Computational and Structural Biotechnology Journal 21 (2023) 2373–2380

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Mini-review

Single-cell and long-read sequencing to enhance modelling of splicing and cell-fate determination

Siyuan Wu [a,b,c], Ulf Schmitz [a,b,*]

[a] Department of Molecular & Cell Biology, James Cook University, Townsville 4811, Queensland, Australia
[b] Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Cairns 4870, Queensland, Australia
[c] School of Mathematics, Monash University, Melbourne 3800, Victoria, Australia

ARTICLE INFO

Article history:
Received 8 November 2022
Received in revised form 13 March 2023
Accepted 13 March 2023
Available online 20 March 2023

Keywords:
Mathematical modelling
RNA velocity
Alternative splicing
Isoform expression
Pesudotemopral trajectory inference
Transcriptome diversity

ABSTRACT

Single-cell sequencing technologies have revolutionised the life sciences and biomedical research. Single-cell sequencing provides high-resolution data on cell heterogeneity, allowing high-fidelity cell type identification, and lineage tracking. Computational algorithms and mathematical models have been developed to make sense of the data, compensate for errors and simulate the biological processes, which has led to breakthroughs in our understanding of cell differentiation, cell-fate determination and tissue cell composition. The development of long-read (a.k.a. third-generation) sequencing technologies has produced powerful tools for investigating alternative splicing, isoform expression (at the RNA level), genome assembly and the detection of complex structural variants (at the DNA level).

In this review, we provide an overview of the recent advancements in single-cell and long-read sequencing technologies, with a particular focus on the computational algorithms that help in correcting, analysing, and interpreting the resulting data. Additionally, we review some mathematical models that use single-cell and long-read sequencing data to study cell-fate determination and alternative splicing, respectively. Moreover, we highlight the emerging opportunities in modelling cell-fate determination that result from the combination of single-cell and long-read sequencing technologies.

1. Introduction

Cell-fate determination is a complex process regulated by various mechanisms, including transcriptional and post-transcriptional regulation, epigenetic modifications, and cell-cell interactions [1,2]. In 2009, the introduction of single-cell RNA sequencing (scRNA-seq) enabled the measurement of gene expression at the level of individual cells, providing critical insights into the molecular mechanisms that govern cell-fate decisions [3]. This breakthrough technology has facilitated the identification of various cellular states during tissue or organ development and the analysis of the developmental transition pathways between different states, leading to the construction of more detailed models of cell-fate determination that take into account cell-to-cell variability and stochasticity in gene expression.

However, developing more accurate models of cell-fate determination requires a detailed understanding not only of gene expression and regulation, but also of alternative splicing, splicing regulation, transcriptomic complexity, and isoform diversity at the single-cell level. Long-read sequencing has the potential to provide a more comprehensive view of these processes than short-read sequencing, as it can identify complex structural variants (DNA), whole transcript alternative splicing events, and cell-type-specific mRNA isoforms expression. Insights obtained from long-read sequencing can help us better understand the mechanisms underlying cell-fate determination [4,5]. Consequently, developing models that incorporate single-cell long-read sequencing data is the next logical step in studying mechanisms of cell-fate decision-making.

In this review, we introduce recent advances in long-read sequencing data analysis and its application in alternative splicing analysis. In Section 3, we focus on recent reports that apply mathematical modelling to study RNA velocity using single-cell sequencing data. In Section 4, we discuss how third-generation sequencing technologies can enhance RNA velocity modelling and advance research on cell-fate determination. We also deliberate on challenges

* Corresponding author at: Department of Molecular & Cell Biology, James Cook University, Townsville 4811, Queensland, Australia.
E-mail address: ulf.schmitz@jcu.edu.au (U. Schmitz).

https://doi.org/10.1016/j.csbj.2023.03.023

and opportunities for mathematical modelling using long-read data at the single-cell level. Finally, we highlight potential future research directions.

2. Recent advances in long-read sequencing data analysis

Long-read sequencing can generate reads of increasing length, with currently up to 2 megabases [6,7]. Progress in third-generation sequencing is mainly driven by two technologies: (i) Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing, and (ii) Oxford Nanopore Technologies' (ONT) nanopore sequencing. The former generates long-read sequence data by utilising a zero-mode waveguide (ZMW) and a charged-coupled device (CCD) camera to detect nucleotides based on four different fluorescent tags [8]. DNA fragments or RNA full-length transcripts are PCR-amplified in a circular fashion achieved by ligating hairpin adapters to every single molecule. Oxford Nanopore devices, on the other hand, record fluctuations of an electric current caused by different nucleotides, when a single DNA/RNA strand passes through a protein nanopore [9,10]. For a detailed introduction to these two technologies, see Pollard et al. [11].

Long-read sequencing, chosen as the Method of the Year 2022 [12], has revolutionised current life science and biomedical research, offering numerous new opportunities for researchers [13–16]. One such opportunity is the ability to identify gene fusion events, which are considered critical drivers of diseases such as cancer. Unlike short-read sequencing, which cannot provide full-length sequence information, long-read sequencing technologies can capture full-length gene or transcript sequences, enabling the identification of different isoform structures of fusion genes or chimeric transcripts. As a result, several tools have been developed to identify complex, full-length fusion transcripts based on long-read transcriptome sequencing data. For example, Liu et al. developed *LongGF*, a software that detects putative gene fusions from long-read transcriptome sequencing data [17]. Another tool that uses long-read sequencing data for fusion calling is *JAFFAL* [18]. Interestingly, this tool can also be applied to single-cell long-read sequencing data.

Many other algorithms have been developed for the analysis of long-read sequencing data, which can be used for base calling [19–22], quality control [23–26], genome assembly [27–30–33], structural variant detection [28,34–37,38], DNA/RNA modification detection [39–42–45], isoforms discovery [26,46–49], and the analysis of alternative splicing and isoform expression [4,50–53,54]. Current tools for long-read sequencing data analysis have been reviewed and benchmarked [13,14,55–57]. The long-read-tools.org database provides an up-to-date record of software tools for long-read sequencing data analysis [58].

2.1. Alternative splicing analysis meets long-read sequencing

Alternative splicing is the process of producing different mRNA isoforms from the same gene by selecting different combinations of splice sites (Fig. 1). This mechanism is crucial for cell-fate determination but can also facilitate the pathogenesis of diseases [59–62]. Alternative splice site selection can lead to the inclusion or exclusion of exonic and intronic sequences in mature mRNA transcripts. While the inclusion of intronic sequences (a.k.a. intron retention) often leads to the degradation of mRNA transcripts via nonsense-mediated decay [63], the differential inclusion of exonic sequences leads to alternative mRNA isoforms [62]. When translated, proteins with different structural features can be produced, which might have altered functions [64].

As third-generation sequencing technologies evolve, new analysis pipelines are needed to determine alternative splicing patterns in full-length transcript information. Tardaguila et al. have developed a pipeline, *SQANTI* [26], for quality control, isoform detection and classification from long reads. *SQANTI*, currently in its third version (*SQANTI3*), is integrated into the Functional Iso-Transcriptomics framework, together with *IsoAnnot* for transcriptome annotation and *tappAS* for functional alternative splicing analysis [65]. Leung et al. leveraged *SQANTI* to process their long-read sequencing data, which led to the discovery of many novel isoforms in the human and mouse cortex [53]. They also demonstrated widespread changes in alternative splicing and isoform diversity between the foetal and adult cortex [53]. Prjibelski et al. introduced a tool named *IsoQuant*, which can reconstruct isoform transcript structures from PacBio or ONT RNA reads [66].

Long-read sequencing also helps uncover the mechanisms underlying alternative splicing. For example, Wan et al. use long-read RNA sequencing data to support a testable prediction of the proposed model: that spliceosomes make many cuts to remove an intron instead of a single cut [67]. In this study, the authors observed transcriptional bursting behaviour across multiple endogenous human genes with distinct bursting frequencies and similar pre-mRNA dwell times from high-throughput dynamic imaging [67]. They also concluded that the stochasticity of alternative splice site selection is a prevalent mechanism across the human genome and recursive intron removal is the underlying processes required for producing mature mRNA transcripts from a pre-mRNA. Finally, they developed a model, based on chemical master equations, to describe transcription and splicing dynamics.

Long-read RNA sequencing can also contribute to studying protein isoform diversity, which is mostly a consequence of alternative splicing at the RNA level. Miller et al. have developed a long-read proteogenomics pipeline for integrating sample-matched long-read RNA sequencing and mass spectrometry-based proteomics data to enhance the detection and classification of protein isoforms. In this analysis pipeline, the long-read RNA sequencing data is used to predict full-length protein isoforms and generate a database, which can be used as reference for the mass spectrometry data analysis [68].

2.2. Single-cell long-read sequencing

A new frontier in the development of third-generation sequencing technologies is the implementation and data analysis of long-read sequencing at the single-cell level [69]. Although the analysis of single-cell long-read data has attracted much attention, available methods remain sparse, though several single-cell long-read sequencing protocols have been developed: *Smart-seq2* is a single-cell full-length RNA sequencing method that amplifies full-length cDNA from individual cells [70]. *ScISOr-seq* uses a unique molecular identifier (UMI) and a template-switching oligo (TSO) to capture full-length transcripts and identify barcodes for individual cells [71]. *RAGE-seq*, on the other hand, incorporates a 3′-adapter to capture the 3′-end of transcripts, which allows for full-length transcript reconstruction [72]. Philpott et al. developed *scCOLOR-seq*, a method that enables the correction of barcode and unique molecular identifier oligonucleotide sequences and permits standalone cDNA nanopore sequencing of single cells [73]. The authors showed the effectiveness of the method by accurately detecting cell-type-specific isoform expression and identifying previously undetected gene fusions in cancer cell lines. Rebboah et al. introduced *LR-Split-seq*, a method that utilises combinatorial barcoding to sequence single cells with long reads and accurately assign them to their cellular origin [74]. This method facilitates more accurate cell classification and has the advantage of detecting low-abundance isoforms that may be difficult to identify using short-read sequencing. These studies show that single-cell long-read sequencing has the potential to facilitate a more comprehensive understanding of the transcriptomes of individual cells, which could advance research areas such as precision medicine.

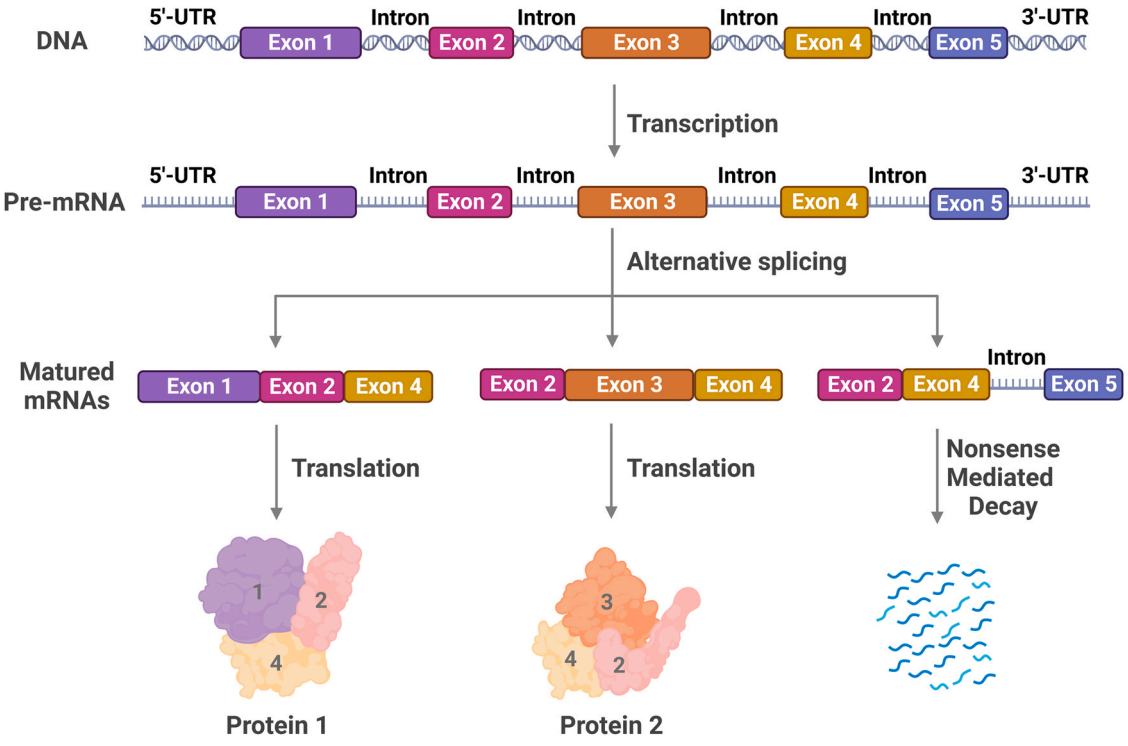

Fig. 1. Schema of alternative splicing. DNA encoding a gene is transcribed into pre-mRNA typically consisting of multiple exons. Alternative splicing leads to different mature mRNA transcripts (isoforms) including possible retained introns and different combinations of exons, which, after translation, produce multiple protein isoforms with potentially different functions. Created with BioRender.com.

Another approach to study cellular diversity and regulatory elements of cell type-specific gene expression is by analysing chromatin accessibility. A widely adopted protocol for this purpose is ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing). Recently, ATAC-seq has also been adopted at the single-cell level (scATAC-seq) [75–77], however, this approach has limitations, particularly when it comes to detecting large-scale structural variations and haplotype phasing. A recent study by Hu et al. demonstrated the utility of combining single-cell ATAC-seq with Nanopore third-generation genome sequencing, with their protocol named *scNanoATAC-seq* for investigating the relationship between chromatin accessibility and genome structure [78]. *scNanoATAC-seq* was shown to accurately capture *allele-specific* chromatin accessibility and detect large-scale structural variations in human cells. The method could be a valuable tool for exploring mechanisms of gene regulation and complex genetic diseases at the single-cell level.

Several tools have been developed for analysing and visualising isoform expression in single-cell long-read sequencing data. One example is the *FLAMES* pipeline developed by Tian et al., which combines the strengths of short-read and long-read sequencing to accurately identify and classify cells, detect novel low-abundance isoforms, analyse splicing events, and identify mutations [79]. It can be utilised with both bulk RNA-seq data and single-cell data. Moreover, *FLAMES* can detect differences in isoform expression between different cell types and discover cell-type-specific isoforms. Gorin and Pachter [80] used the chemical master equation to model transcriptional bursting and splicing processes, and also derived theoretical boundaries (constraints) on the correlation between two transcript counts. They have investigated a set of 500 genes from the single-cell long-read sequencing data obtained via the *FLAMES* pipeline, found that 95.3 % of intra-gene transcript-transcript correlations and 99.7 % of the inter-gene transcript-transcript correlations were consistent with these theoretical constraints (i.e., the sample correlation was less than or equal to the predicted correlation) [80]. However, technical limitations and the presence of intrinsic and extrinsic noise in biological systems prevent the model from fully capturing the dynamics of transcription and splicing processes. Mathematical models that take isoform information and biological noise during transcriptional and splicing processes into account to describe accurate mechanisms of splicing and cell-fate determination are yet to be developed but will certainly lead to a more systematic understanding of the process. Another example in single-cell long-read sequencing data analysis is the *ScisorWiz* R-package [81]. This tool can be used to visualise differential isoform expression, e.g. between cell clusters, by utilising single-cell long-read sequencing data. It can help identifying genes and isoforms that are differentially expressed in specific cell types, leading to new insights into cell-fate determination and the molecular mechanisms that regulate it.

However, a significant challenge in single-cell long-read sequencing analysis is the sparsity of the data. The occurrence of “dropout" events in single-cell long-read sequencing is more severe than in single-cell short-read sequencing. While single-cell long-read sequencing can detect a larger number of isoforms, the lower overall sequencing depth reduces the ability to accurately quantify isoform expression. The problem of sparsity in the data can create a high level of noise, which presents a significant challenge for mathematical modelling approaches, particularly differential equation models [82]. These models can be sensitive to sparsity, resulting in inaccurate predictions. Therefore, it is crucial to consider the impact of sparsity when attempting to model with single-cell long-read sequencing data. Additionally, efforts should be made to improve experimental protocols and develop tools to minimise the negative effects of sparsity on modelling accuracy.

Table 1
Some recent computational packages for trajectory inference, pseudotime analysis and visualisation from single-cell sequencing data.

Packages	Description	Platform	Reference
slingshot (2018)	Cluster-based minimum spanning tree modelling	R	[105]
velocyto (2018)	Original modelling method for estimating RNA velocity	Python and R	[106]
URD (2018)	Simulated diffusion-based modelling	R	[107]
PBA (2018)	Diffusion-drift based modelling	Python	[108]
Waddington-OT (2019)	Optimal transport based modelling	Python and Java	[102]
PAGA (2019)	Connectivity of manifold partitions based modelling	Python	[109]
pseudodynamics (2019)	Reaction-diffusion-advection PDE based modelling	MATLAB	[110]
Palantir (2019)	Markov chain based modelling	Python	[111]
scVelo (2020)	Likelihood based dynamical modelling for estimating RNA velocity	Python	[112]
CellRank (2022)	RNA velocity based toolkit for cell-fate mapping	Python	[113]
Dynamo (2022)	Dynamical systems based modelling for estimating RNA velocity	Python	[114]

PBA, population balance analysis; OT, optimal transport; PAGA, partition-based graph abstraction; PDE, partial differential equation.

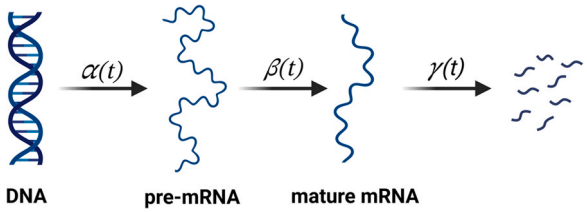

Fig. 2. Schema of mRNA synthesis and degradation. DNA is transcribed into pre-mRNA with a rate, $\alpha(t)$, at time t. The nascent RNA molecule (pre-mRNA) is further processed, via splicing, into mature mRNA with a rate, $\beta(t)$, at time t. Ultimately, the mRNA molecule undergoes degradation with a rate, $\gamma(t)$, at time t. Created with BioRender.com.

3. Modelling cell-fate determination with single-cell omics

3.1. Pseudotemporal analysis and developmental landscape

Pseudotemporal trajectory inference has been one of the grand challenges since single-cell sequencing was selected as Method of the Year in 2013 [82,83]. Single-cell data analysis methods focus largely on assessing transcriptomics profiles [84], often starting with a data matrix that contains read counts (i.e. mRNA expression data) of each gene in each cell. Methods are then applied for dimensionality reduction, trajectory inference, cell ordering and visualisation [85–88–91–94]. These methods use diverse mathematical concepts such as graph theory, statistics, probability theory, statistical mechanics, differential geometry and dynamical systems theory to map discrete snapshots of cell states into a low-dimensional continuous manifold that reflects the developmental process [91,95,96]. This manifold is called the developmental landscape, which relates to Waddington's epigenetic landscape [97,98]. Single-cell sequencing data enables the estimation and visualisation of transition pathways and fate choices from different cell states during development. Therefore, trajectory inference and developmental landscape reconstruction are strongly debated topics in the field of single-cell data analysis [91,96,99–102,103,104].

In Table 1, we provide an overview of widely used computational methods (published in the last five years) for trajectory inference, pseudotime analysis and landscape visualisation (see Table 1). Some of these tools will be discussed in more detail in the following sections. There are also several published reviews that summarise current trends in computational modelling with single-cell data. One example is the recent review by Teschendorff and Feinberg on using statistical mechanics for a systems-level analysis of single-cell data [96]. Comprehensive summaries of trajectory inference methods were published by Wagner and Klein [91] and Saelens et al. [95]. The latter includes a benchmark comparison of 45 commonly used trajectory inference methods based on four evaluation criteria: accuracy, scalability, stability and usability [95]. The authors concluded that choosing the right method depends on dataset dimensions and trajectory topology, and provide a useful guideline for researchers to choose the appropriate method for their available datasets.

3.2. RNA velocity

The concept of RNA velocity was introduced in 2018 by La Manno et al. [106]. RNA velocity, a model of transcriptional dynamics, determines changes in the relative abundance of unspliced and spliced mRNA to obtain information on the kinetics of gene expression. The model can be used to make predictions regarding the state and direction of cell differentiation. Both nascent and mature mRNA information can be obtained through current single-cell RNA sequencing protocols.

Existing RNA velocity approaches mathematically model the processes of transcription from DNA into pre-mRNA, followed by splicing into mature mRNA, and finally mRNA degradation (see Fig. 2). The original steady-state model assumes time-independent transcription and degradation rates, namely $\alpha(t) = \alpha$ and $\gamma(t) = \gamma$, and a constant unit splicing rate, namely $\beta(t) = 1$ across all genes. *Cis*- and *trans*-regulatory mechanisms of gene expression are generally not considered. The changes in the abundances of unspliced, $U(t)$, and spliced, $S(t)$, mRNA at time t are modelled with a system of ordinary differential equations (ODEs) [106]:

$$\begin{aligned}\frac{dU(t)}{dt} &= \alpha(t) - \beta(t)U(t)\\ &= \alpha - U(t),\end{aligned} \tag{1}$$

$$\begin{aligned}\frac{dS(t)}{dt} &= \beta(t)U(t) - \gamma(t)S(t)\\ &= U(t) - \gamma S(t).\end{aligned} \tag{2}$$

These equations can be solved with initial conditions $U(0) = u_0$ and $S(0) = s_0$,

$$U(t) = u_0 e^{-t} + \alpha(1 - e^{-t}), \tag{3}$$

$$S(t) = s_0 e^{-\gamma t} + \frac{\alpha}{\gamma}(1 - e^{-\gamma t}) + \frac{u_0 - \alpha}{\gamma - 1}(e^{-t} - e^{-\gamma t}). \tag{4}$$

Setting Eqs. (1) and (2) equal to zero, we can determine the steady state solution

$$U_{ss}(t) = \gamma S(t). \tag{5}$$

In this case, the RNA velocity is estimated as the absolute difference between the observed state and the steady state, that is,

$$\begin{aligned}r(t) &= |U(t) - U_{ss}(t)|\\ &= |U(t) - \gamma S(t)|\\ &= \left| \frac{e^{-t}(\alpha - u_0) - e^{-\gamma t}(\alpha + \gamma(\gamma s_0 - s_0 - u_0))}{\gamma - 1} \right|.\end{aligned} \tag{6}$$

The developers of the RNA velocity concept have experimentally verified the predictive power of their model by accurately predicting

the cell state in the neural crest lineage, which uncovered the branching lineage tree of the developing mouse hippocampus and examined the transcriptional dynamics in human embryonic brain [106]. Nevertheless, their model assumptions are also highly susceptible to uncertainties. For example, the method is based on the steady-state assumption. However, the reality is that there is no guarantee that a steady state always exists in the observed experimental data. Since it is difficult to estimate the actual transcription and splicing rates, the authors proposed two alternative assumptions to their RNA velocity estimation: .

1. assumes that the rate of change of the number of spliced mRNAs, $dS(t)/dt$, remains constant, whereby the solution of $S(t)$ will be a linear function.
2. assumes that the number of unspliced mRNAs, $U(t)$ is a constant so that the system of ODEs becomes a single variable equation. This would allow the RNA velocity to be estimated without considering the transcription rate.

La Manno et al. implemented their model in a software tool called *velocyto*, together with methods such as K-nearest neighbour pooling and t-distributed stochastic neighbour embedding to visualise the RNA velocity vector field, which can help describe cell-fate decisions.

To address the limitations associated with the steady-state assumption and the unit splicing rate across all genes, Bergen et al. proposed a new method for estimating RNA velocity [112]. Their model assumes a constant splicing rate for each gene rather than all genes having a unit splicing rate. The second key change is to set the transcription rate to a cell-specific latent variable, $\alpha(t) = \alpha^{k_i}(t_i)$, where k_i is the DNA transcriptional state for the i-th observation, namely an induction phase ($k = 1$) and a repression phase ($k = 0$) with an ON state (ss_{on}) and an OFF state (ss_{off}) for each phase. In addition, the term t_i represents the latent time for the i-th observation. Thus, changes of the abundance of unspliced, $U(t_i)$, and spliced, $S(t_i)$, mRNA at time t_i for the i-th observation are modelled with a system of ODEs as follows [112]:

$$\frac{dU(t_i)}{dt_i} = \alpha^{k_i}(t_i) - \beta U(t_i), \tag{7}$$

$$\frac{dS(t_i)}{dt_i} = \beta U(t_i) - \gamma S(t_i). \tag{8}$$

They define the RNA velocity, $r(t)$, as the change of abundance of spliced mRNA. That is,

$$r(t) = \frac{dS(t)}{dt}. \tag{9}$$

Next, Bergen's method conducts parameter estimation using the Expectation-Maximisation (EM) algorithm by minimising the distance between the observed mRNA value and current phase trajectories to determine the latent time and other parameters for each cell. Finally, the transition probability of each cell is computed from the estimated RNA velocity and then mapped to a low-dimensional space using uniform manifold approximation and projection to visualise the cell differentiation trajectory. The authors implemented these methods in a software called *scVelo* [112].

Although their method relaxes the steady-state and unit splicing rate assumptions of the original RNA velocity model, it still simulates the transcriptional process with deterministic linear models, which cannot capture the non-linearity, heterogeneity and stochasticity of gene expression in individual cells. It also continues to assume that the splicing and degradation rates are time-independent and that genes are independent from each other. Therefore, there are still opportunities to further advance single-cell-based RNA velocity modelling. A detailed summary of current challenges and future directions of RNA velocity modelling is provided by Bergen et al. [115].

Apart from *scVelo*, many other methods have been proposed based on the RNA velocity idea. For example, Qiu et al. developed a method, *Dynamo* [114], which can be used to infer absolute RNA velocity and reconstruct differentiation landscapes that predict cell-fate and explore the underlying mechanism from time-resolved metabolically labelled single-cell RNA sequencing data. Lange et al. used a Markov-chain-based modelling to simulate cell state transitions based on RNA velocity and the stochasticity in cell-fate determination [113]. Their method, *CellRank*, can be used for reconstructing developmental landscapes and inferring differentiation trajectories and reprogramming pathways.

Cell-fate prediction based on the RNA velocity concept is not limited to transcriptomics. The concept can be extended based on single-cell multi-omics data, e.g. involving proteomics, metabolomics or epigenomics data. For example, Gennady et al. combined single-cell mRNA and protein expression data to predict protein velocity-based cell-fate decisions [116]. Tedesco et al. improved cell-fate predictions via chromatin velocity modelling based on single-cell genome and epigenome by transposases sequencing (scGET-seq) [117]. Li et al. developed a model called *MultiVelo*, which improved cell-fate predictions by integrating both transcriptomics and epigenomics data for single-cell velocity estimation [118]. Assuming that the dynamics of chromatin opening and closing are mirrored, they adapted Bergen's RNA velocity model [112] to determine chromatin velocity.

$$\frac{dc(t)}{dt} = k_c\alpha_c - \alpha_c c(t), \tag{10}$$

where k_c corresponds to the chromatin state with the OPENING state (k_c = 1) and the CLOSING state (k_c = 0), and parameter α_c is the chromatin rate. Their methods achieved a better cell-fate prediction than other velocity methods based on transcriptomics data.

4. Challenges and opportunities: modelling cell-fate determination with single-cell long-read sequencing data

As mentioned before, it is now possible to perform long-read sequencing at the single-cell level. Using this technology, we have the opportunity to gain deeper insights into mechanisms of cell-fate determination, e.g. by characterising cell differentiation pathways via transcriptome diversity and isoform expression patterns in individual cells. However, there are currently very few mathematical approaches for cell-fate determination that take advantage of single-cell long-read sequencing data.

4.1. How can single-cell long-read sequencing enhance RNA velocity modelling?

Single-cell long-read sequencing can relax some of the RNA velocity model assumptions by providing more accurate and comprehensive data on gene expression dynamics. The current RNA velocity model relies on the assumption that the splicing rate of pre-mRNAs remains constant over time [106]. This allows the ratio of unspliced to spliced reads to be used for inferring the directed differentiation trajectories of individual cells. However, this assumption does not hold true for genes with complex splicing patterns, where different isoforms may have different splicing rates.

Long-read sequencing provides a good representation of isoform diversity and with sufficient sequencing depth accurate quantification of isoform expression levels, which can help to better estimate the rate of splicing and degradation for each isoform. *MAS-ISO-seq* developed by Al'Khafaji et al. can generate high-depth single-cell long-read sequencing data for single-cell isoform analysis, which also enables the detection of low-abundance transcripts and the

identification of rare isoforms [119]. The data obtained from *MAS-ISO-seq* includes an isoform expression count matrix, which provides information on the abundance of different isoforms, including both novel and annotated isoforms, across cells. This allows to estimate the splicing rate of pre-mRNAs more accurately for different isoforms. Furthermore, isoform expression count matrices can be analysed using common single-cell analysis tools, such as *Seurat*, for clustering, dimensionality reduction, and visualisation [119]. These tools can aid pseudotemporal analysis, as well as in the identification of isoform expression patterns across various cell types, providing valuable insights into the functional diversity of cells.

The RNA velocity model distinguishes newly transcribed unspliced pre-mRNA from mature spliced mRNA to measure changes in mRNA abundance. However, this binary classification of transcripts based on the presence of introns is often contradicted by a phenomenon known as intron retention. Intron retention has been shown to be widespread in mature mRNA transcripts and involved in the regulation of cell differentiation [63,120]. Therefore, improvement in inferring splicing rates using long-read sequencing can be achieved by detecting whether mRNA contains a poly-A tail at the 3′-end and thus determining whether the splicing process is complete. A recent study, proposed the first model of alternative splice site selection and recursive intron removal based on long-read sequencing data [67]. It suggests that there are many intermediate states between newly transcribed nascent RNA and mature fully spliced RNA, as also pointed out by Gorin et al. [121]. Thus, by leveraging long-read sequencing data analysis tools like *FLAMES* for splicing and isoform analysis at the single-cell level, we can gain deeper insights into isoform expression dynamics, allowing for the integration of splicing mechanisms to construct more comprehensive mathematical models of cell-fate determination.

4.2. Modelling cell-fate determination at isoform level

Studies have shown that different mRNA isoforms are produced as a result of alternative splicing, which drives cell differentiation and development [64,120]. Given that different genes can produce varying numbers of isoforms across different cell types, each cell type therefore has a distinct transcriptome diversity, which is a significant indicator of a cell's differentiation potency [122]. Long-read data enables us to identify the transcriptome diversity of individual cells as well as the expression levels of cell-specific gene isoforms. This capability enables the prediction of future cell states by analysing the differential expression across cell states. The biggest challenge is that new single-cell long-read sequencing protocols detected plenty of unannotated isoforms. For mathematical modelling, it is difficult to account for all unannotated isoforms. Therefore, it is essential that the concept of transcriptome diversity be rigorously defined. To date, there have only a few mathematical models been develeoped that incorporate single-cell transcriptome diversity. Gulati et al. developed *CytoTRACE*, a framework that predicts cell differentiation states based on the number of expressed genes in each cell [122]. García-Nieto et al. leveraged Shannon entropy to describe the transcriptome diversity to explain variability in gene expression [123]. Moreover, identifying the splicing mechanisms that cause transcriptome diversity is essential for improving efficiency and accuracy in modelling cell-fate determination and formulating valid model assumptions and parameter settings. Furthermore, benchmarking analyses are required for various single-cell long-read sequencing protocols before adopting datasets for model development and parameter estimation.

5. Outlook and conclusion

Long-read sequencing can also be used in the context of single-cell spatial transcriptomics. Single-cell spatial transcriptomics has sharpened our understanding of spatially resolved tissue composition [124,125]. However, existing short-read sequencing methods cannot provide isoform expression and transcriptome diversity information within a given tissue [126], which could be resolved by introducing long-read sequencing into spatial transcriptomics protocols. Lebrigand et al. proposed a novel method, *Spatial Isoform Transcriptomics (SiT)*, for characterising spatial isoform information from nanopore sequencing data [126]. Boileau et al. developed a software, *scNaST*, for analysing spatial gene expression from both short-read and long-read sequencing data to explore isoform diversity within a given tissue [127]. Both of these studies employed deconvolution methods from single-cell data analysis and applied these to spatial transcriptomics data. Deconvolution methods are crucial for accurately interpreting spatial transcriptomics data, enabling the identification of cell types and spatial heterogeneity. They are essential for advancing our understanding of the spatial organisation of cells in tissues and organs, with implications for developing new therapeutic strategies. Benchmark analyses are necessary to determine the suitability of applying deconvolution methods from short-read sequencing to long-read sequencing data. Additionally, it is necessary to develop new methods for improving the accuracy and efficiency of computational tools for processing and analysing single-cell long-read spatial transcriptomics datasets. These advancements will enable researchers to more precisely decipher the spatial organisation and cellular heterogeneity of complex tissues and organs.

In conclusion, long-read sequencing technology has tremendous potential for enabling investigations into the underlying mechanisms of biological systems. The development of stable, reproducible, and accurate computational tools and mathematical models to handle single-cell long-read sequencing data will be a major focus in the years ahead. Novel tools and models may help to further dissect the mechanisms underlying development and cell differentiation and to better understand the role of transcriptome diversity, particularly through alternative splicing, in these processes. This knowledge could ultimately lead to the discovery of new stem cell and regenerative therapeutic strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

U.S. received support from the National Health and Medical Research Council (Investigator Grant #1196405), the Cancer Council NSW (project grant RG20-12), and the Tropical Australian Academic Health Centre (project grant SF0000321).

Author statement

S.W. and U.S. conceived the topic and structure. S.W. drafted the manuscript. U.S. supervised the work. Both authors revised the paper and approved the final version.

References

[1] Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell-cell interactions and communication from gene expression. Nat Rev Genet 2021;22:71–88.
[2] Madrigal P, Deng S, Feng Y, Militi S, Goh KJ, Nibhani R, et al. Epigenetic and transcriptional regulations prime cell fate before division during human pluripotent stem cell differentiation. Nat Commun 2023;14:405.
[3] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009;6:377–82.

[4] Wright DJ, Hall NAL, Irish N, Man AL, Glynn W, Mould A, et al. Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. BMC Genom 2022;23:42.
[5] Ding C, Yan X, Xu M, Zhou R, Zhao Y, Zhang D, et al. Short-read and long-read full-length transcriptome of mouse neural stem cells across neurodevelopmental stages. Sci Data 2022;9:69.
[6] Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics 2019;35:2193–8.
[7] Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol 2021;39:1348–65.
[8] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. DNA sequencing from single polymerase molecules. Science 2009;323:133–8.
[9] Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol 2008;26:1146–53.
[10] Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 2016;17:239.
[11] Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Hum Mol Genet 2018;27:R234–41.
[12] Method of the Year 2022: long-read sequencing; 2023.
[13] Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet 2020;21:597–614.
[14] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020;21:30.
[15] De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. Nat Rev Genet 2021;22:572–87.
[16] Au KF. The blooming of long-read sequencing reforms biomedical research. Genome Biol 2022;23:21.
[17] Liu Q, Hu Y, Stucky A, Fang L, Zhong JF, Wang K. LongGF: computational algorithm and software tool for fast and accurate detection of gene fusions by long-read transcriptome sequencing. BMC Genom 2020;21:793.
[18] Davidson NM, Chen Y, Sadras T, Ryland GL, Blombery P, Ekert PG, et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. Genome Biol 2022;23:10.
[19] Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. Nat Methods 2016;13:751–4.
[20] David M, Dursi LJ, Yao D, Boutros PC, Simpson JT. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. Bioinformatics 2016;33:49–55.
[21] Boža V, Brejová B, Vinar^ T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. PLoS One 2017;12:1–13.
[22] Verhey TB, Castellanos M, Chaconas G. Analysis of recombinational switching at the antigenic variation locus of the Lyme spirochete using a novel PacBio sequencing pipeline. Mol Microbiol 2018;107:104–15.
[23] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 2013;29:1072–5.
[24] De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 2018;34:2666–9.
[25] Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 2018;35:523–5.
[26] Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, DelRisco H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res 2018;28:396–411.
[27] Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10:563–9.
[28] Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 2016;13:1050–4.
[29] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;27:722–36.
[30] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 2019;37:540–6.
[31] Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol 2019;20:278.
[32] Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods 2020;17:155–8.
[33] Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. Nat Biotechnol 2020;38:1044–53.
[34] English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. BMC Bioinform 2014;15:180.
[35] Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature 2015;517:608–11.
[36] Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, vonHaeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15:461–8.
[37] Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics 2019;35:2907–15.
[38] Mahmoud M, Doddapaneni H, Timp W, Sedlazeck FJ. PRINCESS: comprehensive detection of haplotype resolved SNVs, SVs, and methylation. Genome Biol 2021;22:268.
[39] Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 2017;14:407–10.
[40] Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, et al. Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods 2017;14:411–3.
[41] Ni P, Huang N, Zhang Z, Wang D-P, Liang F, Miao Y, et al. DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. Bioinformatics 2019;35:4586–95.
[42] Liu H, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, et al. Accurate detection of m6A RNA modifications in native RNA sequences. Nat Commun 2019;10:4079.
[43] Jenjaroenpun P, Wongsurawat T, Wadley TD, Wassenaar TM, Liu J, Dai Q, et al. Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res 2020;49. e7-e7.
[44] Gao Y, Liu X, Wu B, Wang H, Xi F, Kohnen MV, et al. Quantitative profiling of N^6-methyladenosine at single-base resolution in stem-differentiating xylem of Populus trichocarpa using Nanopore direct RNA sequencing. Genome Biol 2021;22:22.
[45] Abebe JS, Price AM, Hayer KE, Mohr I, Weitzman MD, Wilson AC, et al. DRUMMER-Rapid detection of RNA modifications through comparative nanopore sequencing. Bioinformatics 2022;38:3113–5.
[46] Sahlin K, Tomaszkiewicz M, Makova KD, Medvedev P. Deciphering highly similar multigene family transcripts from Iso-Seq data with IsoCon. Nat Commun 2018;9:4601.
[47] Annaldasula S, Gajos M, Mayer A. IsoTV: processing and visualizing functional features of translated transcript isoforms. Bioinformatics 2021;37:3070–2.
[48] You Y, Clark MB, Shim H. NanoSplicer: Accurate identification of splice junctions using Oxford Nanopore sequencing. Bioinformatics 2022;38:3741–8.
[49] Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, et al. Transcriptome variation in human tissues revealed by long-read sequencing. Nature 2022;608:353–9.
[50] Zhu C, Wu J, Sun H, Briganti F, Meder B, Wei W, et al. Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. Nat Commun 2021;12:4203.
[51] Hu Y, Fang L, Chen X, Zhong JF, Li M, Wang K. LIQA: long-read isoform quantification and analysis. Genome Biol 2021;22:182.
[52] Aw JGA, Lim SW, Wang JX, Lambert FRP, Tan WT, Shen Y, et al. Determination of isoform-specific RNA structure with nanopore long reads. Nat Biotechnol 2021;39:336–46.
[53] Leung SK, Jeffries AR, Castanho I, Jordan BT, Moore K, Davies JP, et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. Cell Rep 2021;37.
[54] Veiga DFT, Nesta A, Zhao Y, Mays AD, Huynh R, Rossi R, et al. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv 2022;8:eabg6711.
[55] Dong X, Du MRM, Gouil Q, Tian L, Baldoni PL, Smyth GK, et al., Benchmarking long-read RNA-sequencing analysis tools using in silico mixtures, bioRxiv; 2022.
[56] Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. Nat Rev Genet 2018;19:329–46.
[57] Wan YK, Hendra C, Pratanwanich PN, Göke J. Beyond sequencing: machine learning algorithms extract biology hidden in Nanopore signal data. Trends Genet 2022;38:246–57.
[58] Amarasinghe SL, Ritchie ME, Gouil Q. Long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. GigaScience 2021;10.
[59] Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. Cell 2016;166:1147–1162.e15.
[60] Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet 2016;17:19–32.
[61] Xu Y, Zhao W, Olson SD, Prabhakara KS, Zhou X. Alternative splicing links histone modifications to stem cell fate decision. Genome Biol 2018;19:133.
[62] Jiang W, Chen L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. Comput Struct Biotechnol J 2021;19:183–95.
[63] Wong JJ-L, Schmitz U. Intron retention: importance, challenges, and opportunities. Trends Genet 2022;38:789–92.
[64] Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 2017;18:437–51.
[65] de la Fuente L, Arzalluz-Luque Á, Tardáguila M, del Risco H, Martí C, Tarazona S, et al. tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. Genome Biol 2020;21:119.
[66] Prjibelski AD, Mikheenko A, Joglekar A, Smetanin A, Jarroux J, Lapidus AL, et al. Accurate isoform discovery with IsoQuant using long reads. Nat Biotechnol 2023.
[67] Wan Y, Anastasakis DG, Rodriguez J, Palangat M, Gudla P, Zaki G, et al. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell 2021;184:2878–2895.e20.
[68] Miller RM, Jordan BT, Mehlferber MM, Jeffery ED, Chatzipantsiou C, Kaur S, et al. Enhanced protein isoform characterization through long-read proteogenomics. Genome Biol 2022;23:69.

[69] Arzalluz-Luque Á, Conesa A. Single-cell RNAseq for the study of isoforms—how is that possible? Genome Biol 2018;19:110.
[70] Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 2014;9:171–81.
[71] Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol 2018;36:1197–202.
[72] Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun 2019;10:3120.
[73] Philpott M, Watson J, Thakurta A, Brown T, Oppermann U, Cribbs AP. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. Nat Biotechnol 2021;39:1517–20.
[74] Rebboah E, Reese F, Williams K, Balderrama-Gutierrez G, McGill C, Trout D, et al. Mapping and modeling the genomic basis of differential RNA isoform expression at single-cell resolution with LR-Split-seq. Genome Biol 2021;22:286.
[75] Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348:910–4.
[76] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–90.
[77] Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS, et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 2019;37:916–24.
[78] Hu Y, Jiang Z, Chen K, Zhou Z, Zhou X, Wang Y, et al. scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell. Cell Res 2023;33:83–6.
[79] Tian L, Jabbari JS, Thijssen R, Gouil Q, Amarasinghe SL, Voogd O, et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. Genome Biol 2021;22:310.
[80] Gorin G, Pachter L. Modeling bursty transcription and splicing with the chemical master equation. Biophys J 2022;121:1056–69.
[81] Stein AN, Joglekar A, Poon C-L, Tilgner HU. ScisorWiz: visualizing differential isoform expression in single-cell long-read data. Bioinformatics 2022;38:3474–6.
[82] Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biol 2020;21:31.
[83] Method of the Year 2013; 2013.
[84] Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol 2018;14:e8046.
[85] Pierson E, Yau C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol 2015;16:1–10.
[86] Ahmed S, Rattray M, Boukouvalas A. GrandPrix: scaling up the Bayesian GPLVM for single-cell data. Bioinformatics 2019;35:47–54.
[87] Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9.
[88] L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint(2018).
[89] Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IW, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2019;37:38–44.
[90] Campbell KR, Yau C. Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. Nat Commun 2018;9:2442.
[91] Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet 2020;21:410–27.
[92] Sun S, Zhu J, Ma Y, Zhou X. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis. Genome Biol 2019;20:269.
[93] Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 2019;15:e8746.
[94] Liu R, Pisco AO, Braun E, Linnarsson S, Zou J. Dynamical systems model of RNA velocity improves inference of single-cell trajectory, pseudo-time and gene regulation. J Mol Biol 2022;434:167606.
[95] Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol 2019;37:547–54.
[96] Teschendorff AE, Feinberg AP. Statistical mechanics meets single-cell biology. Nat Rev Genet 2021;22:459–76.
[97] Waddington C. A discussion of some aspects of theoretical biology. In: Waddington CH, editor. The Strategy of the Genes. London: London George Allen and Unwin; 1957.
[98] Waddington CH. Principles of development and differentiation. Macmillan; 1966.
[99] Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature 2018;553:418–26.
[100] Shi J, Teschendorff AE, Chen W, Chen L, Li T. Quantifying Waddingtonas epigenetic landscape: a comparison of single-cell potency measures. Brief Bioinforma 2018.
[101] Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496–502.
[102] Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell 2019;176:928–943.e22.
[103] Li Y, Jiang Y, Paxman J, O'Laughlin R, Klepin S, Zhu Y, et al. A programmable fate decision landscape underlies single-cell aging in yeast. Science 2020;369:325–9.
[104] Wang S, Lee MP, Jones S, Liu J, Waldhaus J. Mapping the regulatory landscape of auditory hair cells from single-cell multi-omics data. Genome Res 2021;31:1885–99.
[105] Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genet 2018;19:477.
[106] LaManno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature 2018;560:494–8.
[107] Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 2018;360.
[108] Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci USA 2018;115:E2467–76.
[109] Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 2019;20:59.
[110] Fischer DS, Fiedler AK, Kernfeld EM, Genga RMJ, Bastidas-Ponce A, Bakhti M, et al. Inferring population dynamics from single-cell RNA-sequencing time series data. Nat Biotechnol 2019;37:461–8.
[111] 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;37:451–60.
[112] Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 2020;38:1408–14.
[113] Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, et al. CellRank for directed single-cell fate mapping. Nat Methods 2022;19:159–70.
[114] Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell 2022;185:690–711.e45.
[115] Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity–current challenges and future perspectives. Mol Syst Biol 2021;17:e10282.
[116] Gorin G, Svensson V, Pachter L. Protein velocity and acceleration from single-cell multiomics experiments. Genome Biol 2020;21.
[117] Tedesco M, Giannese F, Lazarević D, Giansanti V, Rosano D, Monzani S, et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. Nat Biotechnol 2022;40:235–44.
[118] Li C, Virgilio MC, Collins KL, Welch JD. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. Nat Biotechnol 2022.
[119] Al'Khafaji AM, Smith JT, Garimella KV, Babadi M, Sade-Feldman M, Gatzen M, et al., RNA isoform sequencing using programmable cDNA concatenation, bioRxiv; 2021, 2021.10.01.462818.
[120] Green ID, Pinello N, Song R, Lee Q, Halstead JM, Kwok C-T, et al. Macrophage development and activation involve coordinated intron retention in key inflammatory regulators. Nucleic Acids Res 2020;48:6513–29.
[121] Gorin G, Fang M, Chari T, Pachter L. RNA velocity unraveled. PLoS Comput Biol 2022;18:e1010492.
[122] Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. Science 2020;367:405–11.
[123] García-Nieto PE, Wang B, Fraser HB. Transcriptome diversity is a systematic source of variation in RNA-sequencing data. PLOS Comput Biol 2022;18:e1009939.
[124] Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat Rev Genet 2021;22:627–44.
[125] Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. Genome Med 2022;14:68.
[126] Lebrigand K, BergenstrÅhle J, Thrane K, Mollbrink A, Barbry P, Waldmann R, et al., The spatial landscape of gene expression isoforms in tissue sections, bioRxiv; 2020.
[127] Boileau E, Li X, Naarmann-de Vries IS, Becker C, Casper R, Altmüller J, et al. Full-length spatial transcriptomics reveals the unexplored isoform diversity of the myocardium Post-MI. Front Genet 2022;13.