
Single-Cell RNA Sequencing Procedures and Data Analysis

Markus Wolfien¹ • Robert David^{2,3} • Anne-Marie Galow⁴

¹Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany; ²Department of Cardiac Surgery Rostock University Medical Center, Rostock, Germany; ³Department Life, Light & Matter of the Interdisciplinary Faculty, University of Rostock, Germany; ⁴Institute of Genome Biology, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany

Author for correspondence: Robert David, Department of Cardiac Surgery Rostock University Medical Center, Rostock, Germany. Email: robert.david@med.uni-rostock.de

Doi: <https://doi.org/10.36255/exonpublications.bioinformatics.2021.ch2>

Abstract: Single-cell and single-nuclei sequencing experiments reveal previously unseen molecular details. The number of sequencing procedures and computational data analysis approaches have been increasing rapidly in recent years. This chapter provides an overview of the current developments in single-cell analysis. An introduction and practical guidance for choosing the most suitable sequencing procedure to match individual experimental demands in the course of investigating biological hypotheses are presented. Basic data analysis approaches are highlighted, followed by a discussion on advanced downstream approaches to enrich the information obtained from single-cell experiments; for example, trajectory analysis, pseudotime assumptions, and network inference. Currently unsolved challenges are discussed to allow the reader to avoid the most common pitfalls.

Keywords: network integration; pseudotemporal reconstruction; RNA velocity; single-cell RNA-sequencing; scRNA-seq data

In: *Bioinformatics*. Nakaya HI (Editor). Exon Publications, Brisbane, Australia.

ISBN: 978-0-6450017-1-6; Doi: <https://doi.org/10.36255/exonpublications.bioinformatics.2021>

Copyright: The Authors.

License: This open access article is licenced under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) <https://creativecommons.org/licenses/by-nc/4.0/>

INTRODUCTION

A typical single-cell sequencing workflow involves initial tissue and cell preparation, cell capturing and library preparation, sequencing and raw data processing, as well as visualization and downstream analyses. A plethora of protocols are available for the preparation of single-cell suspensions because the optimal procedure differs for each tissue and cell type to be isolated (1–3). Hence, this chapter focusses on the steps of data generation and data analysis, with emphasis on various capture and sequencing techniques, which are foundations for the subsequent computational data analyses.

THE CAPTURE TECHNIQUE SIGNIFICANTLY DETERMINES THE QUANTITY OF MEASURABLE CELLS

Several protocols for single cell RNA-sequencing (scRNA-seq) have been published over the last few years, and it remains a rapidly evolving field (4). The capture technique determines throughput, sorting options, and the type of additional information that can be obtained. The most widely used options are microwell- and droplet-based. These techniques differ in their strategies for tagging transcripts based on cell origin, and in the ways libraries are generated for sequencing. Table 1 summarizes the most common techniques to date and provides an overview of their main characteristics.

TABLE 1
Single-cell sequencing techniques

Technique	Method	Detected cells	Sensitivity	Costs/cell	Time
CEL-seq2 Hashimshony <i>et al.</i> (5)	Microwell-based	< 400	Very high	~ 3	25h
Drop-seq Macosko <i>et al.</i> (6)	Droplet-based	5,000 – 10,000	Moderate	< 0.1€	10h
ICELL8 Goldstein <i>et al.</i> (7)	Microwell-based	1,000 – 1,800	NA	NA	NA
InDrop-seq Klein <i>et al.</i> (8)	Droplet-based	5,000 – 10,000	Moderate	< 0.1€	10h
MARS-seq Jaitin <i>et al.</i> (9)	Microwell-based	100 – 1,000	Low	~ 0.50€	NA
Seq-well Gierahn <i>et al.</i> (10)	Microwell-based	> 10,000	Moderate	< 0.1€	10h
SmartSeq2 Picelli <i>et al.</i> (11)	Microwell-based	< 400	Very high	~ 10€	25h
10x Genomics Chromium Zheng <i>et al.</i> (12)	Droplet-based	1,000 – 10,000	High	~ 0.25€	9h

The most common single cell sequencing techniques listed together with protocol-dependent key data, such as the range of detectable cell numbers, sensitivity in terms of gene detection rates, and economic factors (average costs, amount of time to complete the procedure). NA, not available.

Microwell-based techniques allow for visual inspection

For well-based platforms, cells are usually transferred into micro- or nano-well plates using pipette or laser capture methods, such as fluorescent activated cell sorting (FACS) based on surface markers. This option renders well-based platforms particularly useful when isolation of a specific subset of cells is required, for example to explore rare cell types. Another advantage is the ability to visually inspect captured cells, allowing for identification of wells containing damaged or no cells and/or providing additional morphological information. The main drawback of well-based platforms is that they are often low-throughput and require a considerable amount of hands-on work per cell in contrast to other methods. These drawbacks are overcome to some extent by utilization of microfluidic platforms, such as Fluidigm C1 (13), which can be integrated in the workflow of some microwell-based platforms, providing a higher throughput. However, only around 10% of cells are typically captured in a microfluidic platform, rendering it inappropriate for the detection of rare-cell types. The C1 system also allows for visual inspection under the microscope, thereby enabling the user to exclude empty wells and wells containing damaged cells or doublets prior to downstream library preparation. The high cost of the microfluidic cartridges can limit the sample size used in each project, but expenses can be reduced on reagents since reactions can be carried out in a smaller volume.

Droplet-based techniques allow for high throughput

Droplet-based methods use microfluidics to encapsulate each individual cell together with a bead inside a nanoliter droplet that includes specific enzymes required to construct the library. The bead carries primers with a unique barcode, which bind the cell's mRNA and thus will be attached to all reads originating from that cell. All droplets can be pooled to produce a sequencing library. After sequencing, the reads can be assigned to the cell of origin based on the barcodes. Since the library preparation costs are comparably low, and the downstream processes are less elaborate due to the pooling step, droplet platforms typically have the highest throughput. Usually, the costs for the subsequent sequencing become the limiting factor, so that in typical experiments the coverage is rather low with only a few thousand different transcripts detected per cell. One major drawback is that these protocols offer little control over the cell input and thus are susceptible to selection bias, leading to inaccurate reflection of the biology of the studied system.

THE SEQUENCING TECHNIQUE DETERMINES THE OPTIONS FOR DATA ANALYSES

Once single-cell resolution is achieved via one of the approaches mentioned above, the individual transcriptomes must be sequenced. There are two main forms of sequencing techniques: full-length and tag-based protocols. Full-length based protocols aim to achieve a uniform read coverage of each transcript, whereas tag-based protocols only capture either the 5'- or 3'-end of each RNA molecule.

The choice of capture and quantification method has important implications on the types of analyses the data can be used for.

Tag-based protocols can be combined with unique molecular identifiers (UMIs), which permit multiplexing and improve quantification. However, the restriction to one end of the transcript may hamper the alignment and renders these protocols unsuitable for studies on allele-specific expression or isoform usage (14). To diminish these limitations, paired-end sequencing can be conducted, which involves sequencing both ends of cDNA fragments in a library and aligning the forward and reverse reads as read pairs. This procedure facilitates the detection of genomic rearrangements, such as insertions, deletions, and inversions, allowing for the discovery of gene fusions, novel transcripts, and novel splice isoforms. Moreover, tag-based methods have been established that are able to detect the co-occurrence of a specific transcription start site and a polyadenylation site (15). However, the generation of full-length cDNAs from very long transcripts still poses a technical limitation for any 5'-3'-sequencing method.

In contrast, full-length protocols provide an even coverage of transcripts and are suitable for the discovery of alternative-splicing events and allele-specific expression using single-nucleotide polymorphisms. A disadvantage of these protocols is that it is not possible to incorporate UMIs and barcodes for exact gene level quantification or multiplexing, leading to increased complexity of downstream processing.

For the sequencing step, the Illumina platform is widely used (e.g., HiSeq4000, NextSeq500, and NovaSeq™6000), being responsible for more than 90% of the world's sequencing data (16). All Illumina platforms use a sequencing by synthesis approach, yielding reliable base calls for highly repetitive sequences.

HOW TO CHOOSE THE RIGHT APPROACH?

The choice of method depends primarily on the individual scientific question and is further influenced by the compromise between cell numbers, information depth, and overall cost. For example, a droplet-based method will be most suitable for the characterization of the composition of a tissue, since it allows for large numbers of cells to be captured. In contrast, for in-depth analysis of rare cell types, it is probably best to enrich them using FACS, if there is a known surface marker, and then sequencing a smaller number of cells. Full-length transcript quantification will be more appropriate for studying different isoforms, since tagged protocols are much more limited. By contrast, UMIs can facilitate gene-level quantification, but they can only be used with tagged protocols. The low cost and high throughput of tag-based approaches has led to their widespread application in studies of gene expression levels, cell-type discovery, and tissue composition. However, it is recommended to consider the different techniques critically before starting the experiment. Svensson *et al.* compared the accuracy and sensitivity of different protocols and reported substantial differences between them (17). Figure 1 is based on current benchmarking studies (18-20) and should serve as a rough orientation for decision-making between the different experimental techniques based on individual experimental demands.

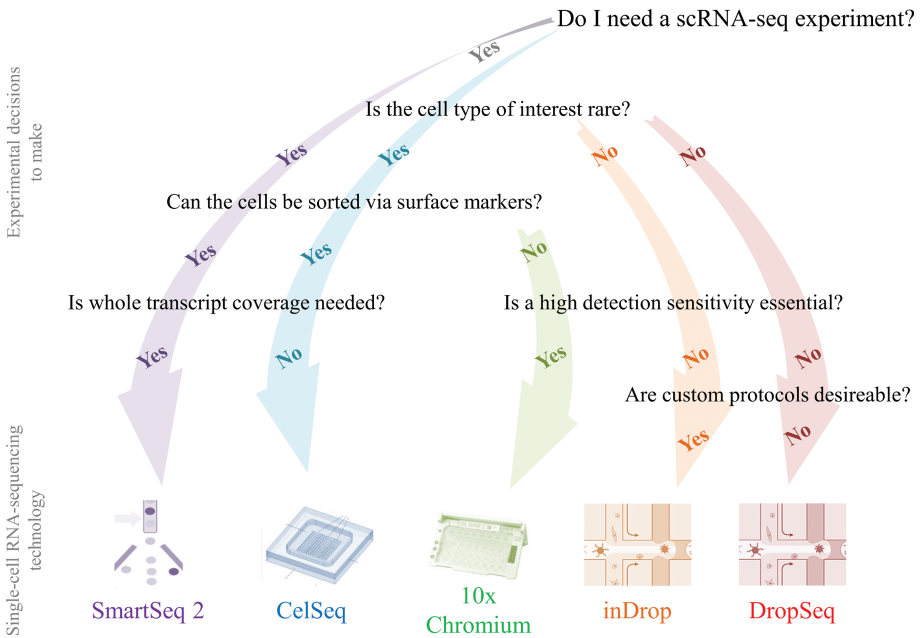


Figure 1. Flow chart guide to the most suitable scRNA-seq technique depending on the scientific problem and the boundary conditions.

COMPUTATIONAL ANALYSIS

In the following, state-of-the-art computational components of scRNA-seq data analysis are presented, and underlying methods are discussed to contribute an update to previous single-cell analysis reviews (21, 22). We highlight well-executed benchmarking studies for additional in-depth reading and seek to guide new users through the landscape of scRNA-seq analysis tools with regards to data processing, downstream, and network analyses.

An introduction to the broad variety of data analysis platforms

A large number of data analysis platforms (web-based, stand-alone workflows, or integrated into computational frameworks like R or Python) are available to analyze scRNA-seq data. A comprehensive list is continuously updated in the scRNA-Tools database (<https://www.scRNA-tools.org/> [accessed on 13 January 2021]). There are commercially available analysis software packages, some of which are developed by single-cell sequencing companies and service providers, such as Cell Ranger and Loupe Cell Browser (10X Genomics) (23), as well as SeqGeq (BD Biosystems) (24). Others are designed by companies specializing in software solutions, such as Partek Flow (Partek) (25). While commercially available packages are user-friendly, open-source analysis packages are usually more powerful, transparent, reproducible, and flexible. Bioconductor, an open-development software

project for the analysis of high-throughput genomics data, provides powerful analysis tools, such as Scanpy (26) or Scater (27). Currently, the R package “Seurat” (28) is one of the most popular toolboxes for general single cell sequencing analysis, consistently performing well in benchmarking studies. However, the need for computational basic skills in either R or Python poses a hurdle for many scientists. Galaxy (<https://usegalaxy.eu> [accessed on 13 January 2021]) serves the same purpose as Bioconductor, but the developers aimed for an enhanced accessibility of the tools and easy usability for scientists with little or no bioinformatics background. An extensive Galaxy online training module (<https://galaxyproject.github.io/training-material/> [accessed on 13 January 2021]) is offered, and it is possible to use the most common workflows without having to master any command line-based tools (29).

Foundations of processing raw scRNA-seq data

Standard data processing can be classified into six stages: (i) raw data alignment; (ii) quality control and data normalization; (iii) data integration and correction; (iv) expression recovery; (v) feature-selection of data; and (vi) dimensionality reduction and visualization. The subsequently performed downstream analyses may use different levels of processed data as input (Figure 2). Depending on the experimental setup or data used, it is also possible to skip certain levels or to have slight alterations in their order, for example, data integration and correction might not be needed for single-batch datasets.

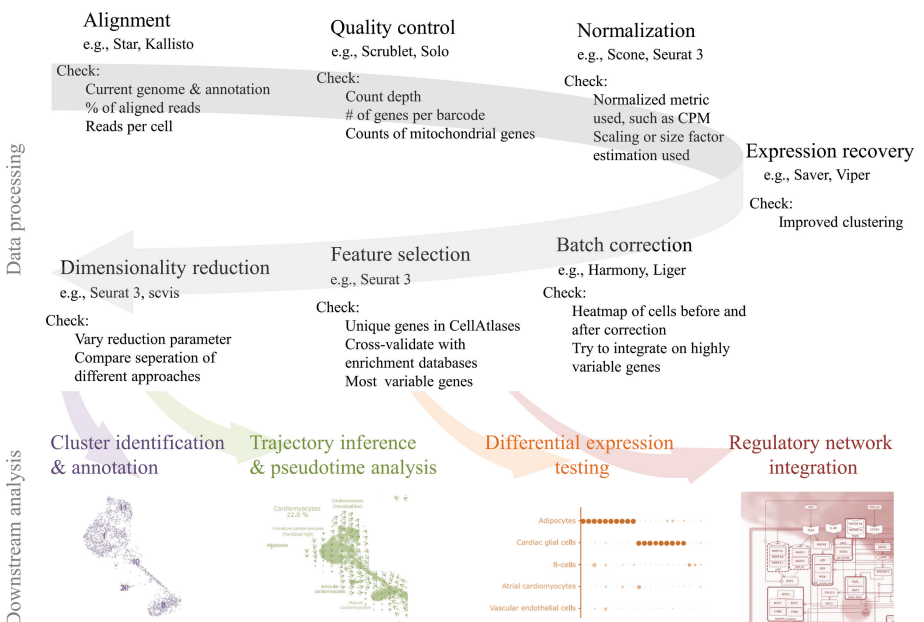


Figure 2. Flow chart summarizing current scRNA-seq data analysis principles.

Alignment of scRNA-Seq raw data

Alignment is the first and one of the most critical steps of the scRNA-seq analyses. In general, the aim of the alignment step is to find the original transcriptomic location of the experimentally obtained sequencing reads. Thus, the choice of the alignment tool and its parametrization directly affects the count matrix, all subsequent downstream analysis steps, and, finally, the biological findings. The two most popular tools for alignment are the splice-aware aligner STAR (30) and the pseudoalignment approach Kallisto (31). Recent benchmark studies evaluated the performance of these two methods using real datasets obtained from different platforms (DropSeq, Fluidigm, and 10xChromium) (32, 33). They conclude that Kallisto's use of computing resources is much less demanding than STAR when only cDNA sequences are used as the reference; however, such efficiency gain is at the cost of loss of information.

Quality control to determine cell viability and sequencing outcome

An important aspect of scRNA-seq protocols is that captured cells, irrespective of the method used, are often stressed, damaged, or broken. In addition, some capture sites can be empty, and some may contain multiple cells. All these events refer to “low quality” cells, which may lead to misinterpretation of the data and, therefore, need to be corrected (34). In general, cell quality control (QC) is commonly performed based on three QC covariates: (i) the number of counts per barcode (count depth); (ii) the number of genes per barcode; and (iii), the fraction of counts from mitochondrial genes per barcode (35). Cells that show an aberrant behavior for these characteristics are typically removed from further analysis, although care must be taken when studying a heterogeneous population of cells as total mRNA content and other features can vary substantially. On the one hand, barcodes with only low count depths, few detected genes, and a high fraction of mitochondrial counts may indicate cells whose cytoplasmic mRNA leaked out through a broken membrane, and thus, only mRNA located in the mitochondria is still conserved (22). On the other hand, cells with very high counts and a large amount of transcripts may represent doublets, which is why these have to be filtered with specific tools like Scrublet (36) or Solo (37).

Generation and normalization of the count matrix to ensure comparability

Counts in a count matrix represent the successful quantification of a sequencing read to a specific genomic location. There are multiple tools that generate a count matrix, e.g., Cell Ranger (38), indrops (39), SEQC (40), or bustools (41). Due to the technical variability inherent in the count matrix generation, the count depths for the same cell can differ. Thus, differences in gene expression between cells based on count data may have been introduced during sampling (42). Normalization addresses these differences to obtain correct relative gene expression abundances between cells, for example, via scaling of count data. The most common normalization approach is count depth scaling, referred to as “counts per million” (CPM). This approach was adapted from bulk RNA-seq analysis to

normalize count data towards a size factor proportional to the count depth per cell. Weinreb *et al.* introduced an extension of CPM that excludes genes accounting for less than 5% of the total counts in any cell, which allows for molecular count variability in only a few highly expressed genes (43). More cellular heterogeneity is taken into consideration by a pooling-based size factor estimation method that can be applied for more heterogeneous samples in order to increase the validity of the biological conclusions (44). To acquire a well-suited normalization, one may use Scone, which is a tool that provides graphical summaries and quantitative reports, as well as trade-offs and ranks of normalization methods by panel performance (45).

Batch-effect correction and external data integration may enhance the biological outcome

Single-cell data is often acquired based on multiple experiments with varying capturing times, consumables, and technology platforms. These differences can lead to large variations or so-called “batch-effects” in the data and may confound biological variations of interest. Tran *et al.* compared 14 methods in terms of computational runtime, the ability to handle large datasets, and batch-effect correction efficacy, while preserving cell type purity (46). Based on their results, Harmony (47), Liger (48), and Seurat 3 are the recommended methods for batch integration. Due to its significantly shorter runtime, Harmony is recommended as the first method to try, with the other methods as viable alternatives. Luecken *et al.* found in their data representing >1.2 million cells that highly variable gene selection improves the performance of data integration methods, whereas scaling pushes methods to prioritize batch removal over conservation of biological variation (49).

Data correction accounting for unusual droplets

For droplet-based methods only a fraction of droplets will contain an intact cell. Since biological experiments are never flawless, some RNA will leak out of dead or damaged cells, thereby producing ambient RNA. Droplets without an intact cell can still contain this ambient RNA, which in turn will contribute to the sequencing library and final reads. Variations in droplet size, amplification efficiency, and sequencing cause a wide range of library sizes for both “RNA background” and real cells. Most methods try to distinguish between them by utilizing the total molecules/reads per barcode to find an “inflection point” between larger libraries representing cells and smaller libraries assumed to be only background. Using knee plots one can visualize this inflection point where the total number of molecules per barcode suddenly drops. The R package DropletsUtils uses the complete count matrix of all droplets to assess the profile of ambient RNA from those droplets with extremely low counts. Gene-expression profiles deviating from this background are considered as originating from intact cells. Since background RNA often looks similar to the expression profile of the largest cell population, this is combined with an inflection point method. By this means, EmptyDrops can verify barcodes for very small cells in highly diverse samples.

Expression recovery corrects for zero or low read counts

An additional challenge during the analysis of scRNA-seq data derives from the low transcript capture and sequencing efficiency of current methods. This leads to a large proportion of genes (often more than 90%) with zero or low read counts (50). Although many of the observed zero counts reflect a true absence of expression, a considerable fraction is due to technical factors that can vary between less than 1% and more than 60% across cells (17). Early expression recovery approaches pooled data for each gene across similar cells, but this may lead to over-smoothing and can disturb the natural cell-to-cell stochasticity in gene expression. For this reason, more advanced approaches such as single-cell analysis via expression recovery (Saver) (50), were introduced. Saver assumes that the count of each gene in each cell follows a negative binomial model, which is used to estimate the recovered expression module. In contrast to approaches that impute dropout events by borrowing information across only genes or cells, scTSSR simultaneously leverages information from both similar genes and similar cells using a two-side sparse self-representation model and shows superior use in specific cases (51). Another approach showing an actual benefit over existing ones is Viper, which is based on nonnegative sparse regression models (52). Viper can progressively infer a sparse set of local neighborhood cells instead of only using similarly expressed genes and cells.

Feature selection filters the most important genes

Feature selection is the process of choosing genes that contain useful information about the underlying biology of the sample, while removing genes that contain no useful information or random noise. This reduces the data size to facilitate the computationally time-consuming steps, while aiming to preserve the relevant biological structure between cells. A simple approach for feature selection is to pick the most variable genes based on their expression between the identified or clustered cell populations. This assumes that actual biological differences occur as increased variation of highly regulated genes, in contrast to genes that are not changed at baseline level or are only slightly regulated via technical noise. There are built-in standard approaches for feature selection in Seurat 3, but also further advanced ones available based on a multinomial model (53) or ensemble feature selection and similarity measurements (54). After the technical process of choosing appropriate features per cellular group, manual curation and cross-validation with existing enrichment databases, for example, Enrichr (55), or gprofiler (56), may help to evaluate the *in silico* findings. For a first description of novel cell populations or subpopulations, experimental validation of highly expressed genes is still considered the gold standard.

Dimensionality reduction highly facilitates data interpretability

Dimensionality reduction is an essential tool required to tame the highly complex information content in scRNA-Seq data analysis. A proper reduction of the dimensions allows for effective noise removal and is pivotal for many downstream analyses that include cell clustering or lineage reconstruction. Principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform

manifold approximation and projection (UMAP), as well as many extensions of these three, are commonly used algorithms in scRNA-seq. Sun *et al.* compared 18 different dimensionality reduction methods on 30 publicly available scRNA-seq datasets (57). They suggest that applying sophisticated gene filtering approaches prior to running dimensionality reduction will help to improve their performance. In addition, they see a benefit in even more stringent gene filtering approaches because these result in a smaller subset of genes and, therefore, facilitate the application of some of the slow dimensionality reduction methods to larger data sets. A major problem during dimensionality reduction is to preserve the global structure of the data because removing dimensions might likewise suppress some information. Some algorithms, such as the scvis algorithm, try to overcome this limitation by computing low-dimensional embeddings of scRNA-seq data while preserving global structure of the high-dimensional measurements (58). Recently, Heiser *et al.* presented an unbiased framework that defines metrics of global and local structure preservation in dimensionality reduction transformations (59).

Important visualizations to share results

Common visualizations in single cell related publications include 2D and 3D clustering, heatmaps of highly expressed genes, violin plots, and dotplots (60, 61). A recent comparison of interactive single cell visualization tools (for example cellxgene, Loom-viewer, iSEE, single-cell explorer) was made by Cakir *et al.* who regard this specific type of visualization as beneficial to the whole research community, facilitating scientific progress (62). One can also consider using R-Shiny approaches to turn one's own research results into interactively explorable plots (63). To enrich one's own research results with additional data, one can utilize scAVI (<http://amp.pharm.mssm.edu/scavi/> [accessed on 13 January 2021]), which is a web-based platform developed to enable users to analyze and visualize published and unpublished scRNA-seq datasets with state-of-the-art algorithms and visualization methods. The scAVI platform supports the analysis and visualization of 463 publicly available scRNA-seq studies from GEO covering 194,653 single cells. Once there is a very high number of cells within an analysis, algorithms like "TooManyCells", a suite of graph-based algorithms, can be used to efficiently identify and visualize cell populations (64).

DOWNSTREAM ANALYSES

After utilizing the main processing steps and visualizing the scRNA-seq data, more in-depth analyses should be carried out at a cellular level and at the gene level to fully decipher the identified single-cell profiles. Here, we present further advanced approaches that will help to exploit the information content of a sample of interest.

Cluster analysis and cell cluster annotation

A major use case of scRNA-seq is to identify, quantify, and characterize cell populations in heterogeneous samples or tissues. From a biological perspective,

such cell populations include different cell types, or may refer to different states of identical cell types, for example, stimulated and unstimulated cells, or cells in a different maturation state. A *de novo* identification of cells, mathematically speaking, is an unsupervised clustering problem, which has been widely studied with machine learning algorithms, and there are several well-established strategies that have been adapted for scRNA-seq data (65). However, the annotation of the newly identified clusters is still a bottleneck in terms of time consumption and expertise needed because one has to manually curate the clusters using so-called cell atlases, such as “CellAtlas” (<https://data.humancellatlas.org/> [accessed on 13 January 2021]) or the “Single Cell Expression Atlas” (<https://www.ebi.ac.uk/gxa/sc/home> [accessed on 13 January 2021]). To overcome these limitations, promising automated annotation tools, such as SCSA (66) or further ML-based oversampling techniques of Bej *et al.* (67) have already shown significant potential.

Trajectory analysis and inference to investigate cellular origins

The newly acquired resolution of scRNA-seq allows researchers to distinguish between closely related cell populations, potentially revealing functionally distinct groups with complex relationships (68). For many cellular investigations, there are no distinct borders between cellular states, but instead a smooth transition, where individual cells represent points along a continuum or lineage, in which cells change states by undergoing gradual transcriptional changes, representing a temporal variable or pseudotime (69). The inference of lineage structures is considered as pseudotemporal reconstruction of a sample of interest that finally infers changing cell states and cell fate decisions (70). In addition, many cell populations contain several lineages that share a common initial group branching into different further subgroups, which requires additional analyses to distinguish between cells that fall along those different lineages (71). The two most popular approaches for pseudotemporal reconstruction are Monocle (70) and Slingshot (69), which have been recently compared and benchmarked (54, 69). These benchmarks showed that dimensionality reduction results based on Monocle3 are in line with recommendations by the Monocle3 software itself, which uses UMAP as the default dimensionality reduction method (72). Moreover, the set of the best dimensionality reduction methods for Monocle3 are consistent with those for Slingshot, with only one method difference between the two (GLMPCA [generalized principal component analysis] in place of common PCA).

Gene expression dynamics and pseudotime can reveal cell fate decisions

A central challenge in trajectory inference is the destructive nature of scRNA-seq, which reveals only static snapshots of cellular states; additional information is required to constrain possible dynamics that could make a reasonable prediction towards the same trajectory (73). The concept of RNA velocity enabled the investigation of such dynamic information by assuming that newly transcribed (unspliced pre-mRNAs containing introns) and mature (spliced mRNAs) can be distinguished in common scRNA-seq protocols (74). However, errors in velocity estimates may arise if the central assumptions of a common splicing rate and the

observation of the full splicing dynamics with steady-state mRNA levels are violated (75). Thus, Bergen *et al.* (75) developed “scVelo”, as a further extension of “RNAvelocity” and solving its limitations, by utilizing the full transcriptional dynamics of splicing kinetics using a likelihood-based dynamical model. This model generalizes RNA velocity computations with transient cell states, which are common in the development of and the response to perturbations. Another interesting approach to simulate time-series trajectories is proposed by Yeo *et al.* (76), who are using a generative model framework that is able to predict trajectories for cells, which are not found in the model’s training set (including cells in which genes or sets of genes have been perturbed).

Differential expression testing as a major hurdle in data analysis

In general, scRNA-seq and bulk RNA-seq data have different characteristics that require a new differential expression (DE) analysis definition beyond the common nonzero difference in average expression, which is not adequately addressed yet. Due to the small amount and low capture efficiency of RNA molecules in single cells, many transcripts tend to be missed during reverse transcription. As a result, one observes that some transcripts are highly expressed in one cell but are not expressed in another cell of the same population, which is defined as a “drop-out” event (77). In addition, multimodality, heterogeneity, and sparsity (many zero counts) are the major hurdles for an effective DE calling. A comprehensive, comparative study of differential gene expression analysis tools for single-cell RNA sequencing data was recently made by Wang *et al.* (78). They observed that the agreement of tools calling DE genes is not high (~10%) and concluded that there is a trade-off between true-positive rates and the precision of calling DE genes. Methods with higher true positive rates tend to show low precision due to introducing false positives, whereas methods with high precision show low true positive rates due to identifying few DE genes. One solution could be to use different DE testing models, such as methods that can capture multimodality (for example, scDD (79)) and model-based approaches (for example, Monocle) designed for handling zero counts.

Gene regulatory networks and disease maps

Gene expression is highly regulated by transcription factors, co-factors, and signaling molecules that span cross-related networks. An improved understanding of these networks is a major goal in biology and medicine because it determines essential factors that are responsible for healthy and disease related phenotypes. So far, molecular networks have been solely based on microarray and bulk RNA-seq data, but are now being refined with single-cell resolution. A recent systematic evaluation of state-of-the-art algorithms for inferring gene regulatory networks from single-cell transcriptional data contributes to the use of single-cell data to parametrize novel networks and improves already existing ones (80). In addition, further algorithms are used to infer global, large-scale regulatory networks on organ scale and perturbed systems, such as diabetes and Alzheimer’s disease (81), or reconstruct networks by using scRNA-seq data of barcoded genotypes (82). Novel disease oriented network applications, like the inflammation

resolution disease map, are able to integrate various kinds of omics data, such as scRNA-Seq, to simulate all relevant molecular processes and serve as a comprehensive knowledge base on single-cell level (83).

CONCLUSION

Taken together, scRNA-seq already offers huge potential for many biological and biomedical areas and will be further applied to decipher currently undiscovered molecular processes. Nevertheless, one must be aware of current sequencing technologies and must determine if scRNA-seq is necessary or can be circumvented with bulk RNA-seq of purified or individual cell types. From a computational perspective, a lot of data analysis steps have potential for optimization and require an even higher awareness of complexity than bulk RNA-seq analysis. The presented procedures and technologies are a current snapshot in the lively field of scRNA-seq, which may rapidly develop in the near future.

Acknowledgement: This work was supported by the EU Social Fund (ESF/14-BM-A55-0024/18, ESF/14-BM-A55-0027/18, ESF/14-BM-A55-0028/18), the “Deutsche Forschungs- gemeinschaft” - DFG (DA1296/6-1), the German Heart Foundation (F/01/12), the FORUN Program of Rostock Medical University (889001/889003), the Josef and Käthe Klinz Foundation (T319/29737/2017), the Damp Foundation (2016-11), and the Federal Ministry of Education and Research - BMBF (VIP+00240, 031L0106C).

Conflict of interest: The authors declare no potential conflict of interest with respect to research, authorship, and/or publication of this chapter.

Copyright and permission statement: The authors confirm that the materials included in this chapter do not violate copyright laws. Where relevant, appropriate permissions have been obtained from the original copyright holder(s), and all original sources have been appropriately acknowledged or referenced.

REFERENCES

1. Bykov Y, Kim SH, Zamarin D. Chapter Sixteen - Preparation of single cells from tumors for single-cell RNA sequencing. In: Galluzzi L, Rudqvist N-P, Methods in Enzymology Academic Press; 2020. S. 295–308. (Tumor Immunology and Immunotherapy - Cellular Methods Part B; Bd. 632). <https://doi.org/10.1016/bs.mie.2019.05.057>
2. Liao J, Yu Z, Chen Y, Bao M, Zou C, Zhang H, et al. Single-cell RNA sequencing of human kidney. *Sci Data*. 2020;7(1):4. <https://doi.org/10.1038/s41597-019-0351-8>
3. Li, H. Single-cell RNA sequencing in Drosophila: Technologies and applications. *WIREs Dev Biol*. 2020;e396. <https://doi.org/10.1002/wdev.396>
4. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc*. 2018;13(4):599–604. <https://doi.org/10.1038/nprot.2017.149>
5. Hashimshony T, Senderovitch N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*. 2016;17(1):77. <https://doi.org/10.1186/s13059-016-0938-8>

6. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202–14. <https://doi.org/10.1016/j.cell.2015.05.002>
7. Goldstein LD, Chen Y-JJ, Dunne J, Mir A, Hubschle H, Guillory J, et al. Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics*. 2017;18(1):519. <https://doi.org/10.1186/s12864-017-3893-1>
8. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*. 2015;161(5):1187–201. <https://doi.org/10.1016/j.cell.2015.04.044>
9. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*. 2014;343(6172):776–9. <https://doi.org/10.1126/science.1247651>
10. Gierahn TM, Wadsworth MH, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods*. 2017;14(4):395–8. <https://doi.org/10.1038/nmeth.4179>
11. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*. 2013;10(11):1096–8. <https://doi.org/10.1038/nmeth.2639>
12. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8(1):14049. <https://doi.org/10.1038/ncomms14049>
13. Fluidigm | Products | C1. [cited 6. July 2020]. <https://www.fluidigm.com/products/c1-system>
14. Archer N, Walsh MD, Shahrezaei V, Hebenstreit D. Modeling Enzyme Processivity Reveals that RNA-Seq Libraries Are Biased in Characteristic and Correctable Ways. *Cell Syst*. 2016;3(5):467–479.e12. <https://doi.org/10.1016/j.cels.2016.10.012>
15. de Klerk E, den Dunnen JT, 't Hoen PAC. RNA sequencing: from tag-based profiling to resolving complete transcript structure. *Cell Mol Life Sci*. 2014;71(18):3537–51. <https://doi.org/10.1007/s00018-014-1637-9>
16. Single-Cell Sequencing Workflow: Critical Steps and Considerations. :35.
17. Svensson V, Natarajan KN, Ly L-H, Miragaia RJ, Labalette C, Macaulay IC, et al. Power analysis of single-cell RNA-sequencing experiments. *Nat Methods*. 2017;14(4):381–7. <https://doi.org/10.1038/nmeth.4220>
18. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*. 2020;38(6):737–46. <https://doi.org/10.1038/s41587-020-0465-8>
19. Mereu E, Lafzi A, Moutinho C, Ziegenhain C, McCarthy DJ, Álvarez-Varela A, et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat Biotechnol*. 2020; (38):747–755. <https://doi.org/10.1038/s41587-020-0469-4>
20. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, et al. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell*. 2019;73(1):130–142.e5. <https://doi.org/10.1016/j.molcel.2018.10.020>
21. Hie B, Peters J, Nyquist SK, Shalek AK, Berger B, Bryson BD. Computational Methods for Single-Cell RNA Sequencing. *Annu Rev Biomed Data Sci*. 2020;3(1):339–64. <https://doi.org/10.1146/annurev-biodatasci-012220-100601>
22. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*. 2019;15(6):e8746. <https://doi.org/10.15252/msb.20188746>
23. What is Loupe? -Software -Genome & Exome -Official 10x Genomics Support. [cited 16. September 2020]. <https://support.10xgenomics.com/genome-exome/software/visualization/latest/what-is-loupe>
24. Take scRNA-seq analysis into your own hands with SeqGeq. | FlowJo, LLC. [cited 16. September 2020]. <https://www.flowjo.com/solutions/seqgeq>
25. Gosche K. Partek Flow Genomic Analysis Software - Lab and Enterprise Solutions. Partek Inc. [cited 16. September 2020]. <https://www.partek.com/partek-flow/>

26. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15. <https://doi.org/10.1186/s13059-017-1382-0>
27. McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinforma Oxf Engl.* 15 2017;33(8):1179–86. <https://doi.org/10.1093/bioinformatics/btw777>
28. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411–20. <https://doi.org/10.1038/nbt.4096>
29. Batut B, Hiltmann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, et al. Community-Driven Data Analysis Training for Biology. *Cell Syst.* 2018;6(6):752–758.e1. <https://doi.org/10.1016/j.cels.2018.05.012>
30. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>
31. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7. <https://doi.org/10.1038/nbt.3519>
32. Du Y, Huang Q, Arisdakessian C, Garmire LX. Evaluation of STAR and Kallisto on Single Cell RNA-Seq Data Alignment. *G3.* 2020;10(5):1775–83. <https://doi.org/10.1534/g3.120.401160>
33. Melsted P, Boeshaghi AS, Gao F, Beltrame E, Lu L, Hjorleifsson KE, et al. Modular and efficient pre-processing of single-cell RNA-seq. *bioRxiv.* 2019;673285. <https://doi.org/10.1101/673285>
34. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 2016;17(1):29. <https://doi.org/10.1186/s13059-016-0888-1>
35. Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol Syst Biol.* 2018;14(4):e8046. <https://doi.org/10.15252/msb.20178046>
36. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* 2019;8(4):281–291.e9. <https://doi.org/10.1016/j.cels.2018.11.005>
37. Bernstein NJ, Fong NL, Lam I, Roy MA, Hendrickson DG, Kelley DR. Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. *Cell Syst.* 2020;11(1):95–101.e5. <https://doi.org/10.1016/j.cels.2020.05.010>
38. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8(1):14049. <https://doi.org/10.1038/ncomms14049>
39. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell.* 2015;161(5):1187–201. <https://doi.org/10.1016/j.cell.2015.04.044>
40. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell.* 2018;174(5):1293–1308.e36. <https://doi.org/10.1016/j.cell.2018.05.060>
41. Melsted P, Ntranos V, Pachter L. The barcode, UMI, set format and BUStools. *Bioinformatics.* 2019;35(21):4472–3. <https://doi.org/10.1093/bioinformatics/btz279>
42. Massoni-Badosa R, Iacono G, Moutinho C, Kulis M, Palau N, Marchese D, et al. Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol.* 2020;21(1):112. <https://doi.org/10.1186/s13059-020-02032-0>
43. Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics.* 2018;34(7):1246–8. <https://doi.org/10.1093/bioinformatics/btx792>
44. Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 2016;17(1):75. <https://doi.org/10.1186/s13059-016-0947-7>
45. Cole MB, Risso D, Wagner A, DeTomaso D, Ngai J, Purdom E, et al. Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq. *Cell Syst.* 2019;8(4):315–328.e8. <https://doi.org/10.1016/j.cels.2019.03.010>

46. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 2020;21(1):12. <https://doi.org/10.1186/s13059-019-1850-9>
47. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16(12):1289–96. <https://doi.org/10.1038/s41592-019-0619-0>
48. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell.* 2019;177(7):1873–1887. e17. <https://doi.org/10.1016/j.cell.2019.05.006>
49. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, et al. Benchmarking atlas-level data integration in single-cell genomics. *bioRxiv.* 2020;2020.05.22.111161. <https://doi.org/10.1101/2020.05.22.111161>
50. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat Methods.* 2018;15(7):539–42. <https://doi.org/10.1038/s41592-018-0033-z>
51. Jin K, Ou-Yang L, Zhao X-M, Yan H, Zhang X-F. scTSSR: gene expression recovery for single-cell RNA sequencing using two-side sparse self-representation. *Bioinformatics.* 2020;36(10):3131–8. <https://doi.org/10.1093/bioinformatics/btaa108>
52. Chen M, Zhou X. VIPER: variability-preserving imputation for accurate gene expression recovery in single-cell RNA sequencing studies. *Genome Biol.* 2018;19(1):196. <https://doi.org/10.1186/s13059-018-1575-1>
53. 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(1):295. <https://doi.org/10.1186/s13059-019-1861-6>
54. Jeong H, Khunlertgit N. Effective single-cell clustering through ensemble feature selection and similarity measurements. *Comput Biol Chem.* 2020;87:107283. <https://doi.org/10.1016/j.compbiolchem.2020.107283>
55. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377>
56. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019;47(W1):W191–8. <https://doi.org/10.1093/nar/gkz369>
57. 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(1):269. <https://doi.org/10.1186/s13059-019-1898-6>
58. Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat Commun.* 2018;9(1):2002. <https://doi.org/10.1038/s41467-018-04368-5>
59. Heiser CN, Lau KS. A Quantitative Framework for Evaluating Single-Cell Data Structure Preservation by Dimensionality Reduction Techniques. *Cell Rep.* 2020;31(5):107576. <https://doi.org/10.1016/j.celrep.2020.107576>
60. Cui Y, Zheng Y, Liu X, Yan L, Fan X, Yong J, et al. Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep.* 2019;26(7):1934–1950.e5. <https://doi.org/10.1016/j.celrep.2019.01.079>
61. Galow AM, Wolfien M, Müller P, Bartsch M, Brunner RM, Hoeflich A, et al. Integrative Cluster Analysis of Whole Hearts Reveals Proliferative Cardiomyocytes in Adult Mice. *Cells.* 2020;9(5):1144. <https://doi.org/10.3390/cells9051144>
62. Cakir B, Prete M, Huang N, van Dongen S, Pir P, Kiselev VY. Comparison of visualization tools for single-cell RNAseq data. *NAR Genomics Bioinforma.* 2020;2(3). <https://doi.org/10.1093/nargab/lqaa052>
63. SCHNAPPs - Single Cell sHiNy APPLication(s) | bioRxiv [cited 16. September 2020]. <https://www.biorxiv.org/content/10.1101/2020.06.07.127274v1.full>

64. Schwartz GW, Zhou Y, Petrovic J, Fasolino M, Xu L, Shaffer SM, et al. TooManyCells identifies and visualizes relationships of single-cell clades. *Nat Methods*. 2020;17(4):405–13. <https://doi.org/10.1038/s41592-020-0748-5>
65. Andrews TS, Hemberg M. Identifying cell populations with scRNASeq. *Mol Aspects Med*. 2018;59:114–22. <https://doi.org/10.1016/j.mam.2017.07.002>
66. Cao Y, Wang X, Peng, G. SCSA: A Cell Type Annotation Tool for Single-Cell RNA-seq Data. *Frontiers in genetics*, 2020; 11: 490. <https://doi.org/10.3389/fgene.2020.00490>
67. Bej S, Davtayan N, Wolfien M, Nassar M, Wolkenhauer O. LoRAS: An oversampling approach for imbalanced datasets. *Machine Learning*. 2021;110: 279–301. <https://doi.org/10.1007/s10994-020-05913-4>
68. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol*. 2016;34(11):1145–60. <https://doi.org/10.1038/nbt.3711>
69. 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 Genomics*. 2018;19(1):477. <https://doi.org/10.1186/s12864-018-4772-0>
70. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381–6. <https://doi.org/10.1038/nbt.2859>
71. Haghverdi L, Büttner M, Wolf FA, Büttner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods*. 2016;13(10):845–8. <https://doi.org/10.1038/nmeth.3971>
72. 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(7745):496–502. <https://doi.org/10.1038/s41586-019-0969-x>
73. Tritschler S, Büttner M, Fischer DS, Lange M, Bergen V, Lickert H, et al. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*. 2019; 146(12):dev170506. <https://doi.org/10.1242/dev.170506>
74. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature*. 2018;560(7719):494–8. <https://doi.org/10.1038/s41586-018-0414-6>
75. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;1–7. <https://doi.org/10.1101/820936>
76. Yeo GHT, Saksena SD, Gifford DK. Generative modeling of single-cell population time series for inferring cell differentiation landscapes. *Systems Biology*; 2020. <https://doi.org/10.1101/2020.08.26.269332>
77. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods*. 2014;11(7):740–2. <https://doi.org/10.1038/nmeth.2967>
78. Wang T, Li B, Nelson CE, Nabavi S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinformatics*. 2019;20(1):40. <https://doi.org/10.1186/s12859-019-2599-6>
79. Korthauer KD, Chu L-F, Newton MA, Li Y, Thomson J, Stewart R, et al. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol*. 2016;17(1):222. <https://doi.org/10.1186/s13059-016-1077-y>
80. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*. 2020;17(2):147–54. <https://doi.org/10.1038/s41592-019-0690-6>
81. Iacono G, Massoni-Badosa R, Heyn H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol*. 2019;20(1):110. <https://doi.org/10.1186/s13059-019-1713-4>
82. Jackson CA, Castro DM, Saldi G-A, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *eLife*. 2020;9:e51254. <https://doi.org/10.7554/eLife.51254>
83. Serhan CN, Gupta SK, Perretti M, Godson C, Brennan E, Li Y, et al. The Atlas of Inflammation Resolution (AIR). *Mol Aspects Med*. 2020;74:100894. <https://doi.org/10.1016/j.mam.2020.100894>

