations, previous benchmarks [13], and our personal experience, with an emphasis on the widely used methods within the last 2 years.

Trajectory inference from snapshot data

Trajectory inference methods starting from snapshot scRNA-seq data, represented as a gene-by-cell matrix of the transcriptome, assign a numerical value to each cell referred to as pseudotime. Based on pseudotime, cells can be ordered and their progression along this pseudotemporal axis may recapitulate biological developmental processes.

Trajectory inference methods can be broadly divided into two categories (Figure 1c). In the first, pseudotime is calculated by projecting cells onto an explicit trajectory graph (e.g. Monocle [14–16], TSCAN [17], Slingshot [18], and STREAM [19]). The trajectory graph is learned through graph or curve fitting in a low-dimensional space using a minimum spanning tree (MST), principal curve, principal graph, or similar methods. In the second, pseudotime is directly inferred from cell coordinates in a low-dimensional space, using *k*-nearest neighbors (*k*-NN)-based random walks or shortest path (e.g. Wishbone [20], DPT [21], PAGA [22], and Palantir [23]). These methods vary by their ability in detecting different trajectory topologies: linear, tree (such as bifurcation and multifurcation), graph with cycles, and disconnected graphs. Importantly, several of these methods are comprehensively reviewed and benchmarked in Ref. [13].

Moreover, some methods can account for additional information such as gene expression variability. Variations of gene expression levels have multiple origins, which contribute differently to the phenotypic variability of single cells at varying efficiency levels (see Ref. [24] for a comprehensive review). One important source of gene expression variation is the discrete bursts of mRNA transcription [25]. Such transcriptional bursting may confound the result of trajectory inference. To address that, CALISTA [26] is based on SABEC [27] and can explicitly model the transcriptional bursting in trajectory inference to simultaneously obtain a mechanistic understanding of gene transcription.

However, trajectory inference methods based on snapshot data generate pseudotemporal ordering of cells from static transcriptional states and cannot incorporate time information even if it is available, a problem also termed time ambiguity. Consequently, this may lead to misinterpretation of cellular developmental processes as well as the mechanisms that drive their dynamics [28,29]. Some RNA-based features (e.g. the number of expressed genes) have been leveraged to improve the prediction of developmental potential [30]. These limitations can also be compensated by distinguishing between pre-mRNAs and mature mRNAs, introducing external information such as sampling timepoints, or using clonal history recovered by lineage tracing assays, as illustrated in the following sections.

RNA velocity

ScRNA-seq readouts contain both (mature) mRNA and pre-mRNA reads; however, the latter are not used in traditional trajectory inference methods. This extra information can be used to determine the chemical kinetic parameters for RNA processing for each gene and to predict the mature mRNA's rate of change, termed RNA velocity as described in both bulk [31] and single-cell [32] RNA-seq studies.

Velocyto [32], the first method to estimate single-cell RNA velocity, is based on a steady-state model that uses the disproportion of pre-mRNA compared to mRNA to estimate the expected mature mRNA profile of the corresponding future cell (Figure 1c). ScVelo [33] fits mature mRNA and pre-mRNA levels with a system of differential equations describing transcription, mRNA processing, degradation, and latent time. Dynamo [34] can also integrate metabolic-labeling information for pre-mRNA age profiling to improve the RNA velocity estimation.

The gene-level RNA velocities are then combined to a single-cell-level RNA velocity based on a transition probability matrix between each predicted future cell and observed cells. These cell velocity estimates are then visualized in a low-dimensional state space as vectors to represent the expected 'motion' in terms of direction and magnitude. CellRank [35] and Dynamo both extend RNA velocity to long-term extrapolation of cell fate, although extrapolation accuracy tends to decline on longer time scales as uncertainty accumulates with each cell–cell transition prediction.

These strategies are based on predefined mRNA dynamics and can offer a potential resolution to the time ambiguity problem faced by trajectory inference of snapshot scRNA-seq measurements. However, RNA velocity measurements in single cells are inherently noisy because of limited read counts of pre-mRNA. Consequently, RNA velocity vectors are usually smoothed between nearby cells to obtain a more robust, representative, and collective motion. In addition, RNA velocity depends heavily on the subset of genes (e.g. variable genes) chosen before dimension reduction as well as on general preprocessing [36]. Despite these challenges, RNA velocity is widely applicable as the mature and pre-mRNA read counts are inherent to conventional scRNA-seq.

Table 1

A brief summary of lineage tracing technologies that can measure other modalities in parallel.

Category	Protocol	Barcode	Description
CRISPR	Gestalt, CARLIN [46,48]	Evolving barcode	These assays use CRISPR genome editing to create 'scars' in genomic DNA, which can be used to reconstruct lineage information during development by targeted DNA sequencing.
CRISPR & lentivirus	macsGESTALT [9]	Evolving & static barcode	macsGESTALT couples CRISPR genome editing and lentivirus-based barcoding to track the progression of clones and sub-clones in cancer.
Lentivirus	CellTag [43]	Static barcode	In CellTag, GFP is lentivirally integrated with short barcodes encoded in its 3' UTR. This enables the parallel capture of lineage information coming from the GFP expression and transcriptomic data. Multiple infections over time allow for increased lineage tracing resolution in time-series experiments.
	LARRY [7,8]		LARRY uses a similar same strategy as CellTag; however, only one lentiviral infection is performed.
Mutation	RETrace [45]	Microsatellite loci mutations	Retrace jointly measures DNA methylation along with DNA microsatellite mutations at single-cell resolution.
	ScATAC-seq [11,50,51]	Mitochondria mutations	Single-cell ATAC-seq can be repurposed to study the phylogenetic relationship across cells by exploiting the presence of mitochondrial mutations.

Trajectory inference with time information

Time-series experiments, where single cells are profiled at multiple time points, enable the reconstruction of dynamic processes from static measurements. Although each measurement remains destructive, this design can mitigate its shortcomings by providing physical time anchors to cell pseudotime (Figure 1). A growing number of computational methods have been proposed to take advantage of time series samples. Some use physical time to improve the accuracy in pseudotime inference (e.g. pseudodynamics [37] psupertime [38]). Tempora additionally uses clusters' pathway enrichment profiles to connect related cell types and states across time points [39].

More interestingly, physical time opens up the possibility to infer the probabilistic differentiation fate of individual cells. Waddington-OT [40] and TrajectoryNet [41] summarize observations of cell states with a distribution at each time point and minimize the distribution density relocation across time points, known originally as 'optimal transport' in transportation theory. These methods can derive a transition probability matrix (as in RNA velocity) for short-term extrapolation and induce likely trajectories [40]. PRESCIENT [42] learns the underlying differentiation dynamics based on a generative deep learning model from time-series gene expression data and can predict long-term cell fate by simulating differentiation with stochastic Markov process.

Despite these advantages, time-series experiments typically require a multibatch design and therefore incur the added potential for batch effects. In addition, a finer sampling resolution in physical time also demands higher costs in labor and resources. On the computational side, extrapolations based on time-series information are also subject to a decline in accuracy at longer time scales.

Cell fate modeling with lineage tracing

Lineage tracing provides a parallel modality to mitigate the drawbacks of inferring the long-term dynamics from destructive single-cell measurements. Lineage tracing uses artificial (e.g. cell barcode) or natural (e.g. somatic mutations) variations in cellular DNA to reconstruct cell lineages. Because variations in genomic DNA are heritable, daughter cells can be traced back to their shared ancestor despite the limitations of destructive measurements. Combined with other single-cell measurements, these lineage tracing technologies provide supplemental information to improve trajectory inference and to interrogate developmental processes (Figure 1). Orthogonally, these experiments serve as potential benchmarks for existing methods for trajectory inference.

Table 2

A shortlist of computational methods and software to reconstruct single-cell trajectories.

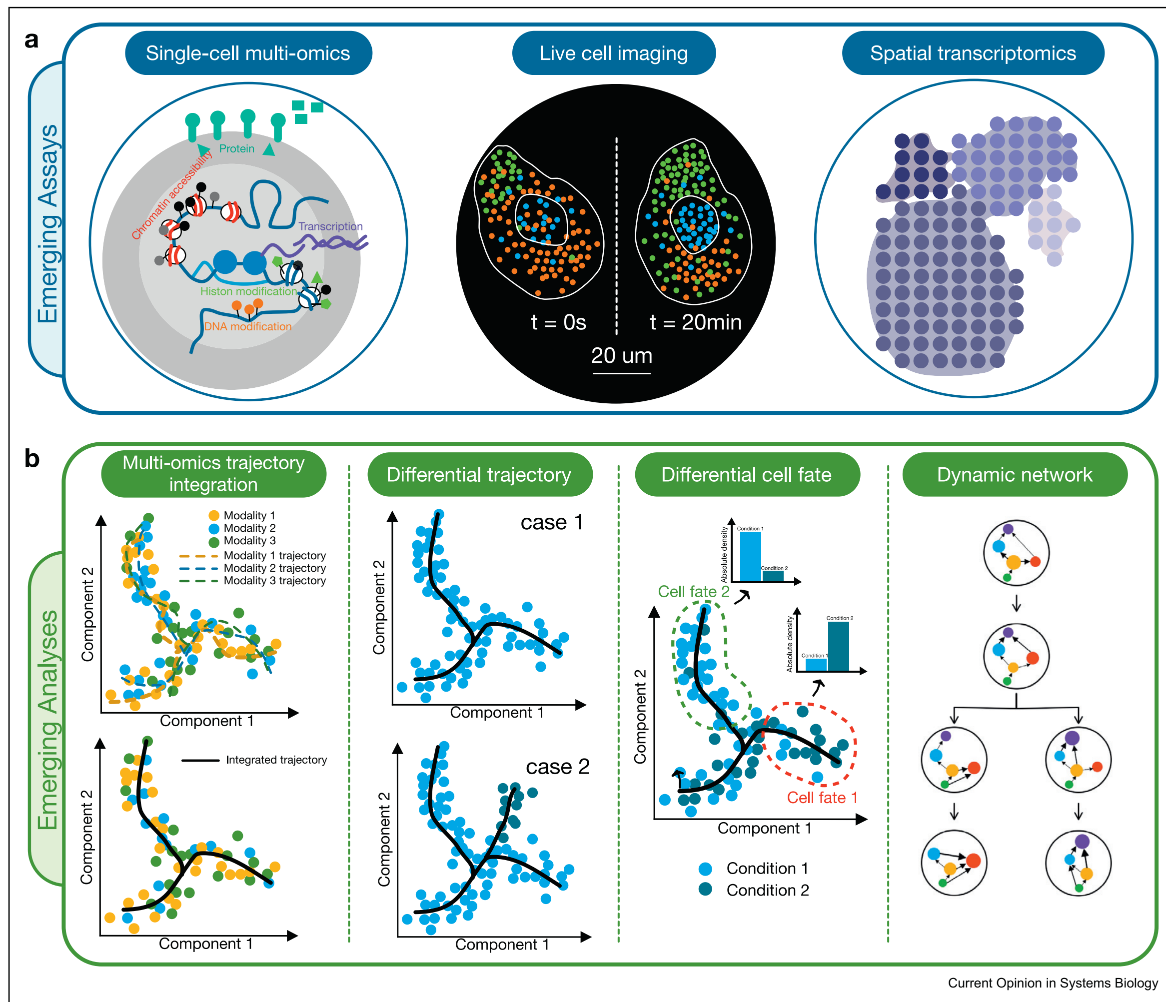
	Name	Input	Output	URL	Description
Trajectory inference from snapshot data (see Ref. [13] for a comprehensive list)	Wishbone [20]	Transcriptome	Pseudotime	https://github.com/dpeerlab/wishbone	Wishbone orders cells and identifies branches by building a *k*-NN graph and calculating the shortest paths between states based on a set of 'waypoints.'
	DPT [21]			https://www.helmholtz-muenchen.de/icb/research/groups/machine-learning/projects/dpt/index.html	DPT reconstructs cellular developmental progression and branching decisions by measuring transitions between cells based on diffusion-like random walks.
	Palantir [23]			https://github.com/dpeerlab/Palantir/	Palantir assigns pseudo-time to cells by modelling cell fates as a probabilistic process using a Markov chain.
	PAGA [22]		Pseudotime Trajectory graph	https://github.com/theislab/paga	PAGA learns the topology of cells at a chosen resolution by partitioning a *k*-NN graph and ordering cells using random-walk-based distances in the PAGA graph.
	Monocle [14–16]			https://cole-trapnell-lab.github.io/monocle3/	Monocle infers trajectories by learning a minimum spanning tree, a DDRtree-based principal tree, and a SimplePPT-based principal graph, respectively. Cells are ordered by projecting them to the learnt graph.
	TSCAN [17]			https://github.com/zji90/TSCAN	TSCAN infers cellular trajectories by learning a minimum spanning tree and pseudotime is obtained by projecting each cell onto the learnt tree.
	Slingshot [18]			https://github.com/kstreet13/slingshot	Slingshot infers cell lineages using simultaneous principal curves and pseudotime is obtained by projecting cells onto these curves.
	STREAM [19]			https://github.com/pinellolab/STREAM	STREAM infers trajectories by learning an elastic principal graph and orders cells by projecting them onto the learned graph.
	CALISTA [26]			https://www.cabselab.com/calista	CALISTA uses a steady-state model of transcriptional bursting for trajectory inference.
RNA velocity	Velocyto [32]	mRNA Pre-mRNA	Pseudotime RNA velocity	http://velocyto.org/	Velocyto uses a steady-state model to describe mRNA dynamics.
	ScVelo [33]			https://github.com/theislab/scvelo	ScVelo fits a dynamic model to infer RNA velocity and pseudotime.
	CellRank [35]		Pseudotime RNA velocity Cell fate	https://github.com/theislab/cellrank	CellRank extrapolates RNA velocity to long-term cell fate with Markov processes.
	Dynamo [34]	mRNA pre-mRNA metabolic labelling		https://github.com/aristoteleo/dynamo-release	Dynamo can handle metabolic labeling of pre-mRNA for RNA velocity, and extrapolate it to long-term cell fate.
Trajectory inference with time information	pseudodynamics [37]	Transcriptome Time	Pseudotime	https://github.com/theislab/pseudodynamics	Pseudodynamics is a mathematical probabilistic framework that combines cell state dynamics and pseudotemporal ordering to infer underlying developmental trajectories.

(*continued on next page*)

Table 2. (*continued*)

	Name	Input	Output	URL	Description
	psupertime [38]			https://github.com/wmacnair/psupertime	Psupertime is a regression-based supervised pseudotime assignment approach that can not only project each cell onto a pseudotemporal axis but also identify genes regulated during biological processes.
	Waddington-OT [40]			https://broadinstitute.github.io/wot/	Waddington-OT is an optimal transport-based approach that uses time course scRNA-seq data to infer ancestor-descendant fate relationships and the dynamic changes (of cell identity via gene expression) probability distributions over time.
	TrajectoryNet [41]			https://www.krishnaswamylab.org/projects/trajectory-net	TrajectoryNet captures dynamic optimal transport between distributions of time series data to model dynamic trajectory continuously along time.
	Tempora [39]	Transcriptome Time Clustering		https://github.com/BaderLab/Tempora	Tempora is an information-theoretic approach that combines time labels and biological pathway information to infer trajectory at the cell cluster level.
	PRESCIENT [42]	Transcriptome Time	Cell fate	https://github.com/gifford-lab/prescient	PRESCIENT infers trajectory by integrating gene expression and physical time with or without lineage information using a deep generative model.
Single-cell lineage tracing	LineageOT [54]	Transcriptome Lineage	Phylogenetic tree	None	LineageOT infers trajectory and lineage trees by optimal transport, optimizing distances of leaves of dynamical phylogenetic trees across timepoints.
	LinTiMaT [53]			https://github.com/jessica1338/LinTIMaT	LinTiMaT infers lineage trees integrating mRNA and lineage with a Bayesian hierarchical clustering.
	PRESCIENT [42]	Transcriptome Time Lineage	Cell fate	https://github.com/gifford-lab/prescient	PRESCIENT uses optional clonal count to inform its model parameters.

Figure 2



Emerging assays (**a**) and current computational opportunities to extend trajectory inference (**b**).

Lineage tracing technologies that are capable of parallel quantifications of lineage and other modalities (e.g. transcriptomics and chromatin accessibility) in the same single cells can be broadly split into three categories (Table 1). In the first [4,7,8,43–45], DNA barcodes are integrated into the genome of cells via a lentiviral vector and later detected in daughter cells. This is accomplished by scRNA-seq read-out of the expressed transcripts along with the barcoded transcripts. In the second [46–48], CRISPR-Cas9 is delivered to cells and targeted to predefined sequences to introduce unique combinations of DNA mutagenesis and create a set of evolving barcodes over time. This is accomplished by targeted DNA sequencing with shared primers specific to the integrated amplicon with or without common scRNA-seq. Notably, these two categories can also be combined for higher resolution lineage tracing [9]. In the third, somatic variations such as short tandem repeats [49] or mitochondrial DNA [11,50,51] are profiled. Mitochondrial DNA mutates at a higher rate than nuclear DNA, enabling lineage tracing. See Ref. [29] for an in-depth review.

Computational methods have been proposed to reconstruct phylogenetic lineage tree based only on lineage tracing information (e.g. DNA barcodes) [52] or by incorporating also gene expression information as in LinTIMaT [53]. For trajectory inference, LineageOT extends Waddington-OT's optimal transport by incorporating lineage information to improve cell fate prediction [54]. PRESCIENT can also leverage lineage information when available to inform its model parameters with clonal cell count [42].

Single-cell lineage tracing is a fast-developing field with several emerging technologies. However, current assays

pose several challenges. For example, lineage tracing assays based on genome editing are difficult to apply to primary human tissues. In addition, the detection of DNA variations is limited by the inherent single-cell data sparsity. At the time of writing, computational methods for trajectory inference with lineage tracing are still in active development with very few off-the-shelf solutions.

Taken together, our brief review of methods shows that they are often based on similar principles, but additional information often leads to more expressive and accurate modeling of trajectory and cell fate prediction. We have categorized and summarized a noncomprehensive list of these methods with required inputs, outputs, core ideas, and software implementations in Table 2. We hope this table can provide a useful resource to select the appropriate analysis method based on available experimental assays and designs.

Future perspectives

The development of methods for trajectory inference has been driven by two parallel frontiers: the upstream technological developments followed by the downstream biological investigation and discovery. As novel technologies and methodologies emerge, the biological questions are being refined. Trajectory inference provides numerous exciting possibilities to take advantage of technological advancements to answer outstanding questions in biology (Figure 2).

More scalable and cost-effective single-cell assays are continuously emerging. These assays combined with genetic perturbations and drug treatments offer unprecedented opportunities to dissect and study cellular specification, differentiation, and development in health and disease.

In this context, trajectory inference offers an important modeling tool to detect 'differential trajectory,' that is, whether and how cells of different conditions are distributed in a putative shared trajectory graph. A straightforward answer may arise from direct comparisons of the inferred trajectory graphs or cell fates, but it remains challenging to account for the trajectory/fate inconsistencies because of the technology used for data generation, data preprocessing measures such as normalization and batch correction, and trajectory inference parameterization. Although these choices have not been systematically assessed in the context of trajectory inference, we believe this is an important area of investigation. For example, a recent study [55] has demonstrated how normalization procedures (e.g. log of counts per million) and feature selection based on highly variable genes may introduce false variability in procedures for dimensionality reduction. This study also introduces a package called GLM-PCA (PCA for generalized linear models) that provides methods for feature selection and dimensionality reduction based on a multinomial modeling of count data. Even though this study showed how these modeling choices can improve clustering as a downstream task, we believe that the proposed procedures can also benefit trajectory inference methods. In addition, another recent study [36] has shown how these preprocessing steps are also important for RNA velocity estimation, as discussed in the RNA velocity section.

In parallel, a cell-level interpretation would ask a 'differential cell fate' question, that is, how individual cell fate can be affected, based on either case−control experiments or in silico perturbations. Initial methodological efforts in these directions (e.g. CellAlign) have been made to align and compare pseudotimes and trajectory graphs [56−59]. Still, this area is in its infancy, and formulating statistical models is difficult but remains crucial.

Differentiation is long believed to be a process of gene regulatory network rewiring that determines cell identity and function. Trajectory inference offers a reference graph along which the rewiring may be recovered at improved time and gene resolutions [60]. Such 'dynamic gene regulatory networks' may reveal the transient and sequential regulatory events that drive or commit cells to differentiation, in parallel with trajectory-based differential gene expression that focuses on molecular events [61].

In addition, single-cell multiomics provide a more comprehensive understanding of intracellular processes, especially in distinguishing the causes from consequences of lineage commitment. Initial efforts toward the multiomics trajectory inference have been made (e.g. STREAM reconstructs developmental trajectories from either single-cell transcriptome or chromatin accessibility). More recently, joint single-cell transcriptome and chromatin accessibility profiles were used to infer cell fate decisions by predicting the cells' future transcriptome states from their current chromatin profiles ('chromatin potential') [11]. This analysis revealed how dynamic changes in chromatin accessibility precede corresponding gene expression in cell lineage commitment [10,11,62]. Protein-mRNA joint measurements have also enabled protein velocity to extrapolate cell states [63]. These assays will continue to unveil opportunities toward more accurate and mechanistically insightful trajectory inference. However, integrating single-cell multiomics remains a current computational challenge [64]. There are two potential strategies to infer joint multimodal trajectories (Figure 2b). One is to do crossmodality integration (e.g. Seurat [65], Conos [66], and MOFA+ [67]) and then infer trajectory directly in the

integrated space. Another is to infer the trajectory separately per modality and later integrate those trajectories into one.

Live-cell imaging can measure and track mRNA and protein molecules within single cells in a real-time, nondestructive manner [68,69]. This technology features the unique capacity to track the abundance and localization of individual molecules in the same cell across multiple time points, rather than choosing a surrogate similar cell at a future time in conventional destructive measurements. Nevertheless, most of these assays are invasive (require genetic engineering of targeting cells) and can measure only a handful of genes simultaneously in each single cell [68,70]. Live-imaging techniques require further improvement to overcome these limitations.

Finally, single-cell spatial transcriptomic technologies can capture gene expression variations across histologically discrete cell types in a high-throughput manner. The combination of spatial mapping and cell transition inference may unveil the transcriptional changes and mechanisms of anatomically restricted cell populations and provide novel insights into the molecular programs within developing tissues [71].

In conclusion, as single-cell assays and technologies constantly evolve toward multiple modalities and higher resolution, harnessing their full potential poses several computational challenges but also opens new opportunities for more accurate and refined trajectory inference. Trajectory inference is an evolving and exciting field where the fusion of experimental technologies and computational methods is poised to progressively unfold the fundamentals of developmental biology.

Funding

This project has been made possible in part by grant number 2019-202669 from the Chan Zuckerberg Foundation. LP is also partially supported by the National Human Genome Research Institute (NHGRI) Career Development Award (R00HG008399) and Genomic Innovator Award (R35HG010717). MEV is supported by the National Cancer Institute (NCI) Ruth L. Kirschstein NRSA Individual Predoctoral Fellowship (1F31CA257625-01). NT is partially supported by the European Hematology Association (EHA) Research Mobility Grant Award.

Conflict of interest statement

Nothing declared.

Acknowledgements

We would like to thank Guanlan Dong and Jayoung Ryu for their invaluable feedback on the manuscript.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

1. Baron CS, Oudenaarden A van: **Unravelling cellular relationships during development and regeneration using genetic lineage tracing**. *Nat Rev Mol Cell Biol* 2019, **20**:753–765.

2. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, Biffi A, Klein AM, Biasco L: **A comprehensive single cell transcriptional landscape of human hematopoietic progenitors**. *Nat Commun* 2019, **10**:1–15.

3. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ: **Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation**. *Cell* 2018, **173**: 1535–1548. e16.

4. Kim IS, Wu J, Rahme GJ, Battaglia S, Dixit A, Gaskell E, Chen H, Pinello L, Bernstein BE: **Parallel single-cell RNA-seq and genetic recording reveals lineage decisions in developing embryoid bodies**. *Cell Rep* 2020, **33**:108222.

5. Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, Waisman A, Huh JR, Klein AM, Socolovsky M: **Population snapshots predict early haematopoietic and erythroid hierarchies**. *Nature* 2018, **555**:54–60.

6. González-Silva L, Quevedo L, Varela I: **Tumor functional heterogeneity unraveled by scRNA-seq technologies**. *Trends in Cancer* 2020, **6**:13–19.

7. Rodriguez-Fraticelli AE, Weinreb C, Wang S-W, Migueles RP, Jankovic M, Usart M, Klein AM, Lowell S, Camargo FD: **Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis**. *Nature* 2020, **583**:585–589.

8. * Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM: **Lineage tracing on transcriptional landscapes links state to fate during differentiation**. *Science* 2020:367.

The authors developed a new sequencing-based lineage tracing technique by using a lentivirally-integrated barcoding strategy, named LARRY together with the inDrop single cell sequencing to measure gene expression and transcribed barcodes, this provides a ground-truth dataset for future lineage tracing studies through the addition of a barcode to trace the real cell trajectory in real time. This study highlights the necessity of improved methods for understanding cell fate, particularly as it relates to gene expression.

9. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, McKenna A, Shendure J, Lengner CJ: **Single-cell lineage and transcriptome reconstruction of metastatic cancer reveals selection of aggressive hybrid EMT states**. *bioRxiv* 2020, https://doi.org/10.1101/2020.08.11.245787.

10. Chen S, Lake BB, Zhang K: **High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell**. *Nat Biotechnol* 2019, **37**:1452–1457.

11. * Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, *et al.*: **Chromatin potential identified by shared single-cell profiling of RNA and chromatin**. *Cell* 2020, **183**:1103–1116.e20.

The latest in a series of methods that jointly profile gene expression and chromatin accessibility from the same cell. They proposed a method for identifying cis-regulatory interactions. Further, they introduced the concept of “chromatin potential” to quantitatively predict cell fates based on the integration of scATAC-seq and scRNA-seq modalities.

12. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, *et al.*: **Joint profiling of chromatin accessibility and gene expression in thousands of single cells**. *Science* 2018, **361**:1380–1385.

13. * Saelens W, Cannoodt R, Todorov H, Saeys Y: **A comparison of single-cell trajectory inference methods**. *Nat Biotechnol* 2019, **37**:547–554.

The authors benchmarked 45 trajectory inference methods on 110 real and 229 synthetic scRNA-seq datasets and highlight the

complementarity of these methods. Their evaluation pipeline not only provides guidelines for users but also can be helpful for developing new tools.

14. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, *et al.*: **The single-cell transcriptional landscape of mammalian organogenesis**. *Nature* 2019, **566**:496–502.

15. 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.

16. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C: **Reversed graph embedding resolves complex single-cell trajectories**. *Nat Methods* 2017, **14**:979–982.

17. Ji Z, Ji H: **TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis**. *Nucleic Acids Res* 2016, **44**: e117. e117.

18. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S: **Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics**. *BMC Genom* 2018, **19**:477.

19. Chen H, Albergante L, Hsu JY, Lareau CA, Bosco GL, Guan J, Zhou S, Gorban AN, Bauer DE, Aryee MJ, *et al.*: **Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM**. *Nat Commun* 2019, **10**:1–14.

20. * 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.
The authors proposed Palantir a method that can infer pseudo-time ordering of cells and assigns probability for differentiating into each terminal state. This method is shown to be well-suited for characterizing cell fate decisions.

21. Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ: **Diffusion pseudotime robustly reconstructs lineage branching**. *Nat Methods* 2016, **13**:845–848.

22. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ: **PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells**. *Genome Biol* 2019, **20**: 59.

23. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D: **Characterization of cell fate probabilities in single-cell data with Palantir**. *Nat Biotechnol* 2019, **37**:451–460.

24. Raser JM, O'Shea EK: **Noise in gene expression: origins, consequences, and control**. *Science* 2005, **309**:2010–2013.

25. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, Segerstolpe Å, Rivera CM, Ren B, Sandberg R: **Genomic encoding of transcriptional burst kinetics**. *Nature* 2019, **565**:251–254.

26. Papili Gao N, Hartmann T, Fang T, Gunawan R: **CALISTA: clustering and LINEAGE inference in single-cell transcriptional analysis**. *Front Bioeng Biotechnol* 2020, **8**.

27. Ezer D, Moignard V, Göttgens B, Adryan B: **Determining physical mechanisms of gene expression regulation from single cell gene expression data**. *PLoS Comput Biol* 2016, **12**, e1005072.

28. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM: **Fundamental limits on dynamic inference from single-cell snapshots**. *Proc Natl Acad Sci Unit States Am* 2018, **115**: E2467–E2476.

29. * Wagner DE, Klein AM: **Lineage tracing meets single-cell omics: opportunities and challenges**. *Nat Rev Genet* 2020, **21**: 410–427.
The authors review the current progress, challenges, and opportunities in studying cellular differentiation dynamics, providing new insights into the potential improvements of cellular dynamics reconstruction.

30. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S, *et al.*: **Single-cell transcriptional diversity is a hallmark of developmental potential**. *Science* 2020, **367**:405–411.

31. Zeisel A, Köstler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, Rechavi G, Soen Y, Jung S, Yarden Y, *et al.*: **Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli**. *Mol Syst Biol* 2011, **7**:529.

32. Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, *et al.*: **RNA velocity of single cells**. *Nature* 2018, **560**:494–498.

33. * Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ: **Generalizing RNA velocity to transient cell states through dynamical modeling**. *Nat Biotechnol* 2020, **38**:1408–1414.
The authors developed a dynamical model for single-cell RNA velocity extending the initial proposal of single-cell RNA velocity in [29]. Compared to steady-state models, this work improves RNA velocity's accuracy and resolution in many real-world applications.

34. Qiu X, Zhang Y, Yang D, Hosseinzadeh S, Wang L, Yuan R, Xu S, Ma Y, Replogle J, Darmanis S, *et al.*: **Mapping vector field of single cells**. *bioRxiv* 2019, https://doi.org/10.1101/696724.

35. Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, Lickert H, Ansari M, Schniering J, Schiller HB, *et al.*: **CellRank for directed single-cell fate mapping**. *bioRxiv* 2020, https://doi.org/10.1101/2020.10.19.345983.

36. * Soneson C, Srivastava A, Patro R, Stadler MB: **Preprocessing choices affect RNA velocity results for droplet scRNA-seq data**. *PLoS Comput Biol* 2021, **17**, e1008585.
The authors highlight the practical factors in data preprocessing that affect the reconstruction of RNA velocity its biological interpretation. This work informs users of the uncertainties in RNA velocity and in its consequent investigations.

37. Fischer DS, Fiedler AK, Kernfeld EM, Genga RMJ, Bastidas-Ponce A, Bakhti M, Lickert H, Hasenauer J, Maehr R, Theis FJ: **Inferring population dynamics from single-cell RNA-sequencing time series data**. *Nat Biotechnol* 2019, **37**:461.

38. Macnair W, Claassen M: **psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels**. *bioRxiv* 2019, https://doi.org/10.1101/622001.

39. Tran TN, Bader GD: **Tempora: cell trajectory inference using time-series single-cell RNA sequencing data**. *PLoS Comput Biol* 2020, **16**, e1008205.

40. * Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, *et al.*: **Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming**. *Cell* 2019, **176**: 928–943. e22.
The authors proposed a trajectory inference method, Waddington-OT. It is an optimal transport-based approach that uses time course scRNA-seq data to infer ancestor-descendant fate relationships and the dynamic changes of probability distributions (of cell identity via gene expression) over time.

41. Tong A, Huang J, Wolf G, Dijk DV, Krishnaswamy S: **TrajectoryNet: a dynamic optimal transport network for modeling cellular dynamics**. In *International conference on machine learning*. PMLR; 2020:9526–9536.

42. Yeo GHT, Saksena SD, Gifford DK: **Generative modeling of single-cell population time series for inferring cell differentiation landscapes**. *bioRxiv* 2020, https://doi.org/10.1101/2020.08.26.269332.

43. Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, Morris SA: **Single-cell mapping of lineage and identity in direct reprogramming**. *Nature* 2018, **564**:219–224.

44. Eyler CE, Matsunaga H, Hovestadt V, Vantine SJ, van Galen P, Bernstein BE: **Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance**. *Genome Biol* 2020, **21**:174.

45. Pei W, Shang F, Wang X, Fanti A-K, Greco A, Busch K, Klapproth K, Zhang Q, Quedenau C, Sauer S, *et al.*: **Resolving fates and single-cell transcriptomes of hematopoietic stem cell clones by PolyloxExpress barcoding**. *Cell Stem Cell* 2020, **27**:383–395. e8.

46. Bowling S, Sritharan D, Osorio FG, Nguyen M, Cheung P, Rodriguez-Fraticelli A, Patel S, Yuan W-C, Fujiwara Y, Li BE, *et al.*: **An engineered CRISPR-cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells**. *Cell* 2020, **181**:1410–1422. e27.

47. McKenna A, Gagnon JA: **Recording development with single cell dynamic lineage tracing**. *Development* 2019:146.

48. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF: **Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain**. *Nat Biotechnol* 2018, **36**:442–450.

49. Wei CJ-Y, Zhang K: **RETrace: simultaneous retrospective lineage tracing and methylation profiling of single cells**. *Genome Res* 2020, https://doi.org/10.1101/gr.255851.119.

50. * Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A, *et al.*: **Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics**. *Cell* 2019, **176**:1325–1339. e22.
This article systematically explores the possibility of leveraging mitochondrial mutation profiles in single cell ATAC-seq to reconstruct the lineage in hematopoiesis and leukemia.

51. Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, Chang HY: **Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA**. *eLife* 2019, **8**, e45105.

52. Weinreb C, Klein AM: **Lineage reconstruction from clonal correlations**. *Proc Natl Acad Sci Unit States Am* 2020, **117**: 17041–17048.

53. Zafar H, Lin C, Bar-Joseph Z: **Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data**. *Nat Commun* 2020, **11**:1–14.

54. Forrow A, Schiebinger G: **A unified framework for lineage tracing and trajectory inference**. *bioRxiv* 2020, https://doi.org/10.1101/2020.07.31.231621.

55. Townes FW, Hicks SC, Aryee MJ, Irizarry RA: **Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model**. *Genome Biol* 2019, **20**:295.

56. Do VH, Blažević M, Monteagudo P, Borozan L, Elbassioni K, Laue S, Ringeling FR, Matijević D, Canzar S: **Dynamic pseudotime warping of complex single-cell trajectories**. *Bioinformatics* 2019.

57. Alpert A, Moore LS, Dubovik T, Shen-Orr SS: **Alignment of single-cell trajectories to compare cellular expression dynamics**. *Nat Methods* 2018, **15**:267–270.

58. Cacchiarelli D, Qiu X, Srivatsan S, Manfredi A, Ziller M, Overbey E, Grimaldi A, Grimsby J, Pokharel P, Livak KJ, *et al.*: **Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome**. *Cell Systems* 2018, **7**:258–268. e3.

59. Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ: **Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging**. *Genome Res* 2019, **29**:2088–2103.

60. Kamimoto K, Hoffmann CM, Morris SA: **CellOracle: dissecting cell identity via network inference and in silico gene perturbation**. *bioRxiv* 2020, https://doi.org/10.1101/2020.02.17.947416.

61. Berge KV den, Bézieux HR de, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S, Clement L: **Trajectory-based differential expression analysis for single-cell sequencing data**. *Nat Commun* 2020, **11**:1–13.

62. Zhu C, Yu M, Huang H, Juric I, Abnousi A, Hu R, Lucero J, Behrens MM, Hu M, Ren B: **An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome**. *Nat Struct Mol Biol* 2019, **26**:1063–1070.

63. Gorin G, Svensson V, Pachter L: **Protein velocity and acceleration from single-cell multiomics experiments**. *Genome Biol* 2020, **21**:39.

64. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A, *et al.*: **Eleven grand challenges in single-cell data science**. *Genome Biol* 2020, **21**:31.

65. Stuart T, Satija R: **Integrative single-cell analysis**. *Nat Rev Genet* 2019, **20**:257.

66. Barkas N, Petukhov V, Nikolaeva D, Lozinsky Y, Demharter S, Khodosevich K, Kharchenko PV: **Joint analysis of heterogeneous single-cell RNA-seq dataset collections**. *Nat Methods* 2019, **16**:695–698.

67. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, Stegle O: **MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data**. *Genome Biol* 2020, **21**:111.

68. Khuperkar D, Hoek TA, Sonneveld S, Verhagen BMP, Boersma S, Tanenbaum ME: **Quantification of mRNA translation in live cells using single-molecule imaging**. *Nat Protoc* 2020, **15**:1371–1398.

69. Sato H, Das S, Singer RH, Vera M: **Imaging of DNA and RNA in living eukaryotic cells to reveal spatiotemporal dynamics of gene expression**. *Annu Rev Biochem* 2020, **89**:159–187.

70. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N: **Noise in protein expression scales with natural protein abundance**. *Nat Genet* 2006, **38**:636–643.

71. Waylen LN, Nim HT, Martelotto LG, Ramialison M: **From whole-mount to single-cell spatial assessment of gene expression in 3D**. *Commun Biol* 2020, **3**:1–11.