Single-Cell RNA Sequencing-Based Computational Analysis to Describe Disease Heterogeneity

Tao Zeng* and Hao Dai

Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China

The trillions of cells in the human body can be viewed as elementary but essential biological units that achieve different body states, but the low resolution of previous cell isolation and measurement approaches limits our understanding of the cell-specific molecular profiles. The recent establishment and rapid growth of single-cell sequencing technology has facilitated the identification of molecular profiles of heterogeneous cells, especially on the transcription level of single cells [single-cell RNA sequencing (scRNA-seq)]. As a novel method, the robustness of scRNA-seq under changing conditions will determine its practical potential in major research programs and clinical applications. In this review, we first briefly presented the scRNA-seq-related methods from the point of view of experiments and computation. Then, we compared several state-of-the-art scRNA-seq analysis frameworks mainly by analyzing their performance robustness on independent scRNA-seq datasets for the same complex disease. Finally, we elaborated on our hypothesis on consensus scRNA-seq analysis and summarized the potential indicative and predictive roles of individual cells in understanding disease heterogeneity by single-cell technologies.

Keywords: cellular heterogeneity, complex diseases, single-cell RNA sequencing, network, integration

INTRODUCTION

It is known that an adult human body consists of trillion cells of different types and origins, and each of them plays its respective role in the body system. These cells can be viewed as basic but essential biological units supporting different body states, e.g., health, disease, or the response to therapy. Decades ago, the low resolution of cell isolation and measurement technologies limited our understanding of the cell-specific molecular profiles and their importance in cellular systems, causing humans to always underestimate disease heterogeneity.

In recent years, the establishment and the rapid growth of single-cell sequencing technology have led to the efficient and inexpensive identification of molecular profiles of individual cells (Bose et al., 2015; Baran-Gale et al., 2018; Svensson et al., 2018). In particular, the transcription of single cells (Wu et al., 2014; Ziegenhain et al., 2017) is a novel and fast evolving field. Single-cell RNA sequencing (scRNA-seq) attracts increasing attention to the identification and characterization of cells on an individual level rather than on a population level (Saliba et al., 2014; McDavid et al., 2016; Raj et al., 2018; Torre et al., 2018).

The research field of single cells, e.g., identifying cell types, recognizing cell markers, and tracing cell origins, is currently undergoing rapid development. New knowledge on cells can improve our understanding of biological systems by changing our perspective from the traditional population level

to the individual cellular level. It can further provide novel insights into old biological and biomedical questions (Raj et al., 2018). For example, with scRNA-seq data rather than bulk transcriptome data, we can detect genes with conserved expression levels across individual cells (Lin Y. et al., 2017). Single-cell transcriptomics could even uncover the diverse transcriptional states of immune cells and their coordination during immune responses (Vegh and Haniffa, 2018). In addition, simultaneous measurements of transcription along with genomic and epigenetic profiling at the single-cell level (Clark et al., 2016) is expected to be developed soon and will provide groundbreaking biological insights into these basic blocks building the biological body (Hemberg 2018).

In this quickly evolving field, many reviews have focused on the biotechnological applications of scRNA-seq and in silico gene expression analysis. The program goals of the Common Fund-supported Single Cell Analysis Program from the National Institutes of Health point out the impact of resolving tissue heterogeneity at the cellular level (Roy et al., 2018). Different scRNA-seq protocols have their strengths and disadvantages under respective settings (Saliba et al., 2014; Bacher and Kendziorski, 2016). The pre-processing approaches of sparse and row-rank scRNA-seq data (Zhang L. et al., 2018), normalization methods (Vallejos et al., 2017), and batch effect corrections (Dal Molin and Di Camillo, 2018; Haghverdi et al., 2018) have all been carried out for a wide range of comparisons and evaluations. Finally, the cell type clustering algorithms, cell marker identification, and cell trajectory reference also have their target-specific evaluation approaches for the deconvolution of biological system heterogeneity (Menon 2018; Papalexi and Satija, 2018). In addition, integrative impacts of whole scRNA-seq protocols and analysis methodologies have undergone in-depth assessments (Dal Molin et al., 2017; Svensson et al., 2017; Todorov and Saeys, 2018).

These current developments and achievements of scRNA-seq motivated us to investigate the individual cell types, cell signatures, cell origins in time and space, and cell communication strategies. Meanwhile, as a novel method, its robustness under different conditions (e.g., when applied to different datasets) will determine its actual practical potential in major research programs (e.g., the Precision Medicine Initiative or the Human Brain Project) (Poo et al., 2016; Sankar and Parker, 2016) or in clinical applications (e.g., diagnosis or prognosis of complex diseases) (Zeng et al., 2016). Thus, in this review paper, we discussed scRNA-seq from the point of view of experiments and computation. Then, on independent scRNA-seq datasets for the same complex disease (i.e., diabetes), we compared several state-of-the-art scRNA-seq analysis frameworks mainly by the robustness of their performances in the identification of cell types and markers. Lastly, we elaborated on our hypothesis on consensus scRNA-seq analysis and summarized the potential indicative and predictive roles of characteristic cells in understanding disease heterogeneity by single-cell technologies.

MATERIALS AND METHODS

A recent review has demonstrated the principle and potential of scRNA-seq in a wide range of studies, including development,

physiology, and disease (Potter 2018). It concluded that the data noise and cell number are the main limitations in scRNA-seq studies, and many research fields would benefit from its continuous development. In contrast, this work concentrated on the scRNA-seq-based study from the two angles of experiments and computation. Especially, the robustness of scRNA-seq under changing conditions will decide its practical potential, e.g., in precision medicine. Thus, different from a previous report (Potter 2018), we further compared several state-of-the-art scRNA-seq analysis frameworks and included our hypothesis on the performance consensus.

scRNA-seq-Associated Biological Experiments

scRNA-seq is becoming a widely used genome-wide technology to detect cellular identities and dynamics, e.g., cell subpopulations, cell state marker genes and pathways, cell state transitions, and cell trajectories (Nguyen et al., 2018). This sustained improvement of the sensitivity, flexibility, and efficiency of scRNA-seq will help to resolve many biological and biomedical research questions on the individual cell level.

On the one hand, the rapid development of experimental protocols of scRNA-seq expands the measurement of mRNA levels to many associated fields of study (Fuzik et al., 2016; Hashimshony et al., 2016; Ilicic et al., 2016; Bagnoli et al., 2018; Han et al., 2018; Hayashi et al., 2018; Sasagawa et al., 2018). Especially, scRNA-seq applications have provided new insights into conventional biological questions, e.g., cellular heterogeneity. New cell types have been more widely recognized than previously expected (Burns et al., 2015; Usoskin et al., 2015; Rheaume et al., 2018), and gene expression levels corresponding to old and new cell types have uncovered many biological functions and mechanisms that were overlooked in conventional cell population studies (Nelson et al., 2016; Li H. et al., 2017); single-cell transcriptomic characteristics can reveal more time-dependent features of a biological system (Zeisel et al., 2015; Zeng et al., 2017; Lescroart et al., 2018; Liu D. et al., 2018), whereas the pseudo-time of single cells would mimic the actual dynamic biological process (Kowalczyk et al., 2015; Cacchiarelli et al., 2018). Taking all of the above novelties together, we can deepen our understanding on the complex mechanisms underlying cell-to-cell variation. These complex dynamic responses are controlled by regulatory cell-to-cell communication, which is also responsible for cellular heterogeneity (Shalek et al., 2014).

Measuring Regulatory Elements in a Single Cell

Cell-specific transcriptional signals might be regulated by the high-order structural folding of nucleosomes (Nagano et al., 2017; Lando et al., 2018), which can be investigated by combining scRNA-seq with other single-cell approaches (Stevens et al., 2017; Liu T. et al., 2018; Mezger et al., 2018). Of note, current scRNA profiling methods usually destroy cells during the analysis process, hindering the measurement of temporal gene expression changes. However, some information on biological dynamics will always be present in the data. For example, the continuum of molecular states in a population can reflect the trajectory or pseudo-time of a typical cell, so various methods increase their

power by reconstructing the trajectory by quantification of a group of cells in multiple static snapshots (Weinreb et al., 2018).

Measuring Post-transcriptional Regulations in a Single Cell

Understanding nongenetic cellular heterogeneity will help to characterize complete biological mechanisms in live cells, but little knowledge is available on the heterogeneity of regulatory modifications between individual cells. For example, microRNAs (miRNAs) are small RNAs that regulate gene expression in a post-transcriptional manner and might reduce cell-to-cell variability on the protein level by repressing mRNA translation or promoting mRNA degradation. Although the wet experimental evidence for the roles of miRNA in individual cells is limited, great efforts have been made to investigate such regulatory modifications in single cells (Fan et al., 2015). For instance, single-cell Quartz-Seq technology was developed to identify different kinds of nongenetic cellular heterogeneity in a quantitative manner (Sasagawa et al., 2013). Single-cell small RNA sequencing and analysis techniques have supplied much evidence that miRNAs could be potential molecular biomarkers for indicating the type and state of particular cells (Faridani et al., 2016). Moreover, using a combination of scRNA-seq data and mathematical modeling, it is also possible to detect key miRNAs as cell type-specific post-transcriptional regulators (Rzepiela et al., 2018).

Measuring Upstream Regulatory Factors in a Single Cell

Individual cells within different subpopulations can show significant variations when responding to external stresses, but the nature of this cellular heterogeneity is not clear, especially the remarkable alterations in the transcriptional architecture (Xue et al., 2013; Edsgard et al., 2016; Gasch et al., 2017). Fortunately, scRNA-seq provides high resolution to genetics by linking phenotypes to cell-specific gene functions, and the genetic screening of single cells can even be realized now (Birnbaum 2018; Raj et al., 2018). For example, the Perturb-seq was designed to combine scRNA-seq and CRISPR-based perturbations to detect individual perturbations causing target gene changes, gene signature appearances, genetic interaction rewiring, and cell state transitions (Dixit et al., 2016), e.g., discovering previously unknown immune circuits (Jaitin et al., 2016). Next, the allele-sensitive scRNA-seq could recognize clonal and dynamic monoallelic expression patterns (Reinius et al., 2016) or analyze allele-specific cis-control in genome-wide expressions (Deng et al., 2014; Jiang et al., 2017). Besides, focusing on the quantitative trait locus (QTL), the computational tool demuxlet was implemented to perform expression QTL (eQTL) analysis, which can identify natural genetic variation within multiplexed droplet scRNA-seq to evaluate cell type-specific gene expression changes (Kang et al., 2018). Similarly, some new cell type-specific "co-expression QTLs" have even been detected according to the genetic variants, significantly altering co-expression relationships (van der Wijst et al., 2018).

Measuring Downstream Regulation in a Single Cell

The cell-to-cell regulatory communication plays important roles in cellular diversity across diverse biological systems, which is an important factor in the evolution of observed cell types. scRNA-seq provides a powerful tool to analyze particular regulatory mechanisms and their downstream influence in a corresponding subset of cells (Chu et al., 2016; Korthauer et al., 2016; Enge et al., 2017; Severo et al., 2018). For example, the integration of transcription factor expression, chromatin profiling, and sequence motif analysis can be effective to identify the cell-specific genomic regulation underlying cell-specific gene expression (Sebe-Pedros et al., 2018). Similarly, the integration of information about single-cell transcriptomics and cell-free plasma RNA provides the potential to uncover longitudinal cellular dynamics of cells in complex biological processes or pathological development (Tsang et al., 2017). Next, a two-part method combining a generalized linear model and gene set enrichment analysis on single-cell data provided evolutionary insights in gene co-expression by experimental treatments (Finak et al., 2015). In addition, benefitting from time-course data obtained by scRNA-seq, it is possible to characterize the fate decision and transcriptional control of self-renewal, differentiation, and maturation of particular cells (Su et al., 2017), and transient cellular states corresponding to asynchronous cellular responses can be observed under conditional perturbations (Rizvi et al., 2017).

scRNA-seq-Associated Analytic Computations

As seen in the above summary, scRNA-seq technologies are swiftly developing. They are greatly beneficial to the investigation of transcriptional landscapes at the single-cell level, where they are able to profile cell-to-cell variability in cell populations and characterize unexpected heterogeneity of transcription in originally thought homogeneous cell populations. Although many computational methods for analyzing scRNA-seq data have been extensively developed, tested, and validated on simulated datasets, scRNA-seq protocols are still complex so that bias will easily occur in downstream analysis. In fact, computational models and tools available for the design and analysis of scRNA-seq experiments (**Table 1**) have their advantages and disadvantages in various settings, and many questions have yet to be solved in this exciting area (Bacher and Kendziorski, 2016). Similar to other high-throughput sequencing technologies, the general actions on scRNA-seq data include several key steps before the follow-up analysis for single cells (Jia et al., 2017; Li Y. H. et al., 2017; McCarthy et al., 2017; Chen W. et al., 2018; Vu et al., 2018), i.e., pre-procession (e.g., zero imputation) (Li and Li, 2018; Van den Berge et al., 2018), quality control (e.g., variation analysis) (Brennecke et al., 2013; Ding et al., 2015; Jiang et al., 2016; Eling et al., 2018; Lu et al., 2018), normalization (Bacher et al., 2017; Cole et al., 2017; Haghverdi et al., 2018; Tian et al., 2018), and visualization/simulation (Zappia et al., 2017). Although scRNA-seq studies have provided revolutionary tools to assist researchers to address scientific questions previously hard to investigate directly, several computational challenges are beginning to arise.

Challenge of Cluster Analysis of Single Cells

The detection of cell types from heterogeneous cells is an important step in the development of scRNA-seq data analysis

TABLE 1 | List of computational tools for single-cell RNA sequencing (scRNA-seq) analysis.

Category	ID	Access	Code and citation
Pre-procession	scater	Bioconductor	R (McCarthy et al., 2017)
	scPipe	Bioconductor	R (Tian et al., 2018)
	GRM	http://wanglab.ucsd.edu/star/GRM	R (Ding et al., 2015)
Cell clustering	SAFEclustering	http://yunliweb.its.unc.edu/safe/	R (Yang et al., 2018)
	DendroSplit	Github	Python (Zhang J. et al., 2018)
	clusterExperiment	Bioconductor	R (Risso et al., 2018)
	scmap	Bioconductor	R (Kiselev et al., 2018)
	scVDMC	Github	Matlab (Zhang H. et al., 2018)
	CIDR	Github	R (Lin P. et al., 2017)
	scClustBench	http://www.maths.usyd.edu.au/u/SMS/bioinformatics/software.html	R (Kim et al., 2018)
	SNN-Cliq	http://bioinfo.uncc.edu/SNNCliq	Matlab & Python (Xu and Su, 2015)
Cell marking	MAST	Github	R (Finak et al., 2015)
	SC2P	Github	R (Wu et al., 2018)
	DEsingle	Bioconductor	R (Miao et al., 2018)
	powsimR	Github	R (Vieth et al., 2017)
	BPSC	Github	R (Vu et al., 2016)
	Sincell	Bioconductor	R (Julia et al., 2015)
Cell ordering	dynverse	Github	R (Saelens et al., 2018)
	Progra	Github	R (Gong et al., 2018)
	p-Creode	Github	Python (Herring et al., 2018)
Pipeline	SINCERA	https://research.cchmc.org/pbge/sincera.html	R (Guo et al., 2015)
	SCell	Github	Exe (Diaz et al., 2016)
	Falco	Github	Python (Yang et al., 2017)
	ASAP	Github	R & python (Gardeux et al., 2017)
	SIMLR	Github	R & Matlab (Wang et al., 2017; Wang B. et al., 2018)
	SEURAT	http://satijalab.org/seurat/	R (Butler et al., 2018)
	Monocle	Bioconductor	R (Trapnell et al., 2014; Qiu et al., 2017a; Qiu et al., 2017b)
	DPT	http://www.helmholtz-muenchen.de/icb/dpt	R & Matlab (Haghverdi et al., 2016)
B-cell receptor reconstruction	VDJPuzzle	bitbucket	R & Python (Rizzetto et al., 2018)
	bracer	Github	Python (Lindeman et al., 2018)
Network inference	SCODE	Github	R (Matsumoto et al., 2017)
	LEAP	CRAN	R (Specht and Li, 2017)

in biological research (Marinov et al., 2014; Lin C. et al., 2017; Jin et al., 2018; Kiselev et al., 2019). Different methods use distinct characteristics of data and gain varying outcomes in terms of both the number of clusters and the cluster assignment of cells (Ntranos et al., 2016; Kim et al., 2018; Risso et al., 2018). Many approaches, such as SAFE clustering (Yang et al., 2018), DendroSplit (Zhang J. et al., 2018), scmap (Kiselev et al., 2018), MetaNeighbor (Crow et al., 2018), scVDMC (Zhang H. et al., 2018), CIDR (Lin P. et al., 2017), SC3 (Kiselev et al., 2017), scLVM (Buettner et al., 2015), and RaceID (Grun et al., 2015), have been developed to promote the efficiency of clustering single cells. They promote the clustering consensus, interpretability, subjectivity, comparability, and replicability. However, the biological significance, number estimation, and computational speed of such clustering analysis still require significant improvements (Duan et al., 2018).

Challenge of Identity Analysis of Single Cells

scRNA-seq has brought transcriptome research to a higher resolution as the "up or down" expression pattern can be examined at the single-cell level (Chen L. et al., 2018; Xie et al., 2019). The projection of high-dimensional data into a low-dimensional subspace will be a powerful strategy for mining such extensive data (Zeng et al., 2016; Yip et al., 2018; Yu and Zeng, 2018). Statistic-based approaches, such as PowsimR (Vieth

et al., 2017), BPSC (Vu et al., 2016), Linnorm (Yip et al., 2017), and Oscope (Leng et al., 2015), have been established to evaluate differential expression among individual cells. Especially, latent factor-based analysis will be useful to find hidden biological signals and corresponding gene components from scRNA-seq samples (Buettner et al., 2017; Yu, 2018). However, to guarantee the biological meaning of detected cell identities, it is still necessary to discriminate the real and dropout zeros in scRNA-seq data (Miao et al., 2018). It is also essential to identify the combination of binary and continuous regulation in individual cells (Wu et al., 2018) and to integrate the nonlinear projection with prior-known biological knowledge (Li X. et al., 2017).

Challenge of Trajectory Analysis of Single Cells

The single-cell experiments provide a great chance to rebuild a sequence of changes in a dynamical process of the biological system from individual "snapshots" of cells (Matsumoto et al., 2017; Gong et al., 2018). The construction of a pseudo-temporal path as cell orders would be a useful way to characterize dynamical gene expression in a heterogeneous cell population, assuming the existence hypothesis of gradual transition of the cell transcriptome (Specht and Li, 2017; Herring et al., 2018; Shindo et al., 2018; Strauss et al., 2018). For example, based on the minimum spanning tree approach, the Tools for Single Cell

Analysis is developed for in silico pseudo-time reconstruction in scRNA-seq analysis (Ji and Ji, 2016). As an iterative supervised learning algorithm, FateID can recognize the cell fate preference by quantifying the lineage-specific probabilistic biases (Herman et al., 2018). By unsupervisedly selecting feature genes and judging the location and number of branches and loops, SLICER is able to infer highly nonlinear trajectories (Welch et al., 2016). However, many opportunities still exist to develop these current methods, particularly detecting complex trajectory topologies, linking pseudo-time and real-world time, determining baseline points, estimating transition possibility, and recognizing progression trends with tipping point (Zeng et al., 2013).

Challenge of Origin Analysis of Single Cells

The origin and nature of signals leading to pattern formation and self-organization is an essential question in developmental or stem cell biology. The answer would be recovered from the gene expressions of individual cells with spatial locations in a particular tissue (Vergara et al., 2017; Chen Q. et al., 2018). On the one hand, from the technological point of view, several methods have been designed for recording the spatial information of cells. The spatial transcriptomic technology and computational deconvolution can be combined to detect distinct expression profiles corresponding to different tissue components (Berglund et al., 2018). One technique that performs RT-LAMP reactions on a histological tissue section can preserve the original spatial location of the nucleic acid molecules to become an effective tissue analysis tool (Ganguli et al., 2018). Another technique is based on a panel of zonated landmark genes, where the lobule coordinates of mouse liver cells can be inferred according to their transcriptome, whereas the zonation profiles of all liver genes can also be characterized with high spatial resolution (Halpern et al., 2017). On the other hand, from the analytic point of view, supervised methods have been shown to be efficient, inferring the potential spatial distribution of cells. On the foundation of a reference gene expression database, e.g., the gene expression atlas for positional gene expression profiles within cells, an scRNA-seq-based high-throughput method has been applied to identify the spatial origin of cells (Achim et al., 2015). Obviously, spatial labeling technologies still need further technological developments for more easy and accurate testing, and the spatial classification and prediction of cells require more elaborate and efficient mathematical and computational models.

Challenge of Integrative Analysis of Single Cells

Understanding the genetic and cellular processes and programs driving the differentiation of diverse cell types and organ formation is a major challenge in developmental biology (Kelsey et al., 2017, Velten et al., 2017, Duren et al., 2018, Liu L. et al., 2018). Frameworks and software are required to perform dimension reduction, clustering, and visualization on scRNA-seq data to improve biological interpretability (Gardeux et al., 2017; Wang et al., 2017). Numerous methods have been implemented for analyzing scRNA-seq data in a whole life-cycle manner (Guo et al., 2015; Diaz et al., 2016; Leng et al., 2016; Yang et al., 2017). SparseDC solves a unified optimization problem so that it can carry out three tasks simultaneously, e.g., identifying cell types,

tracing expression changes across conditions, and identifying marker genes for these changes (Barron et al., 2018). BigSCale implements a scalable analytical framework to handle millions of cells, so it can overcome large data challenges by the directed down-sampling strategy on index cell transcriptomes (Iacono et al., 2018). In addition to these usual analytic routines for conventional targets, more diverse integration models are required for data-driven, model-driven, hypothesis-driven, and combinatory bioinformatics mining in single-cell data.

Understanding Disease Heterogeneity by scRNA-seq Analysis

For questions in the biological and biomedical fields, human cancers are especially considered complex ecosystems where the basic elements (cells) exist in different disease states characterized by phenotypes and genotypes. As is well known, conventional methods have their limits when measuring and quantifying the diverse tumor (cell) composition in patients, e.g., traditional bulk expression profiles have to average the cells within each tumor. Nowadays, scRNA-seq provides a powerful technique to detect critical cell differences and deconvolve such cellular heterogeneity in disease tissues. Therefore, one important benefit obtained from scRNA-seq is the possibility to decipher tumor architecture (Cloney 2017), so that it might overcome intratumoral heterogeneity, which hampers the success of precision medicine and is therefore a huge challenge in cancer treatment (Patel et al., 2014; Kim et al., 2016; Zong, 2017). Actually, in the context of cancer, mRNA can be used to identify malignant cells and diverse tumor-tissue compositions; such tumor compositions could indicate the cancer-associated cells and types determining tumor characteristics (Young et al., 2018). Thus, scRNA-seq-based methods could be widely applied in clinical decision support (Tirosh et al., 2016a; Filbin et al., 2018; Krieg et al., 2018; Pellegrino et al., 2018).

i) *Tumor mechanism investigation.* One general framework can be used to decipher differences between multiple classes of human tumors by decoupling cancer cell genotypes, phenotypes, and the composition of tumor microenvironment (Venteicher et al., 2017). One single-cell analysis method has provided some insights into the cellular architecture of oligodendrogliomas and their function in development regulation, which potentially is compatible with the cancer stem cell model and its consideration in disease management (Tirosh et al., 2016b).

ii) *Tumor subtype recognition.* To deconvolve the cellular composition of a solid tumor from bulk gene expression data using reference gene expression profiles from tumor-derived scRNA-seq data, many cell types or subtypes must be identified accurately (Schelker et al., 2017). For example, one scRNA-seq study of triple-negative breast cancer identified the individual subpopulations with respective gene expression phenotypes and corresponding genotype driver candidates, whose associated signature genes can predict long-term outcomes (Karaayvaz et al., 2018).

iii) *Tumor immune therapy.* Single-cell analyses have suggested distinct patterns in the tumor microenvironment, e.g., the breast cancer transcriptome has shown a wide range of

intratumoral heterogeneity that is reshaped by both immune and tumor cells in a closely communicated microenvironment at a single-cell resolution (Chung et al., 2017). An unbiased scRNA-seq analysis has detected human dendritic cells and several monocyte subtypes in the human blood to permit more accurate immune monitoring in health and disease (Villani et al., 2017). In a more special field, the single-cell transcriptional information in B-cell lineages might have broad applications involved in vaccine design, antibody development, and cancer treatment (Rizzetto et al., 2018; Upadhyay et al., 2018).

iv) *Tumor virus-environment recognition*. Indeed, the interaction between a host and a pathogen is a highly dynamical process, so the potential association between a pathogen and cancer is worthy of profound investigation. An scRNA-seq-based method, scDual-Seq, has been proposed to capture host and pathogen transcriptomes simultaneously (Avital et al., 2017). In different mouse models, the hypothetical virus-host interaction events have been found to play some key regulatory role in virus phenotypes involved in complex diseases by tracking viral RNA at single-cell resolution within the immune system (Douam et al., 2017).

Of course, the translational usage of scRNA-seq is not limited to the field of tumor biology or complex human diseases; it is expected to have great potential and to enjoy a wide range of applications in biological and biomedical fields, such as infant development, health and wellness, and disease monitoring.

Design of Hypothesis and Theory Study on scRNA-seq Analysis Robustness

As is well known, scRNA-seq analysis is used to compare the expression levels of multiple genes at single-cell resolution (Tang et al., 2009). Different from the conventional population-based biological technologies for gene expression measurement (e.g., bulk gene expression), scRNA-seq is able to distinguish the expression differences between individual cells rather than tissues. With the continuous development of such technology, the testing cost is decreasing, whereas the number of cells that can simultaneously be tested is increasing exponentially. Some recent reviews have summarized these technological developments and protocol improvements in scRNA-seq analyses (Svensson et al., 2017; Ziegenhain et al., 2017; Svensson et al., 2018). An inspiring observation is that the number of tested cells and the number of detected genes can vary significantly depending on the corresponding experimental platforms. For example, SMART-seq2 is able to detect about 10,000 genes and achieve the highest accuracy, but the number of cells analyzed by this method is only 100 to 1,000 (Picelli et al., 2013; Picelli et al., 2014). In contrast, Drop-seq is able to test more than 10,000 cells simultaneously, but the number of genes detected is usually less than 5,000 (Macosko et al., 2015). Recently, several commercial platforms, such as 10X Genomics Chromium, Fluidigm C1, and Wafergen ICELL8, were available for scRNA-seq analysis with the capability to measure hundreds to millions of cells through a simple and fast workflow.

Researchers are usually required to select the suitable experimental protocol to design the follow-up scRNA-seq analysis based on corresponding biological questions:

i) If one aims to discover new cell types with distinct expression patterns, more cells should be tested because it is impossible to find rare cell types from only a few hundred cells by chance.
ii) If one aims to analyze the changes in gene expression between different cell types or developmental stages or to analyze the gene interactions to find some key regulatory genes, more genes have to be measured with high accuracy.
iii) If one aims to analyze particular cell types by isolating a subset of cells for sequencing, fluorescence-activated cell sorting or a similar technology needs to be used to select the cells with cell type-specific cell surface markers.

To evaluate and investigate the robustness of different scRNA-seq analysis methods, we have carried out two comparisons on multiple scRNA-seq datasets.

The aim of the first comparison is to discuss the experimental factors for scRNA-seq analysis. As is well known, the accuracy of RNA-seq data analysis is dependent on the experimental methods, especially the sequencing depth and dropout rate. To test these experimental factors before further evaluation, we compared four datasets on two different experimental platforms: GSE81608 (Xin et al., 2016) and GSE83139 (Wang et al., 2016) on an Illumina HiSeq 2500 and GSE86469 (Lawlor et al., 2017) and GSE81547 (Enge et al., 2017) on an Illumina NextSeq 500. All of these datasets come from the single-cell studies of human pancreatic islet cells so that their computational results will be comparable, and the number of clusters for each method was fixed to be the same as the number of biological classes corresponding to each dataset, as shown in **Table 2**.

The aim of the second comparison is to discuss the analytic approaches for scRNA-seq analysis. The performance of dissimilar methods on different real datasets of the same complex disease is important to evaluate, because performance robustness will be strictly required for biomedical studies and applications. Thus, we have employed several widely used methods in a few public scRNA-seq datasets from complex disease studies, which are listed in **Table 3**. According to the above summary, we actually evaluated the performances on cell cluster, cell identity, and cell trajectory. These methods' parameter settings are listed in the supplementary files (**Supp 1**).

i) For cell clustering analysis, traditional methods, such as hierarchical clustering, k-means, and scRNA-seq-induced SIMLR (Wang et al., 2017; Wang B. et al., 2018), SNN-Cliq (Xu and Su, 2015), and SEURAT (Butler et al., 2018) have been evaluated and compared.
ii) For cell pseudo-time analysis, the Monocle (Trapnell et al., 2014; Qiu et al., 2017a; Qiu et al., 2017b) and diffusion pseudo-time (DPT) (Haghverdi et al., 2016) have been tested and compared.

Of note, to quantitatively measure and compare the analysis accuracy of cell clusters from different methods, the conventional adjusted rand index (ARI) is applied. Given a dataset of *n* cells,

TABLE 2 | Clustering performances of four datasets with different experiment methods represented as adjusted rand index (ARI).

	GSE81547	GSE83139	GSE81608	GSE86469
Experiment platforms	NextSeq 500	HiSeq 2500	HiSeq 2500	NextSeq 500
Number of cells	2,282	635	1,600	617
Number of detected genes per cell on average	3,281	5,638	5,706	8,339
Number of potential cell types*	6	8	4	7
Hierarchical clustering	0.34	0.25	0.46	0.63
k-means	0.34	0.27	0.44	0.48
tSNE+k-means	0.37	0.34	0.54	0.72
SIMLR	0.34	0.32	0.51	0.61
SNN-Cliq	0.10	0.31	0.05	0.61
SEURAT	0.31	0.31	0.45	0.89

*GSE81547 includes alpha cells, beta cells, delta cells, acinar cells, mesenchyme cells, and ductal cells. GSE83139 includes alpha cells, beta cells, delta cells, PP cells, acinar cells, mesenchyme cells, ductal cells, and dropped cells. GSE81608 includes alpha cells, beta cells, delta cells, and PP cells. GSE86469 includes alpha cells, beta cells, delta cells, PP cells, acinar cells, stellate cells, and ductal cells.

TABLE 3 | Summary of evaluation datasets on human complex diseases.

Data ID	Purpose	Platform	#scRNA-Seq	#Class
GSE69405	scRNA-seq identifies subclonal heterogeneity in anticancer drug responses of lung adenocarcinoma cells	HiSeq 2500	176	3
GSE73121	scRNA-seq in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma	HiSeq 2500	118	3
GSE81608	scRNA-seq on human islet cells revealing type 2 diabetes genes	HiSeq 2500	1600	4
GSE83139	scRNA-seq of the human endocrine pancreas	HiSeq 2500	635	8

the experimentally determined cell types are $X_1, X_2, \ldots, X_r$ and the calculated clusters are $Y_1, Y_2, \ldots, Y_s$. The number of cells that belong to cell type X_i is denoted as a_i, the number of cells that belong to cluster Y_j is denoted as b_j, and the number of cells that belong to both X_i and Y_j is denoted as n_{ij}, which means $n_{ij} = |X_i \cap Y_j|$. Then, the ARI is calculated as follows:

$$ARI = \frac{\Sigma_{ij}\binom{n_{ij}}{2} - \left[\Sigma_i\binom{a_i}{2} \cdot \Sigma_j\binom{b_j}{2}\right] \Big/ \binom{n}{2}}{\frac{1}{2}\left[\Sigma_i\binom{a_i}{2} + \Sigma_j\binom{b_j}{2}\right] - \left[\Sigma_i\binom{a_i}{2} \cdot \Sigma_j\binom{b_j}{2}\right] \Big/ \binom{n}{2}}.$$

RESULTS AND DISCUSSION

Experimental Factors for scRNA-seq Analysis

The experimental processes of the four datasets presented in **Table 2** are briefly summarized below.

1) For GSE81608 (Xin et al., 2016), islets were handpicked and enzymatically digested; during RNA *in situ* hybridization, the cells were permeabilized and hybridized with combinations of mRNA probes and a multiplex fluorescent kit was used to amplify the mRNA signal. Sequencing was performed on an Illumina HiSeq2500 in rapid mode by multiplexed single-read run with 50 cycles.

2) For GSE83139 (Wang et al., 2016), human islets require careful sample acquisition and preparation; the SMART-seq

method was used for first-strand cDNA synthesis and polymerase chain reaction (PCR) amplification. All of the libraries were sequenced on the Illumina HiSeq 2500 with 100 bp single-end reads.

3) For GSE86469 (Lawlor et al., 2017), islets are systematically acquired, processed, and dissociated; then, single-cell processing is carried out on the C1 single-cell Autoprep system. All of the sequencing was performed on an Illumina NextSeq500 using the 75-cycle high-output chip.

4) For GSE81547 (Enge et al., 2017), the experimental models and human pancreas or islet samples were conducted in accordance with guidelines; during flow cytometry, isolated human islets were dissociated into single cells by enzymatic digestion using Accumax (Invitrogen). Next, single-cell RNA-seq libraries were generated as described in the literature, and barcoded libraries were pooled and subjected to 75 bp paired-end sequencing on the Illumina NextSeq instrument.

Of course, the whole experimental process should be consistent; however, the scRNA-seq wet experiments in different studies were conducted with different parameters and under different circumstances, which are worthy of future evaluation. Although sequencing platforms are only one part of the scRNA-seq experiment, we tried to include them for the comparison study in this work. In **Table 2**, we see that there is no obvious performance difference between two experiment platforms; however, the accuracy (i.e., ARI) seems to increase when the number of detected genes becomes large for almost all of the tested methods, which is consistent with a previous conclusion (Potter, 2018) and implies that the influence of sequencing depth is very important

in the experimental protocol for follow-up data analysis. Of note, the parameter setting for each compared method in this work is outlined in the supplementary files (**Supp 1**).

Analytic Approaches for scRNA-seq Analysis

First, it can be seen that the datasets after dimension reduction by t-distributed stochastic neighbor embedding (tSNE) (Maaten and Hintton, 2008) exhibit better performances in conventional k-means clustering than the initial dataset, which is due to the noise reduction of scRNA-seq data. Dimension reduction can be used in the visualization of such phenomena, which reduces one dataset from high-dimensional data space to two- or three-dimensional data space. **Figure 1A** illustrates the performances of principal component analysis (PCA) and tSNE on multiple datasets. It is clear that tSNE, a nonlinear method, can usually achieve better visualization effects than PCA, a linear method. This is because tSNE can group the cell points from one class cluster together and keep the cell points from different classes separated from each other. The quantitative measurement of the influence of PCA and tSNE by the Davies-Bouldin index also supported this conclusion, as shown in the supplementary files (**Supp 2**). Of note, due to the large computational complexity of nonlinear methods, the general strategy for large data analysis includes two steps. The first is to reduce the dimension to 20 to 50 by PCA, and the second is to reduce such moderate dimension to 2 to 3 by tSNE. This strategy is expected to achieve a good balance between computational performance and resource consumption.

Second, in the cell clustering analysis, the analyzed genes are selected that exhibit expression in at least three cells, so that most genes have actually been used. For hierarchical clustering, k-means, tSNE+k-means, and SIMLR, the number of clusters for each method was fixed to be the same as the number of biological classes corresponding to each dataset, as shown in **Table 3**. For SNN-Cliq and SEURAT, the parameters were adjusted to guarantee that the number of final clusters was the same as the number of biological classes in those datasets, as shown in **Table 3**. In other words, the number of clusters for every method is the same for one dataset to make different methods fairly comparable to ARI. As seen in **Figure 1B**, it is obvious the performances of tSNE+k-means, SIMLR, and SEURAT were better than those of others with higher ARI values in most scRNA-seq datasets. In addition, although tSNE+k-means, SIMLR, and SEURAT have similar performances with regard to ARI, they usually accurately detected different true classes (**Figure 1C**). This means different methods would have different analysis preferences due to different underlying mathematical or biological frameworks and explanations of scRNA information.

Third, scRNA-seq data follow a time series and the expression of cells may change continuously. For this kind of dataset, some statistical methods can be used to order the cells one by one along a trajectory, which is called pseudo-time or pseudo-trajectory. This mathematical model has been widely applied in developmental biology to reconstruct the differentiation processes and find the key time point of differentiation (Cannoodt et al., 2016). In addition, cell pseudo-time analysis can also be used in studies of cancer and diabetes to reconstruct the occurrence and

transformation processes of complex diseases. Thus, the Monocle and DPT have been carried out for pseudo-time analysis on multiple scRNA-seq datasets; these two computational methods are dependent on entirely different principles. In this cell pseudo-time analysis, the most expression-variable genes are selected as feature genes for downstream analysis. As shown in **Figure 1D**, the feature genes exhibit great differences between datasets with different biological backgrounds; however, the two datasets on similar biological phenotypes still have much overlap (i.e., the feature genes from two datasets related to tumor cells with treatments or those from two datasets associated with diabetes). Of note, using human pancreas scRNA-seq datasets in another platform (i.e., GSE86469 and GSE81547; **Table 2**) as controls, the top 50 selected feature genes from the total four datasets indeed had more overlapping genes, as listed in the supplementary files (**Supp 3**). In **Figure 1E**, it is seen that both Monocle and DPT are able to reconstruct the pseudo-time with branches, and DPT seems to obtain more accurate results as the cells of the same cell type tend to group together. Meanwhile, the pseudo-time and branch point seem to be clearer in the analyses of Monocle. Of note, the performance of pseudo-time analysis will be strongly influenced by the selected feature genes. In this comparison, the most expression-variable genes were used, but usually it would be much better to select the feature genes based on the prior biological knowledge in each case study. Furthermore, the consistency of pseudo-time results from different methods is considered and evaluated. As shown in **Figure 2**, the correlations between the first principal components of the pseudo-time results from Monocle and DPT have been calculated. Then, the estimation similarities of cell orders in particular cell classes from different methods are compared. It is obvious that the cell order correlations have huge variances in a wide range among different prior-known cell classes. In addition, two other pseudo-time methods, Wanderlust (Bendall et al., 2014) and SCUBA (Marco et al., 2014), were also applied to reconstruct the pseudo-time trajectory of single cells without branch, as discussed in the supplementary files (**Supp 4**). The observations and conclusions were similar. Thus, in the pseudo-time analysis, consensus performance of dissimilar methods is weak currently.

CONCLUSION

scRNA-seq has opened a new way to study complex biological phenomena on the single-cell level, which will be especially helpful in the research of complex diseases. However, to enhance its performance in actual applications, e.g., in the clinic, several improvements are still required. For cell clustering and identification, gene networks rather than separate genes would be more important and reliable to characterize cell states (e.g., network biomarkers for disease subtypes) (Zeng et al., 2014; Zeng et al., 2016). For the cell order, the start or end point of pseudo-time is still a manual judgment, and the auto-determination of these time points will render these methods more flexible and applicable (e.g., temporal driving for disease causality) (Yu et al., 2017; Wang et al., 2018; Setty et al., 2019). The branch point of pseudo-time also requires more models on critical transitions (e.g., tipping point for
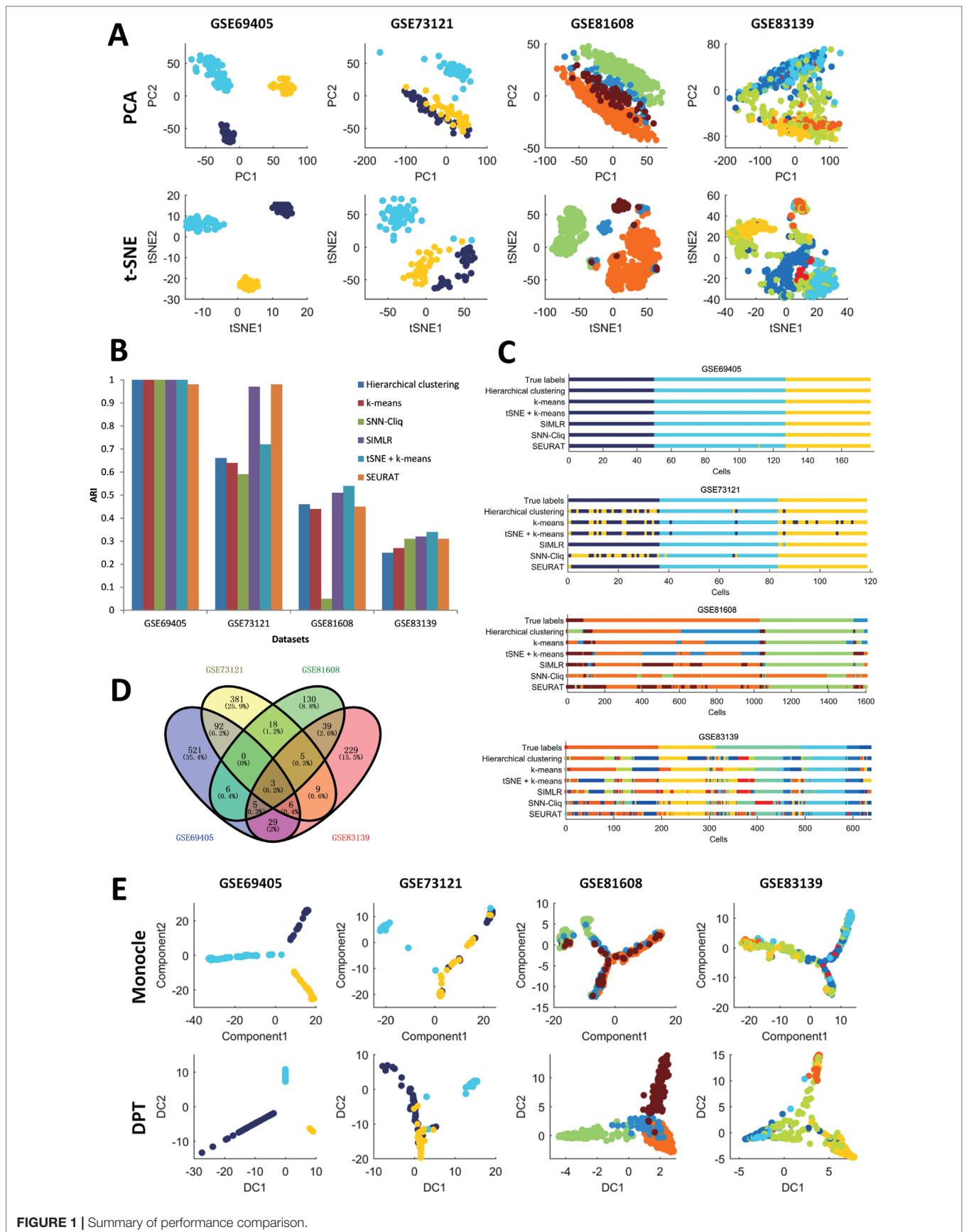
FIGURE 1 | Summary of performance comparison.

ID	GSE69405	GSE73121	GSE81608	GSE83139
Class 1	*H358 human lung cancer cells, 50 cells* **0.15**	*PDX - metastatic renal cell carcinoma, 36 cells* **0.65**	*Pancreatic PP cells, 93 cells* **0.06**	*Acinar cells, 6 cells* **0.77**
Class 2	*Tumor cell-enriched PDX cells, 77 cells* **0.79**	*PDX - primary renal cell carcinoma, 47 cells* **0.84**	*Pancreatic α cells, 946 cells* **0.62**	*α cells, 190 cells* **0.62**
Class 3	*Another lung cancer PDX case, 49 cells* **0.14**	*Parental - metastatic renal cell carcinoma, 35 cells* **0.78**	*Pancreatic β cells, 503 cells* **0.23**	*β cells, 111 cells* **0.65**
Class 4			*Pancreatic δ cells, 58 cells* **0.13**	*δ cells, 9 cells* **0.81**
Class 5				*Dropped cells, 178 cells* **0.40**
Class 6				*Duct cells, 96 cells* **0.26**
Class 7				*Mesenchyme cells, 27 cells* **0.54**
Class 8				*PP cells, 18 cells* **0.80**

FIGURE 2 | Summary of robustness comparison.

disease transition) (Zeng et al., 2013; Li et al., 2014). Particularly, the assembling method with good consensus on different datasets is expected to provide more robust integrative scRNA-seq methods for biological and biomedical studies (e.g., pattern fusion for disease heterogeneity) (Shi et al., 2017; Guo et al., 2018).

AUTHOR CONTRIBUTIONS

TZ conceived the concept and design of the work. HD and TZ performed the experiments. TZ and HD analyzed the results. TZ drafted the manuscript. TZ and HD revised the paper.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00629/full#supplementary-material

SUPP 1 | Parameter settings of scRNA-seq analysis methods.

SUPP 2 | Feature genes of four human pancreas scRNA-seq datasets.

SUPP 3 | Quantitative measurement of PCA and tSNE.

SUPP 4 | Additional pseudo-time methods on four scRNA-seq datasets.

REFERENCES

Achim, K., Pettit, J. B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., et al. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33 (5), 503–509. doi: 10.1038/nbt.3209

Avital, G., Avraham, R., Fan, A., Hashimshony, T., Hung, D. T., and Yanai, I. (2017). scDual-Seq: mapping the gene regulatory program of *Salmonella* infection by host and pathogen single-cell RNA-sequencing. *Genome Biol.* 18 (1), 200. doi: 10.1186/s13059-017-1340-x

Bacher, R., and Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 17, 63. doi: 10.1186/s13059-016-0927-y

Bacher, R., Chu, L. F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., et al. (2017). SCnorm: robust normalization of single-cell RNA-seq data. *Nat. Methods* 14 (6), 584–586. doi: 10.1038/nmeth.4263

Bagnoli, J. W., Ziegenhain, C., Janjic, A., Wange, L. E., Vieth, B., Parekh, S., et al. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat. Commun.* 9 (1), 2937. doi: 10.1038/s41467-018-05347-6

Baran-Gale, J., Chandra, T., and Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Brief Funct. Genom.* 17 (4), 233–239. doi: 10.1093/bfgp/elx035

Barron, M., Zhang, S., and Li, J. (2018). A sparse differential clustering algorithm for tracing cell type changes *via* single-cell RNA-sequencing data. *Nucleic Acids Res.* 46 (3), e14. doi: 10.1093/nar/gkx1113

Bendall, S. C., Davis, K. L., Amir el, A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., et al. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157 (3), 714–725. doi: 10.1016/j.cell.2014.04.005

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstrahle, J., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* 9 (1), 2419. doi: 10.1038/s41467-018-04724-5

Birnbaum, K. D. (2018). Power in numbers: single-cell RNA-seq strategies to dissect complex tissues. *Annu. Rev. Genet.* 23 (52), 203–221. doi: 10.1146/annurev-genet-120417-031247

Bose, S., Wan, Z., Carr, A., Rizvi, A. H., Vieira, G., Pe'er, D., et al. (2015). Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* 16, 120. doi: 10.1186/s13059-015-0684-3

Brennecke, P., Anders, S., Kim, J. K., Kolodziejczyk, A. A., Zhang, X., Proserpio, V., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10 (11), 1093–1095. doi: 10.1038/nmeth.2645

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., et al. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33 (2), 155–160. doi: 10.1038/nbt.3102

Buettner, F., Pratanwanich, N., McCarthy, D. J., Marioni, J. C., and Stegle, O. (2017). f-scLVM: scalable and versatile factor analysis for single-cell RNA-seq. *Genome Biol.* 18 (1), 212. doi: 10.1186/s13059-017-1334-8

Burns, J. C., Kelly, M. C., Hoa, M., Morell, R. J., and Kelley, M. W. (2015). Single-cell RNA-seq resolves cellular complexity in sensory organs from the neonatal inner ear. *Nat. Commun.* 6, 8557. doi: 10.1038/ncomms9557

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420. doi: 10.1038/nbt.4096

Cacchiarelli, D., Qiu, X., Srivatsan, S., Manfredi, A., Ziller, M., Overbey, E., et al. (2018). Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome. *Cell Syst.* 7 (3), 258–268. doi: 10.1016/j.cels.2018.07.006

Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* 46 (11), 2496–2506. doi: 10.1002/eji.201646347

Chen, L., and Zheng, S. (2018). BCseq: accurate single cell RNA-seq quantification with bias correction. *Nucleic Acids Res.* 46 (14), e82. doi: 10.1093/nar/gky308

Chen, Q., Shi, J., Tao, Y., and Zernicka-Goetz, M. (2018). Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. *Nat. Commun.* 9 (1), 1819. doi: 10.1038/s41467-018-04155-2

Chen, W., Li, Y., Easton, J., Finkelstein, D., Wu, G., and Chen, X. (2018). UMI-count modeling and differential expression analysis for single-cell RNA sequencing. *Genome Biol.* 19 (1), 70. doi: 10.1186/s13059-018-1438-9

Chu, L. F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D. T., et al. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* 17 (1), 173. doi: 10.1186/s13059-016-1033-x

Chung, W., Eum, H. H., Lee, H. O., Lee, K. M., Lee, H. B., Kim, K. T., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* 8, 15081. doi: 10.1038/ncomms15081

Clark, S. J., Lee, H. J., Smallwood, S. A., Kelsey, G., and Reik, W. (2016). Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.* 17, 72. doi: 10.1186/s13059-016-0944-x

Cloney, R. (2017). Cancer genomics: single-cell RNA-seq to decipher tumour architecture. *Nat. Rev. Genet.* 18 (2), 2–3. doi: 10.1038/nrg.2016.151

Cole, M. B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., et al. (2017). Performance assessment and selection of normalization procedures for single-cell RNA-Seq. *bioRxiv*. doi: 10.1101/235382. [Epub ahead of print].

Crow, M., Paul, A., Ballouz, S., Huang, Z. J., and Gillis, J. (2018). Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* 9 (1), 884. doi: 10.1038/s41467-018-03282-0

Dal Molin, A., Baruzzo, G., and Di Camillo, B. (2017). Single-cell RNA-sequencing: assessment of differential expression analysis methods. *Front. Genet.* 8, 62. doi: 10.3389/fgene.2017.00062

Dal Molin, A., and Di Camillo, B. (2018). How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. *Brief Bioinform.* doi: 10.1093/bib/bby007. [Epub ahead of print].

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343 (6167), 193–196. doi: 10.1126/science.1245316

Diaz, A., Liu, S. J., Sandoval, C., Pollen, A., Nowakowski, T. J., Lim, D. A., et al. (2016). SCell: integrated analysis of single-cell RNA-seq data. *Bioinformatics* 32 (14), 2219–2220. doi: 10.1093/bioinformatics/btw201

Ding, B., Zheng, L., Zhu, Y., Li, N., Jia, H., Ai, R., et al. (2015). Normalization and noise reduction for single cell RNA-seq experiments. *Bioinformatics* 31 (13), 2225–2227. doi: 10.1093/bioinformatics/btv122

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167 (7), 1853–1866 e1817. doi: 10.1016/j.cell.2016.11.038

Douam, F., Hrebikova, G., Albrecht, Y. E., Sellau, J., Sharon, Y., Ding, Q., et al. (2017). Single-cell tracking of flavivirus RNA uncovers species-specific interactions with the immune system dictating disease outcome. *Nat. Commun.* 8, 14781. doi: 10.1038/ncomms14781

Duan, T., Pinto, J. P., and Xie, X. (2018). Parallel clustering of single cell transcriptomic data with split-merge sampling on Dirichlet process mixtures. *Bioinformatics* 35 (6), 953–961. doi: 10.1093/bioinformatics/bty702

Duren, Z., Chen, X., Zamanighomi, M., Zeng, W., Satpathy, A. T., Chang, H. Y., et al. (2018). Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proc. Natl. Acad. Sci. U. S. A.* 115 (30), 7723–7728. doi: 10.1073/pnas.1805681115

Edsgard, D., Reinius, B., and Sandberg, R. (2016). scphaser: haplotype inference using single-cell RNA-seq data. *Bioinformatics* 32 (19), 3038–3040. doi: 10.1093/bioinformatics/btw484

Eling, N., Richard, A. C., Richardson, S., Marioni, J. C., and Vallejos, C. A. (2018). Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data. *Cell Syst.* 7 (3), 284–294. doi: 10.1016/j.cels.2018.06.011

Enge, M., Arda, H. E., Mignardi, M., Beausang, J., Bottino, R., Kim, S. K., et al. (2017). Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 171 (2), 321–330 e314. doi: 10.1016/j.cell.2017.09.004

Fan, X., Zhang, X., Wu, X., Guo, H., Hu, Y., Tang, F., et al. (2015). Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* 16, 148. doi: 10.1186/s13059-015-0706-1

Faridani, O. R., Abdullayev, I., Hagemann-Jensen, M., Schell, J. P., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* 34 (12), 1264–1266. doi: 10.1038/nbt.3701

Filbin, M. G., Tirosh, I., Hovestadt, V., Shaw, M. L., Escalante, L. E., Mathewson, N. D., et al. (2018). Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360 (6386), 331–335. doi: 10.1126/science.aao4750

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16, 278. doi: 10.1186/s13059-015-0844-5

Fuzik, J., Zeisel, A., Mate, Z., Calvigioni, D., Yanagawa, Y., Szabo, G., et al. (2016). Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* 34 (2), 175–183. doi: 10.1038/nbt.3443

Ganguli, A., Ornob, A., Spegazzini, N., Liu, Y., Damhorst, G., Ghonge, T., et al. (2018). Pixelated spatial gene expression analysis from tissue. *Nat. Commun.* 9 (1), 202. doi: 10.1038/s41467-017-02623-9

Gardeux, V., David, F. P. A., Shajkofci, A., Schwalie, P. C., and Deplancke, B. (2017). ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. *Bioinformatics* 33 (19), 3123–3125. doi: 10.1093/bioinformatics/btx337

Gasch, A. P., Yu, F. B., Hose, J., Escalante, L. E., Place, M., Bacher, R., et al. (2017). Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol.* 15 (12), e2004050. doi: 10.1371/journal.pbio.2004050

Gong, W., Kwak, I. Y., Koyano-Nakagawa, N., Pan, W., and Garry, D. J. (2018). TCM visualizes trajectories and cell populations from single cell data. *Nat. Commun.* 9 (1), 2749. doi: 10.1038/s41467-018-05112-9

Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., et al. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525 (7568), 251–255. doi: 10.1038/nature14966

Guo, M., Wang, H., Potter, S. S., Whitsett, J. A., and Xu, Y. (2015). SINCERA: a pipeline for single-cell RNA-Seq profiling analysis. *PLoS Comput. Biol.* 11 (11), e1004575. doi: 10.1371/journal.pcbi.1004575

Guo, W. F., Zhang, S. W., Liu, L. L., Liu, F., Shi, Q. Q., Zhang, L., et al. (2018). Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. *Bioinformatics* 34 (11), 1893–1903. doi: 10.1093/bioinformatics/bty006

Haghverdi, L., Lun, A. T. L., Morgan, M. D., and Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36 (5), 421–427. doi: 10.1038/nbt.4091

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* 13 (10), 845–848. doi: 10.1038/nmeth.3971

Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542 (7641), 352–356. doi: 10.1038/nature21065

Han, X., Chen, H., Huang, D., Chen, H., Fei, L., Cheng, C., et al. (2018). Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol.* 19 (1), 47. doi: 10.1186/s13059-018-1426-0

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17, 77. doi: 10.1186/s13059-016-0938-8

Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaido, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* 9 (1), 619. doi: 10.1038/s41467-018-02866-0

Hemberg, M. (2018). Single-cell genomics. *Brief Funct. Genom.* 17 (4), 207–208. doi: 10.1093/bfgp/ely025

Herman, J. S., Sagar, , and Grun, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods* 15 (5), 379–386. doi: 10.1038/nmeth.4662

Herring, C. A., Banerjee, A., McKinley, E. T., Simmons, A. J., Ping, J., Roland, J. T., et al. (2018). Unsupervised trajectory analysis of single-cell RNA-Seq and imaging data reveals alternative tuft cell origins in the gut. *Cell Syst.* 6 (1), 37–51 e39. doi: 10.1016/j.cels.2017.10.012

Iacono, G., Mereu, E., Guillaumet-Adkins, A., Corominas, R., Cusco, I., Rodriguez-Esteban, G., et al. (2018). bigSCale: an analytical framework for big-scale single-cell data. *Genome Res.* 28 (6), 878–890. doi: 10.1101/gr.230771.117

Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., et al. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 17, 29. doi: 10.1186/s13059-016-0888-1

Jaitin, D. A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., et al. (2016). Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq. *Cell* 167 (7), 1883–1896 e1815. doi: 10.1016/j.cell.2016.11.039

Ji, Z., and Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44 (13), e117. doi: 10.1093/nar/gkw430

Jia, C., Hu, Y., Kelly, D., Kim, J., Li, M., and Zhang, N. R. (2017). Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. *Nucleic Acids Res.* 45 (19), 10978–10988. doi: 10.1093/nar/gkx754

Jiang, P., Thomson, J. A., and Stewart, R. (2016). Quality control of single-cell RNA-seq by SinQC. *Bioinformatics* 32 (16), 2514–2516. doi: 10.1093/bioinformatics/btw176

Jiang, Y., Zhang, N. R., and Li, M. (2017). SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol.* 18 (1), 74. doi: 10.1186/s13059-017-1200-8

Jin, S., MacLean, A. L., Peng, T., and Nie, Q. (2018). scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 34 (12), 2077–2086. doi: 10.1093/bioinformatics/bty058

Julia, M., Telenti, A., and Rausell, A. (2015). Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. *Bioinformatics* 31 (20), 3380–3382. doi: 10.1093/bioinformatics/btv368

Kang, H. M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36 (1), 89–94. doi: 10.1038/nbt.4042

Karaayvaz, M., Cristea, S., Gillespie, S. M., Patel, A. P., Mylvaganam, R., Luo, C. C., et al. (2018). Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat. Commun.* 9 (1), 3588. doi: 10.1038/s41467-018-06052-0

Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: recording the past and predicting the future. *Science* 358 (6359), 69–75. doi: 10.1126/science.aan6826

Kim, K. T., Lee, H. W., Lee, H. O., Song, H. J., Jeong da, E., Shin, S., et al. (2016). Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* 17, 80. doi: 10.1186/s13059-016-0945-9

Kim, T., Chen, I. R., Lin, Y., Wang, A. Y., Yang, J. Y. H., and Yang, P. (2018). Impact of similarity metrics on single-cell RNA-seq data clustering. *Brief Bioinform.* doi: 10.1093/bib/bby076

Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20 (5), 273–282. doi: 10.1038/s41576-018-0088-9

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14 (5), 483–486. doi: 10.1038/nmeth.4236

Kiselev, V. Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* 15 (5), 359–362. doi: 10.1038/nmeth.4644

Korthauer, K. D., Chu, L. F., Newton, M. A., Li, Y., Thomson, J., Stewart, R., et al. (2016). A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol.* 17 (1), 222. doi: 10.1186/s13059-016-1077-y

Kowalczyk, M. S., Tirosh, I., Heckl, D., Rao, T. N., Dixit, A., Haas, B. J., et al. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 25 (12), 1860–1872. doi: 10.1101/gr.192237.115

Krieg, C., Nowicka, M., Guglietta, S., Schindler, S., Hartmann, F. J., Weber, L. M., et al. (2018). High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat. Med.* 24 (2), 144–153. doi: 10.1038/nm.4466

Lando, D., Stevens, T. J., Basu, S., and Laue, E. D. (2018). Calculation of 3D genome structures for comparison of chromosome conformation capture experiments with microscopy: An evaluation of single-cell Hi-C protocols. *Nucleus* 9 (1), 190–201. doi: 10.1080/19491034.2018.1438799

Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., et al. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* 27 (2), 208–222. doi: 10.1101/gr.212720.116

Leng, N., Choi, J., Chu, L. F., Thomson, J. A., Kendziorski, C., and Stewart, R. (2016). OEFinder: a user interface to identify and visualize ordering effects in single-cell RNA-seq data. *Bioinformatics* 32 (9), 1408–1410. doi: 10.1093/bioinformatics/btw004

Leng, N., Chu, L. F., Barry, C., Li, Y., Choi, J., Li, X., et al. (2015). Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nat. Methods* 12 (10), 947–950. doi: 10.1038/nmeth.3549

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sanchez-Danes, A., et al. (2018). Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* 359 (6380), 1177–1181. doi: 10.1126/science.aao4174

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., et al. (2017). Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* 171 (5), 1206–1220 e1222. doi: 10.1016/j.cell.2017.10.019

Li, M., Zeng, T., Liu, R., and Chen, L. (2014). Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis. *Brief Bioinform.* 15 (2), 229–243. doi: 10.1093/bib/bbt027

Li, W. V., and Li, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* 9 (1), 997. doi: 10.1038/s41467-018-03405-7

Li, X., Chen, W., Chen, Y., Zhang, X., Gu, J., and Zhang, M. Q. (2017). Network embedding-based representation learning for single cell RNA-seq data. *Nucleic Acids Res.* 45 (19), e166. doi: 10.1093/nar/gkx750

Li, Y. H., Li, D., Samusik, N., Wang, X., Guan, L., Nolan, G. P., et al. (2017). Scalable multi-sample single-cell data analysis by partition-assisted clustering and multiple alignments of networks. *PLoS Comput. Biol.* 13 (12), e1005875. doi: 10.1371/journal.pcbi.1005875

Lin, C., Jain, S., Kim, H., and Bar-Joseph, Z. (2017). Using neural networks for reducing the dimensions of single-cell RNA-Seq data. *Nucleic Acids Res.* 45 (17), e156. doi: 10.1093/nar/gkx681

Lin, P., Troup, M., and Ho, J. W. (2017). CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* 18 (1), 59. doi: 10.1186/s13059-017-1188-0

Lin, Y., Ghazanfar, S., Strbenac, D., Wang, A., Patrick, E., Speed, T., et al. (2017). Housekeeping genes, revisited at the single-cell level. *bioRxiv*. doi: 10.1101/229815. [Epub ahead of print].

Lindeman, I., Emerton, G., Mamanova, L., Snir, O., Polanski, K., Qiao, S. W., et al. (2018). BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* 15 (8), 563–565. doi: 10.1038/s41592-018-0082-3

Liu, D., Wang, X., He, D., Sun, C., He, X., Yan, L., et al. (2018). Single-cell RNA-sequencing reveals the existence of naive and primed pluripotency in pre-implantation rhesus monkey embryos. *Genome Res.* 28 (10), 1481–1493. doi: 10.1101/gr.233437.117

Liu, L., Liu, C., Wu, L., Quintero, A., Yuan, Y., Wang, M., et al. (2018). Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *bioRxiv*. doi: 10.1101/316208. [Epub ahead of print].

Liu, T., and Wang, Z. (2018). scHiCNorm: a software package to eliminate systematic biases in single-cell Hi-C data. *Bioinformatics* 34 (6), 1046–1047. doi: 10.1093/bioinformatics/btx747

Lu, H., Li, J., Martinez Paniagua, M. A., Bandey, I. N., Amritkar, A., Singh, H., et al. (2018). TIMING 2.0: High-throughput single-cell profiling of dynamic cell-cell interactions by time-lapse imaging microscopy in nanowell grids. *Bioinformatics* 35 (4), 706–708. doi: 10.1093/bioinformatics/bty676

Maaten, L. v. d., and Hintton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. doi: 10.1007/s10846-008-9235-4

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi: 10.1016/j.cell.2015.05.002

Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L., et al. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. U. S. A.* 111 (52), E5643–E5650. doi: 10.1073/pnas.1408993111

Marinov, G. K., Williams, B. A., McCue, K., Schroth, G. P., Gertz, J., Myers, R. M., et al. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* 24 (3), 496–510. doi: 10.1101/gr.161034.113

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M. S. H., Ko, S. B. H., Gouda, N., et al. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* 33 (15), 2314–2321. doi: 10.1093/bioinformatics/btx194

McCarthy, D. J., Campbell, K. R., Lun, A. T., and Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33 (8), 1179–1186. doi: 10.1093/bioinformatics/btw777

McDavid, A., Finak, G., and Gottardo, R. (2016). The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. *Nat. Biotechnol.* 34 (6), 591–593. doi: 10.1038/nbt.3498

Menon, V. (2018). Clustering single cells: a review of approaches on high-and low-depth single-cell RNA-seq data. *Brief Funct. Genom.* 17 (4), 240–245. doi: 10.1093/bfgp/elx044

Mezger, A., Klemm, S., Mann, I., Brower, K., Mir, A., Bostick, M., et al. (2018). High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* 9 (1), 3647. doi: 10.1038/s41467-018-05887-x

Miao, Z., Deng, K., Wang, X., and Zhang, X. (2018). DEsingle for detecting three types of differential expression in single-cell RNA-seq data. *Bioinformatics* 34 (18), 3223–3224. doi: 10.1093/bioinformatics/bty332

Nagano, T., Lubling, Y., Varnai, C., Dudley, C., Leung, W., Baran, Y., et al. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 547 (7661), 61–67. doi: 10.1038/nature23001

Nelson, A. C., Mould, A. W., Bikoff, E. K., and Robertson, E. J. (2016). Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal-foetal interface during pregnancy. *Nat. Commun.* 7, 11414. doi: 10.1038/ncomms11414

Nguyen, Q. H., Lukowski, S. W., Chiu, H. S., Senabouth, A., Bruxner, T. J. C., Christ, A. N., et al. (2018). Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. *Genome Res.* 28 (7), 1053–1066. doi: 10.1101/gr.223925.117

Ntranos, V., Kamath, G. M., Zhang, J. M., Pachter, L., and Tse, D. N. (2016). Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. *Genome Biol.* 17 (1), 112. doi: 10.1186/s13059-016-0970-8

Papalexi, E., and Satija, R. (2018). Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18 (1), 35–45. doi: 10.1038/nri.2017.76

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344 (6190), 1396–1401. doi: 10.1126/science.1254257

Pellegrino, M., Sciambi, A., Treusch, S., Durruthy-Durruthy, R., Gokhale, K., Jacob, J., et al. (2018). High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res.* 28 (9), 1345–1352. doi: 10.1101/gr.232272.117

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098. doi: 10.1038/nmeth.2639

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Prot. 9, 171–181. doi: 10.1038/nprot.2014.006

Poo, M. M., Du, J. L., Ip, N. Y., Xiong, Z. Q., Xu, B., and Tan, T. (2016). China Brain Project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron* 92 (3), 591–596. doi: 10.1016/j.neuron.2016.10.050

Potter, S. S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* 14 (8), 479–492. doi: 10.1038/s41581-018-0021-7

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* 14 (3), 309–315. doi: 10.1038/nmeth.4150

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017b). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14 (10), 979–982. doi: 10.1038/nmeth.4402

Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., et al. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* 36 (5), 442–450. doi: 10.1038/nbt.4103

Reinius, B., Mold, J. E., Ramskold, D., Deng, Q., Johnsson, P., Michaelsson, J., et al. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* 48 (11), 1430–1435. doi: 10.1038/ng.3678

Rheaume, B. A., Jereen, A., Bolisetty, M., Sajid, M. S., Yang, Y., Renna, K., et al. (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. *Nat. Commun.* 9 (1), 2759. doi: 10.1038/s41467-018-05134-3

Risso, D., Purvis, L., Fletcher, R. B., Das, D., Ngai, J., Dudoit, S., et al. (2018). clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *PLoS Comput. Biol.* 14 (9), e1006378. doi: 10.1371/journal.pcbi.1006378

Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., et al. (2017). Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* 35 (6), 551–560. doi: 10.1038/nbt.3854

Rizzetto, S., Koppstein, D. N. P., Samir, J., Singh, M., Reed, J. H., Cai, C. H., et al. (2018). B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle. *Bioinformatics* 34 (16), 2846–2847. doi: 10.1093/bioinformatics/bty203

Roy, A. L., Conroy, R., Smith, J., Yao, Y., Beckel-Mitchener, A. C., Anderson, J. M., et al. (2018). Accelerating a paradigm shift: the Common Fund Single Cell Analysis Program. *Sci. Adv.* 4 (8), eaat8573. doi: 10.1126/sciadv.aat8573

Rzepiela, A. J., Ghosh, S., Breda, J., Vina-Vilaseca, A., Syed, A. P., Gruber, A. J., et al. (2018). Single-cell mRNA profiling reveals the hierarchical response of miRNA targets to miRNA induction. *Mol. Syst. Biol.* 14 (8), e8266. doi: 10.15252/msb.20188266

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2018). A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. *bioRxiv*. doi: 10.1101/276907. [Epub ahead of print].

Saliba, A. E., Westermann, A. J., Gorski, S. A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42 (14), 8845–8860. doi: 10.1093/nar/gku555

Sankar, P. L., and Parker, L. S. (2016). The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet. Med.* 19 (7), 743–750. doi: 10.1038/gim.2016.183

Sasagawa, Y., Danno, H., Takada, H., Ebisawa, M., Tanaka, K., Hayashi, T., et al. (2018). Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol.* 19 (1), 29. doi: 10.1186/s13059-018-1407-3

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., et al. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* 14 (4), R31. doi: 10.1186/gb-2013-14-4-r31

Schelker, M., Feau, S., Du, J., Ranu, N., Klipp, E., MacBeath, G., et al. (2017). Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat. Commun.* 8 (1), 2032. doi: 10.1038/s41467-017-02289-3

Sebe-Pedros, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhe, M. P., Renno, J., et al. (2018). Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell* 173 (6), 1520–1534 e1520. doi: 10.1016/j.cell.2018.05.019

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* 37 (4), 451–460. doi: 10.1038/s41587-019-0068-4

Severo, M. S., Landry, J. J. M., Lindquist, R. L., Goosmann, C., Brinkmann, V., Collier, P., et al. (2018). Unbiased classification of mosquito blood cells by single-cell genomics and high-content imaging. *Proc. Natl. Acad. Sci. U. S. A.* 115 (32), E7568–E7577. doi: 10.1073/pnas.1803062115

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510 (7505), 363–369. doi: 10.1038/nature13437

Shi, Q., Zhang, C., Peng, M., Yu, X., Zeng, T., Liu, J., et al. (2017). Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. *Bioinformatics* 33 (17), 2706–2714. doi: 10.1093/bioinformatics/btx176

Shindo, Y., Kondo, Y., and Sako, Y. (2018). Inferring a nonlinear biochemical network model from a heterogeneous single-cell time course data. *Sci. Rep.* 8 (1), 6790. doi: 10.1038/s41598-018-25064-w

Specht, A. T., and Li, J. (2017). LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics* 33 (5), 764–766. doi: 10.1093/bioinformatics/btw729

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544 (7648), 59–64. doi: 10.1038/nature21429

Strauss, M. E., Reid, J. E., and Wernisch, L. (2018). GPseudoRank: a permutation sampler for single cell orderings. *Bioinformatics* 35 (4), 611–618. doi: 10.1093/bioinformatics/bty664

Su, X., Shi, Y., Zou, X., Lu, Z. N., Xie, G., Yang, J. Y. H., et al. (2017). Single-cell RNA-seq analysis reveals dynamic trajectories during mouse liver development. *BMC Genomics* 18 (1), 946. doi: 10.1186/s12864-017-4342-x

Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14 (4), 387–387. doi: 10.1038/nmeth.4220

Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Prot.* 13 (4), 599–604. doi: 10.1038/nprot.2017.149

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6 (5), 377–382. doi: 10.1038/nmeth.1315

Tian, L., Su, S., Dong, X., Amann-Zalcenstein, D., Biben, C., Seidi, A., et al. (2018). scPipe: a flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. *PLoS Comput. Biol.* 14 (8), e1006361. doi: 10.1371/journal.pcbi.1006361

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., et al. (2016a). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352 (6282), 189–196. doi: 10.1126/science.aad0501

Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., et al. (2016b). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 539 (7628), 309–313. doi: 10.1038/nature20123

Todorov, H., and Saeys, Y. (2018). Computational approaches for high-throughput single-cell data analysis. *FEBS J.* 286 (8), 1451–1467. doi: 10.1111/febs.14613

Torre, E., Dueck, H., Shaffer, S., Gospocic, J., Gupte, R., Bonasio, R., et al. (2018). Rare cell detection by single-cell RNA sequencing as guided by single-molecule RNA FISH. *Cell Syst.* 6 (2), 171–179 e175. doi: 10.1016/j.cels.2018.01.014

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32 (4), 381–386. doi: 10.1038/nbt.2859

Tsang, J. C. H., Vong, J. S. L., Ji, L., Poon, L. C. Y., Jiang, P., Lui, K. O., et al. (2017). Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 114 (37), E7786–E7795. doi: 10.1073/pnas.1710470114

Upadhyay, A. A., Kauffman, R. C., Wolabaugh, A. N., Cho, A., Patel, N. B., Reiss, S. M., et al. (2018). BALDR: a computational pipeline for paired heavy and light chain immunoglobulin reconstruction in single-cell RNA-seq data. *Genome Med.* 10 (1), 20. doi: 10.1186/s13073-018-0528-3

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lonnerberg, P., Lou, D., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18 (1), 145–153. doi: 10.1038/nn.3881

Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat. Methods* 14 (6), 565–571. doi: 10.1038/nmeth.4292

Van den Berge, K., Perraudeau, F., Soneson, C., Love, M. I., Risso, D., Vert, J. P., et al. (2018). Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol.* 19 (1), 24. doi: 10.1186/s13059-018-1406-4

van der Wijst, M. G. P., Brugge, H., de Vries, D. H., Deelen, P., Swertz, M. A., LifeLines Cohort, S., et al. (2018). Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* 50 (4), 493–497. doi: 10.1038/s41588-018-0089-9

Vegh, P., and Haniffa, M. (2018). The impact of single-cell RNA sequencing on understanding the functional organization of the immune system. *Brief Funct. Genom.* 17 (4), 265–272. doi: 10.1093/bfgp/ely003

Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* 19 (4), 271–281. doi: 10.1038/ncb3493

Venteicher, A. S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M. G., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 355, 6332. doi: 10.1126/science.aai8478

Vergara, H. M., Bertucci, P. Y., Hantz, P., Tosches, M. A., Achim, K., Vopalensky, P., et al. (2017). Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of Platynereis dumerilii. *Proc. Natl. Acad. Sci. U. S. A.* 114 (23), 5878–5885. doi: 10.1073/pnas.1610602114

Vieth, B., Ziegenhain, C., Parekh, S., Enard, W., and Hellmann, I. (2017). powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinformatics* 33 (21), 3486–3488. doi: 10.1093/bioinformatics/btx435

Villani, A. C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356, 6335. doi: 10.1126/science.aah4573

Vu, T. N., Wills, Q. F., Kalari, K. R., Niu, N., Wang, L., Pawitan, Y., et al. (2018). Isoform-level gene expression patterns in single-cell RNA-sequencing data. *Bioinformatics* 34 (14), 2392–2400. doi: 10.1093/bioinformatics/bty100

Vu, T. N., Wills, Q. F., Kalari, K. R., Niu, N., Wang, L., Rantalainen, M., et al. (2016). Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics* 32 (14), 2128–2135. doi: 10.1093/bioinformatics/btw202

Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., and Batzoglou, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* 14 (4), 414–416. doi: 10.1038/nmeth.4207

Wang, B., Ramazzotti, D., Sano, L. D., Zhu, J., Pierson, E., and Batzoglou, S. (2018). SIMLR: a tool for large-scale genomic analyses by multi-kernel learning. *Proteomics* 18, 2. doi: 10.1002/pmic.201700232

Wang, L., Yu, X., Zhang, C., and Zeng, T., (2018). Detecting personalized determinants during drug treatment from Omics big data. *Curr. Pharm. Des.* doi: 10.2174/1381612824666181106102111

Wang, Y. J., Schug, J., Won, K. J., Liu, C., Naji, A., Avrahami, D., et al. (2016). Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 65 (10), 3028–3038. doi: 10.2337/db16-0405

Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M., and Klein, A. M. (2018). Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U. S. A.* 115 (10), E2467–E2476. doi: 10.1073/pnas.1714723115

Welch, J. D., Hartemink, A. J., and Prins, J. F. (2016). SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* 17 (1), 106. doi: 10.1186/s13059-016-0975-3

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., et al. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11 (1), 41–46. doi: 10.1038/nmeth.2694

Wu, Z., Zhang, Y., Stitzel, M. L., and Wu, H. (2018). Two-phase differential expression analysis for single cell RNA-seq. *Bioinformatics* 34 (19), 3340–3348. doi: 10.1093/bioinformatics/bty329

Xie, P., Gao, M., Wang, C., Zhang, J., Noel, P., Yang, C., et al. (2019). SuperCT: a supervised-learning framework for enhanced characterization of single-cell transcriptomic profiles. *Nucleic Acids Res.* 47 (8), e48 doi: 10.1093/nar/gkz116

Xin, Y., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., et al. (2016). RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.* 24 (4), 608–615. doi: 10.1016/j.cmet.2016.08.018

Xu, C., and Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* 31 (12), 1974–1980. doi: 10.1093/bioinformatics/btv088

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500 (7464), 593–597. doi: 10.1038/nature12364

Yang, A., Troup, M., Lin, P., and Ho, J. W. (2017). Falco: a quick and flexible single-cell RNA-seq processing framework on the cloud. *Bioinformatics* 33 (5), 767–769. doi: 10.1101/064006

Yang, Y., Huh, R., Culpepper, H. W., Lin, Y., Love, M. I., and Li, Y. (2018). SAFE-clustering: Single-cell Aggregated (From Ensemble) clustering for single-cell RNA-seq data. *Bioinformatics*. 35 (8), 1269–1277. doi: 10.1093/bioinformatics/bty793

Yip, S. H., Sham, P. C., and Wang, J. (2018). Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. *Brief Bioinform.* doi: 10.1093/bib/bby011. [Epub ahead of print].

Yip, S. H., Wang, P., Kocher, J. A., Sham, P. C., and Wang, J. (2017). Linnorm: improved statistical analysis for single cell RNA-seq expression data. *Nucleic Acids Res.* 45 (22), e179. doi: 10.1093/nar/gkx1189

Young, M. D., Mitchell, T. J., Vieira Braga, F. A., Tran, M. G. B., Stewart, B. J., Ferdinand, J. R., et al. (2018). Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361 (6402), 594–599. doi: 10.1126/science.aat1699

Yu, T. (2018). A new dynamic correlation algorithm reveals novel functional aspects in single cell and bulk RNA-seq data. *PLoS Comput. Biol.* 14 (8), e1006391. doi: 10.1371/journal.pcbi.1006391

Yu, X., Zhang, J., Sun, S., Zhou, X., Zeng, T., and Chen, L. (2017). Individual-specific edge-network analysis for disease prediction. *Nucleic Acids Res.* 45 (20), e170. doi: 10.1093/nar/gkx787

Yu, X. T., and Zeng, T. (2018). Integrative analysis of omics big data. *Methods Mol. Biol.* 1754, 109–135. doi: 10.1007/978-1-4939-7717-8_7

Zappia, L., Phipson, B., and Oshlack, A. (2017). Splatter: simulation of single-cell RNA sequencing data. *Genome Biol.* 18 (1), 174. doi: 10.1186/s13059-017-1305-0

Zeisel, A., Munoz-Manchado, A. B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347 (6226), 1138–1142. doi: 10.1126/science.aaa1934

Zeng, C., Mulas, F., Sui, Y., Guan, T., Miller, N., Tan, Y., et al. (2017). Pseudotemporal ordering of single cells reveals metabolic control of postnatal beta cell proliferation. *Cell Metab.* 25 (5), 1160–1175 e1111. doi: 10.1016/j.cmet.2017.04.014

Zeng, T., Sun, S. Y., Wang, Y., Zhu, H., and Chen, L. (2013). Network biomarkers reveal dysfunctional gene regulations during disease progression. *FEBS J.* 280 (22), 5682–5695. doi: 10.1111/febs.12536

Zeng, T., Wang, D. C., Wang, X., Xu, F., and Chen, L. (2014). Prediction of dynamical drug sensitivity and resistance by module network rewiring-analysis based on transcriptional profiling. *Drug Resist. Updat.* 17 (3), 64–76. doi: 10.1016/j.drup.2014.08.002

Zeng, T., Zhang, W., Yu, X., Liu, X., Li, M., and Chen, L. (2016). Big-data-based edge biomarkers: study on dynamical drug sensitivity and resistance in individuals. *Brief Bioinform.* 17 (4), 576–592. doi: 10.1093/bib/bbv078

Zhang, H., Lee, C. A. A., Li, Z., Garbe, J. R., Eide, C. R., Petegrosso, R., et al. (2018). A multitask clustering approach for single-cell RNA-seq analysis in recessive dystrophic epidermolysis bullosa. *PLoS Comput. Biol.* 14 (4), e1006053. doi: 10.1371/journal.pcbi.1006053

Zhang, J. M., Fan, J., Fan, H. C., Rosenfeld, D., and Tse, D. N. (2018). An interpretable framework for clustering single-cell RNA-Seq datasets. *BMC Bioinform.* 19 (1), 93. doi: 10.1186/s12859-018-2092-7

Zhang, L., and Zhang, S. (2018). Comparison of computational methods for imputing single-cell RNA-sequencing data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2018.2848633. [Epub ahead of print].

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., et al. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65 (4), 631–643 e634. doi: 10.1016/j.molcel.2017.01.023

Zong, C. C. (2017). Single-cell RNA-seq study determines the ontogeny of macrophages in glioblastomas. *Genome Biol.* 18 (1), 235. doi: 10.1186/s13059-017-1375-z