



Exploring Additional Valuable Information From Single-Cell RNA-Seq Data

Yunjin Li¹, Qiyue Xu¹, Duoqiao Wu² and Geng Chen^{1*}

¹ Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, ² Institute of Clinical Science, Zhongshan Hospital, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:

Cornelia Braicu,
Iuliu Hațieganu University of Medicine
and Pharmacy, Romania

Reviewed by:

Guibo Li,
Beijing Genomics Institute (BGI),
China

Vincent Gardeux,
École Polytechnique Fédérale
de Lausanne, Switzerland
Huanming Yang,
Beijing Genomics Institute (BGI),
China

*Correspondence:

Geng Chen
gchen@bio.ecnu.edu.cn;
chengeng66666@gmail.com

Specialty section:

This article was submitted to
Molecular Medicine,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 04 September 2020

Accepted: 26 October 2020

Published: 01 December 2020

Citation:

Li Y, Xu Q, Wu D and Chen G
(2020) Exploring Additional Valuable
Information From Single-Cell
RNA-Seq Data.
Front. Cell Dev. Biol. 8:593007.
doi: 10.3389/fcell.2020.593007

Single-cell RNA-seq (scRNA-seq) technologies are broadly applied to dissect the cellular heterogeneity and expression dynamics, providing unprecedented insights into single-cell biology. Most of the scRNA-seq studies mainly focused on the dissection of cell types/states, developmental trajectory, gene regulatory network, and alternative splicing. However, besides these routine analyses, many other valuable scRNA-seq investigations can be conducted. Here, we first review cell-to-cell communication exploration, RNA velocity inference, identification of large-scale copy number variations and single nucleotide changes, and chromatin accessibility prediction based on single-cell transcriptomics data. Next, we discuss the identification of novel genes/transcripts through transcriptome reconstruction approaches, as well as the profiling of long non-coding RNAs and circular RNAs. Additionally, we survey the integration of single-cell and bulk RNA-seq datasets for deconvoluting the cell composition of large-scale bulk samples and linking single-cell signatures to patient outcomes. These additional analyses could largely facilitate corresponding basic science and clinical applications.

Keywords: single-cell RNA-seq, cell-to-cell communication, RNA velocity, copy number variations, non-coding RNAs, cell-type deconvolution

INTRODUCTION

In recent years, single-cell RNA-seq (scRNA-seq) technologies and related bioinformatics methods have been developing and innovating rapidly, which significantly revolutionized our understanding of the expression heterogeneity and transcriptome dynamics of individual cells for diverse species including human (Quadrato et al., 2017), mouse (Brown et al., 2016), zebrafish (Wagner et al., 2018), and *Drosophila* (Karaiskos et al., 2017). The data generated by scRNA-seq can be generally grouped into read-based and unique molecular identifier (UMI)-based, depending on the full-length transcript sequencing [e.g., Smart-seq2 (Picelli et al., 2014)] or 3'/5'-end capturing [such as 10× Chromium (Zheng et al., 2017), and Drop-seq (Macosko et al., 2015)] protocols used (Chen et al., 2019). A series of preprocessing steps are required for overcoming the high noise of raw scRNA-seq data to obtain robust results from downstream analysis. Quality control (QC) of scRNA-seq data is important to remove the low-quality cells resulting from RNA degradation, break of the cell membrane, or multicells to avoid misinterpretation of downstream results, which have been reviewed recently (Luecken and Theis, 2019). Then normalization is needed to eliminate the influence of technical effects on molecular counts (e.g., sequencing depth) to make gene

expression comparable between cells. For the two main types of data generated from the full-length transcript and 3'/5'-end enrichment scRNA-seq protocols, distinct normalization methods are needed. It is recommended to take gene length into account for full-length transcript scRNA-seq data (such as the common approach of TPM normalization), while disparate methods like scran (Lun et al., 2016) are required for normalizing 3'/5'-tag scRNA-seq data (Luecken and Theis, 2019). However, normalization cannot directly address the biases of technical noises (e.g., batch effect and dropout) and biological covariates (such as cell cycle); further data processing like batch effect correction and imputation may be needed to mitigate such effects according to the data properties and research goals.

After data preprocessing, a range of common analyses can be conducted, like cell type/state identification and annotation, trajectory inference, alternative splicing detection, gene regulatory network (GRN) reconstruction, which has been reviewed by us and other colleagues (Chen et al., 2019; Luecken and Theis, 2019). Because scRNA-seq data usually involve many cells and thousands of genes, feature selection and dimensionality reduction methods are needed to reduce the dimensionality of high-dimensional datasets to lighten the computational burden of downstream analysis (Andrews and Hemberg, 2018). Generally, 500–5,000 highly variable genes are often used depending on the data complexity in feature selection approaches (Yip et al., 2019). Linear [e.g., principal component analysis (PCA)] or non-linear [such as t-distributed stochastic neighbor embedding [t-SNE (van der Maaten and Hinton, 2008)]] and uniform approximation and projection (UMAP) (Diaz-Papkovich et al., 2019) dimensionality reduction methods can be used to further reduce the data dimension and visualize the data in two or three dimensions (Moon et al., 2018).

Based on the data with reduced dimensions, the cell clusters are typically identified in single-cell analysis. Methods for clustering (such as *k*-means) or community detection (e.g., *K*-nearest neighbor graph) are often applied to determine the clusters according to the expression similarity of genes (Duò et al., 2018; Kiselev et al., 2019). Once the clusters of single cells are determined, marker genes can be identified through differential expression (DE) analysis to annotate the clusters with meaningful biological insight. Moreover, for the scRNA-seq data generated from full-length transcript sequencing protocols, the alternative splicing changes between distinct cell clusters can be further investigated as we summarized previously (Chen et al., 2019). On the other hand, for the single-cell datasets involving developmental or differentiation process, trajectory inference methods can be utilized to infer the order of cells along developmental trajectories. Saelens et al. (2019) benchmarked dozens of trajectory inference tools and revealed that these methods are complementary with variable performances depending on the dataset dimensions and trajectory topology. Additionally, cellular differentiation and cell state transition processes are controlled by the underlying GRNs. An increasing number of approaches have been developed to infer the GRNs from scRNA-seq data generally based on the assumption that the genes with similar expression profiles could be regulated by a common transcription factor [such as SCENIC

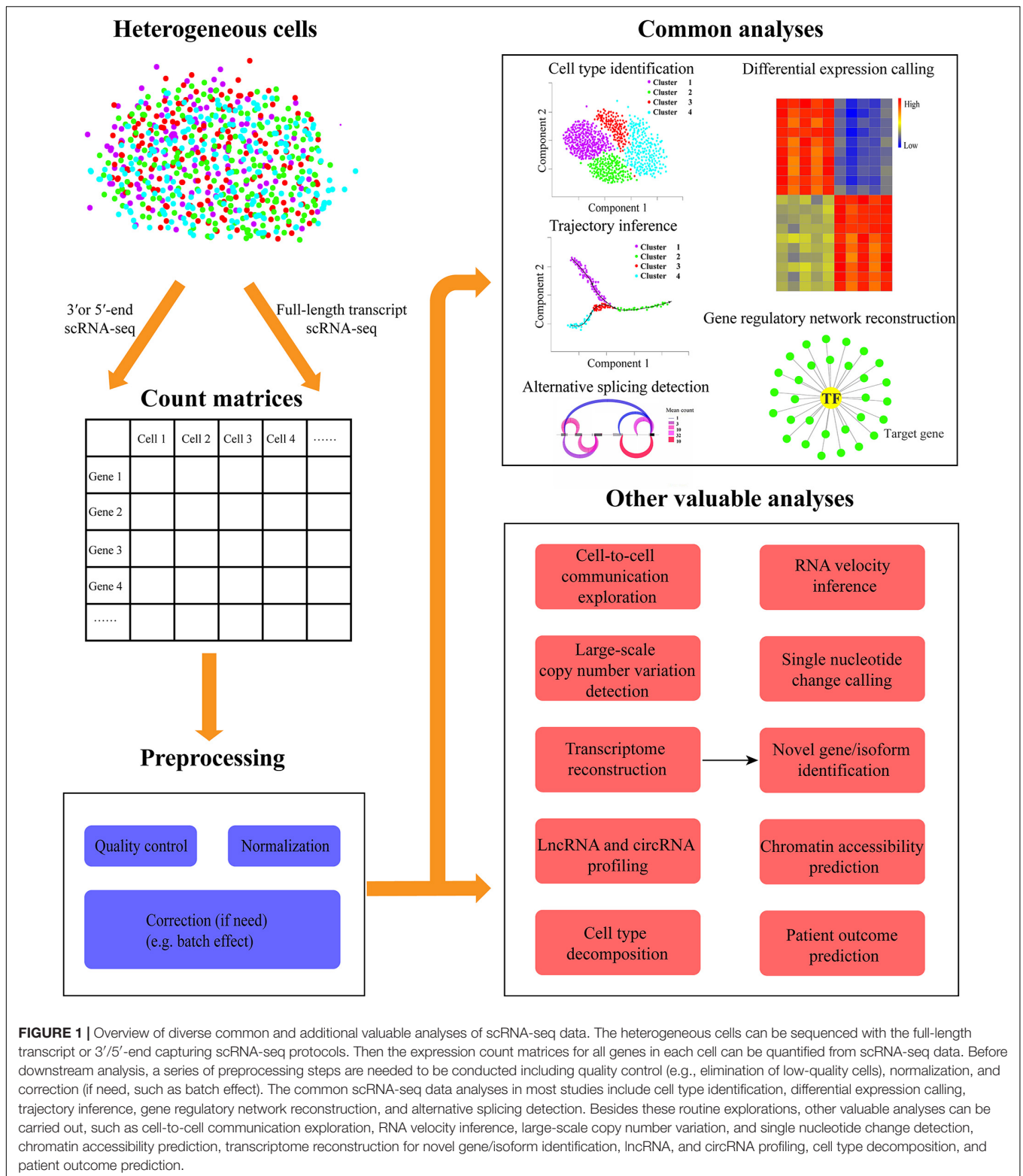
(Aibar et al., 2017)], but more efforts are needed to improve the accuracy of these analytical approaches (Chen and Mar, 2018; Fiers et al., 2018; Pratapa et al., 2020).

However, besides those common analyses, many other valuable explorations can be conducted to gain additional insights into scRNA-seq data (Figure 1). In this review, we first describe the progress and related methods for cell-cell communication network inference, RNA velocity analysis, interrogation of chromosomal-scale copy number variations (CNVs) and single nucleotide variations, as well as novel gene/isoform identification. Then we summarize the integration of single-cell and bulk RNA-seq data to cost-effectively analyze a large sample size. In particular, we discuss their implications and potential challenges as well as future directions.

CELL-TO-CELL COMMUNICATION NETWORK INFERENCE

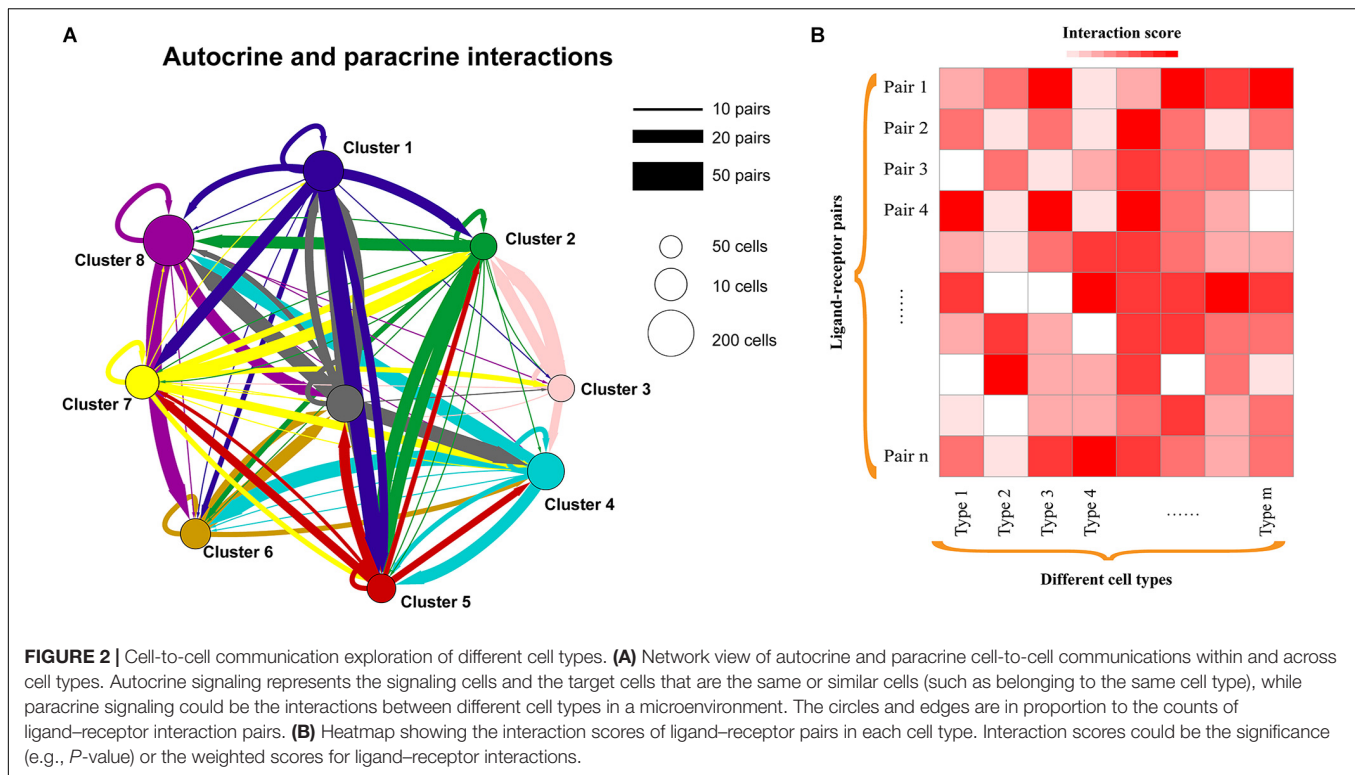
Cells often do not function independently but can communicate with each other and change their behaviors by transmitting and receiving signals within their environment. In multicellular organisms, cell signaling is critical for joining different cell types together to form tissues (e.g., brain, lung, muscle, and liver). Specifically, autocrine (interact with the same or similar cells) and paracrine (communicate with nearby cells) signaling networks within and across cell types play fundamental roles for cells working together to coordinate diverse organismal processes. Moreover, an abundance of cell fate decisions are made to react to extracellular signals from the interactions between secreted ligands and cell-surface receptors in the local environment (Watabe and Miyazono, 2009). Especially for cancers, the tumor microenvironment is typically composed of various cell types (including malignant, immune, and stromal cells). Understanding the cell-to-cell communication/interaction network among distinct cell populations can facilitate the elucidation of underlying mechanisms for tumorigenesis, tumor progression, metastasis, therapy resistance, and immune infiltration (Hanahan and Weinberg, 2011). Defects in cell-to-cell interaction have been demonstrated to be associated with different cancers (Haass et al., 2004), autoimmune (Gorelik and Flavell, 2000), and metabolic diseases (Hotamisligil, 2006).

ScRNA-seq enables expression quantification of transcripts encoding ligands and their cognate receptors in each cell, which provides unprecedented opportunities for decoding the diversity, complexity, and dynamics of intercellular communication networks (Figure 2). An increasing number of studies investigated the cell-to-cell communications between distinct cell populations and uncovered meaningful biological insights. For example, interlineage communications mediated by ligand-receptor complexes among single cells can regulate liver bud development (Camp et al., 2017), and functionally important ligand-receptor interactions associated with cancer metastasis were recently identified in head and neck squamous cell carcinoma (Puram et al., 2017). We also detected a set of intercellular communications between macrophages and cancer stem-like cells (CSCs) in glioma that the expression levels



of involved ligands and receptors are significantly correlated with the survival of patients (Yuan et al., 2019). Moreover, lung basophils were found to widely communicate with both immune and non-immune compartments (Cohen et al., 2018),

and cell-cell interactions were useful in identifying the cell types of human placenta (Pavlicev et al., 2017). Interaction network analysis between distinct cell types within the melanoma microenvironment highlighted that tumor cell composition is



critical for diagnostic and therapeutic strategies (Tirosh et al., 2016a). Additionally, extensive intercellular communication networks were observed between diverse mouse heart cell types, which contributed to the transcriptional program of sexual dimorphism (Skelly et al., 2018).

To identify the potential interactions within or between cell subpopulations from scRNA-seq data, an increasing number of computational methods have been developed based on the expression abundance of ligand and receptor pairs (Table 1). For instance, Kumar et al. (2018) proposed a computational approach to characterize cell-cell communications across the cell types in a microenvironment (such as tumor ecosystem) by scoring the ligand-receptor interactions between two cell types as the product of average expression of ligands and receptors in corresponding cell types. PyMINer integrates ligand and receptor information, protein-protein interactions as well as pathway analyses to build the autocrine-paracrine signaling networks (Tyler et al., 2019). scTensor defines the cell-cell interactions as directed hypergraphs (nodes are cell types, and edges represent ligand-receptor pairs) and can infer many-to-many interactions with tensor decomposition (Tsuyuzaki et al., 2019). iTALK identifies the intercellular crosstalk signals based on curated ligand-receptor pairs and can visualize the results in different plot formats like Circos, network, and errorbar (Wang Y. et al., 2019). Moreover, CellPhoneDB provides a repository of curated receptors, ligands, and their interactions, and can allow users to search particular ligand/receptor or predict enriched cellular interactions with inquired scRNA-seq data efficiently (Efremova et al., 2020). CellChat quantitatively infers intercellular communication networks using mass action models,

which also enables the visualization of cellular interactions (Jin et al., 2020a). Additionally, SingleCellSignalR allows the assessment of the confidence of predicted ligand-receptor (Cabello-Aguilar et al., 2020), while NicheNet can enable a functional understanding of cell-cell communications by providing the information on how ligand-receptor interactions influence the target gene expression (Browaeys et al., 2020). However, the study for systematic performance evaluation of these methods is currently lacking. Moreover, the available approaches for inferring cell-cell interactions are generally based on the known and/or curated ligand-receptor pairs; the interactions mediated by unknown ligand-receptor pairs will be missed. When interpreting the resulting cell-cell communications between cell types, especially the number of interactions, it would be better to consider the missing interactions. Therefore, dissecting the cellular communications in health and disease states will largely benefit the elucidation of the underlying molecular mechanisms.

RECONSTRUCTION OF SPATIAL CELLULAR COMMUNICATIONS AND GENE EXPRESSION

Additionally, the spatial organization of cells is closely associated with diverse cell functions and behaviors including cell-cell interactions, but such information is usually missing from scRNA-seq data as the cells are needed to be dissociated before sequencing. Interestingly, novoSpaRc was recently developed to enable *de novo* spatial reconstruction of gene expression using

TABLE 1 | Computational approaches for additional analyses of scRNA-seq data.

Categories	Suitable scRNA-seq protocols	Tools	URL	References
Cell to cell communication	Full-length transcript or 3'/5'-tag sequencing	PyMINER	https://bitbucket.org/scotttyler892/pyminer_release	Tyler et al., 2019
		scTensor	https://rdrr.io/bioc/scTensor/	Tsuyuzaki et al., 2019
		iTALK	https://github.com/Coolgenome/iTALK	Wang Y. et al., 2019
RNA velocity	Full-length transcript or 3'/5'-tag sequencing	CellPhoneDB	https://github.com/Teichlab/cellphonedb	Efremova et al., 2020
		velocyto	https://github.com/velocyto-team/velocyto.R	La Manno et al., 2018
Copy number variation	Full-length transcript or 3'/5'-tag sequencing	scVelo	https://github.com/theislab/scvelo	Bergen et al., 2019
		inferCNV	https://github.com/broadinstitute/inferCNV	Patel et al., 2014
Chromatin accessibility	Full-length transcript sequencing or 3'/5'-tag sequencing	HoneyBADGER	https://github.com/JEFworks/HoneyBADGER	Fan et al., 2018
		BIRD	https://github.com/WeiQiangZhou/BIRD	Zhou et al., 2017
Single nucleotide variants	Full-length transcript sequencing	SAMtools	http://samtools.sourceforge.net/	Li, 2011
		Strelka2	https://github.com/Illumina/strelka	Kim et al., 2018
		FreeBayes	https://github.com/ekg/freebayes	Garrison and Marth, 2012
RNA editing	Full-length transcript sequencing	GIREMI	https://github.com/zhqingit/giremi	Zhang and Xiao, 2015
		REDIttools	https://github.com/BioinfoUNIBA/REDIttools	Picardi and Pesole, 2013
Transcriptome reconstruction	Full-length transcript sequencing	TransComb (genome-guided)	https://zenodo.org/record/61994#.XIEfaOgzaUk	Liu J. T. et al., 2016
		StringTie (genome-guided and <i>de novo</i>)	https://ccb.jhu.edu/software/stringtie/	Pertea et al., 2015
		Cufflinks (genome-guided)	http://cole-trapnell-lab.github.io/cufflinks/	Trapnell et al., 2010
		Trinity (<i>de novo</i>)	https://github.com/trinityrnaseq/trinityrnaseq/wiki	Grabherr et al., 2011
		Trans-ABYSS (<i>de novo</i>)	https://github.com/bcgsc/transabyss	Robertson et al., 2010
		rnaSPAdes (<i>de novo</i>)	http://cab.spbu.ru/software/rnaspades/	Bushmanova et al., 2019
Coding potential assessment	Full-length transcript sequencing	CPAT	http://rna-cpat.sourceforge.net/	Wang et al., 2013
		LncRNA-ID	https://github.com/zhangy72/LncRNA-ID	Achawanantakun et al., 2015
		LGC	http://bigd.big.ac.cn/biocode/tools/BT000004	Wang G. Y. et al., 2019
Circular RNA identification	Total RNA (poly (A+) and poly (A-) RNAs) sequencing	find_circ2	https://github.com/rajewsky-lab/find_circ2	Memczak et al., 2013
		CircExplorer2	https://circexplorer2.readthedocs.io/en/latest/	Zhang et al., 2016
Cell composition deconvolution	Full-length transcript or 3'/5'-tag sequencing	CIRI2	https://sourceforge.net/projects/ciri/	Gao et al., 2018
		CMP	https://cran.r-project.org/web/packages/scBio/index.html	Frishberg et al., 2019
		MuSiC	https://github.com/xuranw/MuSiC	Wang X. R. et al., 2019
Survival analysis	Full-length transcript or 3'/5'-tag sequencing	DWLS	https://bitbucket.org/yuanlab/dwls	Tsoucas et al., 2019
		CIBERSORTx	https://cibersortx.stanford.edu/	Newman et al., 2019
		Cox regression	https://github.com/therneau/survival	Li, 2003

scRNA-seq data alone (Nitzan et al., 2019). Specifically, CSOmap cannot only predict the cellular interactions but also can infer the cell spatial organizations *de novo* from single-cell transcriptomic data (Ren et al., 2020). Furthermore, the sequencing-based or image-based spatial technologies that can preserve the spatial coordinates of cells have achieved great progress (Mayr et al.,

2019). Integrative analysis of the spatial and scRNA-seq data may enable us to gain novel insights into cell–cell communications by constructing the spatial expression patterns of signaling ligands and receptors using transfer learning or deep learning approaches (Efremova and Teichmann, 2020). For instance, SpaOTsc can allow the inference of spatial gene expression patterns and spatial

cell–cell communications by incorporating scRNA-seq and spatial data (Cang and Nie, 2020). With the innovation of scRNA-seq and spatial transcriptomics as well as the computational algorithms, the accuracy of intercellular communication network inference will be improved as well. Specifically, such analysis may shed light on the signaling mechanisms of cellular behaviors and responses under various conditions like tumor progression, development, or differentiation.

IDENTIFICATION OF LARGE-SCALE COPY NUMBER VARIATIONS

Besides cellular communication detection, scRNA-seq data can be used to identify different types of genomic variations. Intratumoral heterogeneity is a ubiquitous feature for various cancer types, which contributes to tumor progression and therapy failure (Kreso and Dick, 2014). One of the well-studied sources of intratumoral heterogeneity is genetic variation, such as single nucleotide variations and CNVs that are the gains or losses of genomic sequences larger than one kilobase in size (Vogelstein et al., 2013). CNVs play an essential role in generating both physiological and pathological phenotypes through altering corresponding gene transcription or disrupting neighboring or distant non-coding regulatory regions; some of them could have pathogenic roles in common and rare cancers (Shlien and Malkin, 2009).

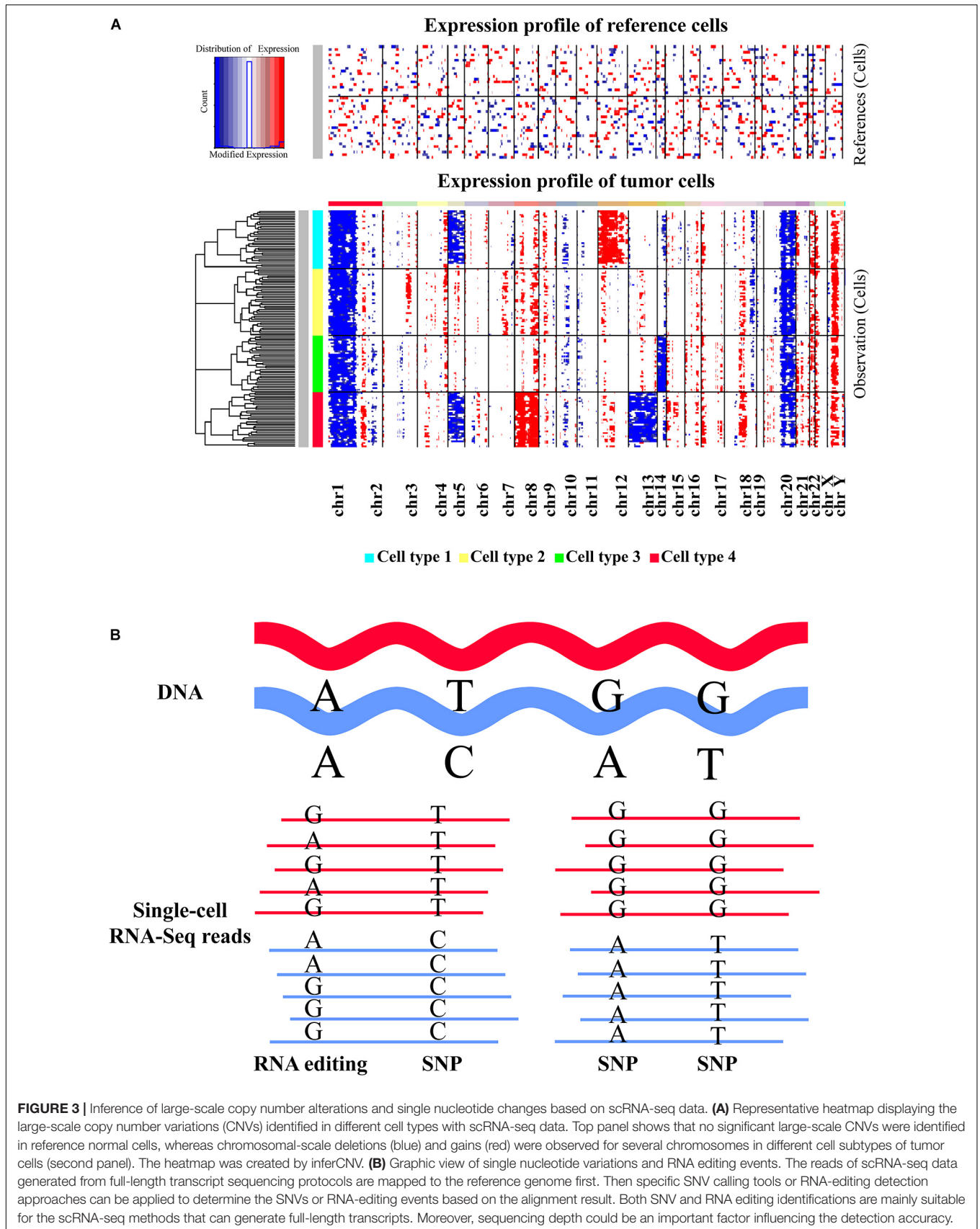
As large-scale CNVs may cause the gain or loss of many genes, they can result in the upregulation or downregulation of the genes in the affected regions. It has been shown that scRNA-seq data can provide informative large-scale CNV evidence for corresponding cells (**Figure 3A**). For instance, Patel et al. (2014) revealed coherent chromosomal-scale CNV pattern in glioblastoma by averaging relative expression levels of genes over large chromosomal regions and comparing with a set of reference normal cells using their method of inferCNV. With a similar approach, somatic large-scale CNVs were examined in metastatic melanoma (Tirosh et al., 2016a), oligodendroglioma (Tirosh et al., 2016b), as well as head and neck cancer (Puram et al., 2017) at single-cell resolution, which allowed researchers to effectively distinguish malignant cells from non-malignant ones. Recently, another computational method that integrated the hidden Markov model with a Bayesian approach, called HoneyBADGER, has also been developed for identifying the CNVs and loss of heterozygosity in single cells based on the allele and expression information inferred from scRNA-seq data (Fan et al., 2018) (**Table 1**). Since genomic instability is a hallmark of diverse cancers (Negrini et al., 2010; Ferguson et al., 2015), detecting the somatic large-scale CNVs in single cells could discriminate tumor cells from normal ones and gain insights into their roles in tumorigenesis. However, attention should be paid to the sparsity and noise of scRNA-seq data because currently available scRNA-seq approaches are generally with high-dropout property, which may result in false positives and influence the CNV detection. Collectively, scRNA-seq provides an alternative and cost-effective way for exploring large-scale CNVs in individual cells. It is valuable for unraveling the

evolutionary complexity of tumors and understanding cancer development and progression.

ANALYSIS OF SINGLE NUCLEOTIDE VARIANTS AND RNA EDITING

In addition to CNV detection, single nucleotide variants (SNVs) and RNA editing events could also be inferred from single-cell transcriptomic data. SNVs are the most prevalent type of genetic variation and are closely associated with diverse normal and disease phenotypes. The influences of SNVs could manifest on gene expression by *cis* and/or *trans* effects (Bryois et al., 2014), and a multitude of SNVs have been linked to tumor evolution (Navin et al., 2011). Importantly, SNVs in progenitors could be inherited by all the daughter cells during DNA replication, thus systematic SNV calling in single cells is one promising strategy for delineating cellular heterogeneity and phylogenetic relationships, especially for cancer evolution (Navin et al., 2010; Abbosh et al., 2017; Ju et al., 2017; Martincorena et al., 2017). Although single-cell exome sequencing or whole-genome sequencing technologies can be used to interrogate SNVs, such approaches could introduce substantial error rates due to inherent technical limitations (Xu et al., 2012; Zafar et al., 2016), and they are highly expensive for sequencing a large number of cells. By contrast, scRNA-seq is more affordable, and the SNVs detected from single-cell transcriptomic data could be interesting since they are expressed, and their functions are easier to elucidate. A range of studies have revealed intriguing findings by exploring SNVs from scRNA-seq data using the tools originally developed for bulk sequencing data (Tirosh et al., 2016b; Enge et al., 2017; Fan et al., 2018; Poirion et al., 2018; Ding et al., 2019). For example, Enge et al. (2017) gained insights into aging-related genetic and transcriptional processes of the human pancreas by analyzing the somatic mutation patterns with single-cell transcriptomic data. A linear modeling framework, SSrGE, was recently proposed to detect the effective and expressed SNVs that are associated with gene expression from scRNA-seq data, which could facilitate the subpopulation identification and genotype–phenotype relationship determination (Poirion et al., 2018). Moreover, Ding et al. (2019) developed a method for trajectory inference based on the SNPs inferred from scRNA-seq data.

Currently, few tools were specially designed for SNV calling based on single-cell transcriptomic data. However, Liu et al. (2019) systematically evaluated the performance of traditional variant callers on scRNA-seq datasets and recommended SAMtools (Li, 2011), Strelka2 (Kim et al., 2018), and FreeBayes (Garrison and Marth, 2012) to call SNVs for the data with low supporting reads, with sufficient read depths, and with high variant allele frequencies, respectively (**Table 1**). SAMtools calls the SNVs directly based on the sequencing data with a statistic model, while Strelka2 employs a mixture model to alleviate the effects of context-specific variation, and FreeBayes uses a Bayesian statistical framework to model multiallelic loci. With these tools, the SNVs in each cell can be predicted by treating each cell as a sample like bulk data. Notably, low read depths that



resulted from the biologically low expressions and/or technical bias (e.g., dropout events) could reduce the sensitivity of SNV detection. Therefore, the innovation of scRNA-seq strategies to minimize the dropout events will greatly improve the accuracy of SNV inference (Liu et al., 2019). Moreover, novel SNV calling methods that are specifically designed for scRNA-seq are also crucial for correcting the technical bias and increase the sensitivity and specificity of variant calling. Overall, detecting SNVs from single-cell transcriptomic data could provide another layer of cellular heterogeneity among single cells besides gene expression (**Figure 3B**), which could be useful for lineage tracing and subpopulation identification as well as genotype–phenotype linkage inference (Poirion et al., 2018; Tang, 2020).

Unlike genomic SNVs, RNA editing is a posttranscriptional process that made nucleotide changes on RNA sequences, and adenosine-to-inosine (A-to-I) editing is the most common type in general (Nishikura, 2010) (**Figure 3B**). RNA editing has been considered as a crucial mechanism for increasing the molecular diversity and regulating the function of proteins (Maas et al., 2006; Park et al., 2012). The known functional impacts of RNA editing mainly include amino acid sequence changes, alternative splicing alteration, RNA stability influence, and alternations on miRNA sequence or miRNA targeting sequence (Nishikura, 2016). Furthermore, aberrant RNA editing events could be correlated with the etiology or progression of various diseases, such as amyotrophic lateral sclerosis, astrocytoma, hepatocellular carcinoma, and metastatic melanoma (Slotkin and Nishikura, 2013; Kung et al., 2018; Kanata et al., 2019). Although sequencing the genome and transcriptome from the same sample/cell can theoretically enable more accurate RNA editing detection, such data are relatively uncommon and costly. Several computational tools are available for robustly identifying RNA editing sites using bulk RNA-Seq data alone, such as GIREMI (Zhang and Xiao, 2015), the pipeline proposed by Ramaswami et al. (2013), and REDIttools (Picardi and Pesole, 2013) (**Table 1**). However, the approaches specifically developed for scRNA-seq data are currently lacking, and a few studies investigated the RNA editome in individual cells. Recently, Ding et al. (2019) suggested that an abundance of SNVs identified from scRNA-seq data by their method are likely to be RNA-editing events. Since aberrant RNA editing events could be correlated with the etiology or progression of many diseases including cancers (Slotkin and Nishikura, 2013; Kung et al., 2018; Kanata et al., 2019), exploring the RNA editome in single cells can facilitate a better understanding of their functional implications to cellular heterogeneity and clinical utility in diseases. Considering that RNA editing detection depends closely on the sequencing depth, applying the tools originally designed for bulk data to single-cell data should be careful due to the inherent technical noise and low sequencing depth of current scRNA-seq protocols. There is an urgent need to develop robust methods for identifying RNA editing events with single-cell data. Consequently, exploring the RNA editome in single cells will be more feasible with the improvement of single-cell sequencing and specialized algorithms, which will benefit the elucidation of the functional implications of RNA editing to cellular variations and disease development.

EXPLORING RNA VELOCITY

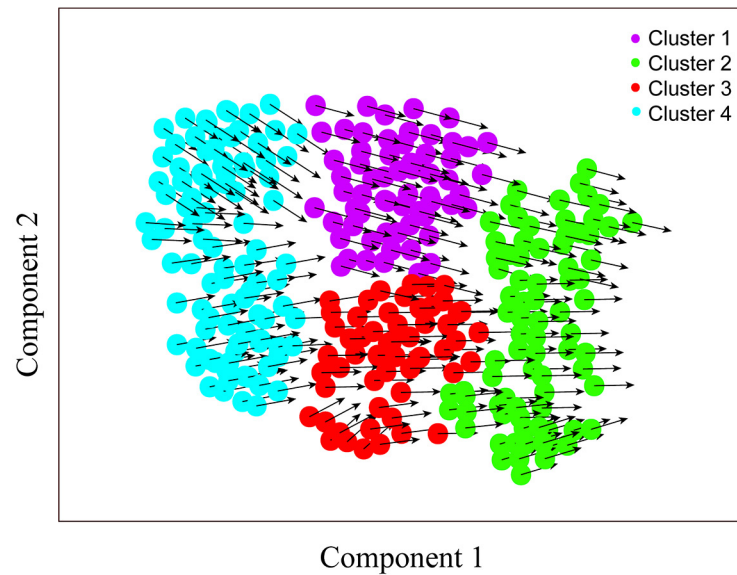
scRNA-seq data have also been used to predict the future transcriptional state of single cells (termed RNA velocity) by deducing their directed dynamic transcriptome changes (**Figure 4A**). RNA regulation involves multiple stages including transcription, RNA maturation, and RNA degradation; thus, the abundance of RNAs is a strong indicator of cell state. Previous bulk RNA-seq study has shown that gene splicing and degradation can be effectively estimated based on the relative abundance of unspliced and spliced RNAs (Zeisel et al., 2011; Gaidatzis et al., 2015). Thus, similar signals could be also decoded from individual cells with single-cell transcriptomic data (Svensson and Pachter, 2018). La Manno et al. (2018) proposed a model named *velocyto* (**Table 1**) to estimate the rate of change in mRNA abundance (RNA velocity) to predict the future transcriptional state of individual cells by distinguishing between spliced and unspliced mRNAs with scRNA-seq data. This RNA velocity inference method has been applied to an increasing number of researches. For instance, RNA velocity analysis revealed dynamic transcriptional changes of immune cells in hepatocellular carcinoma (Zhang et al., 2019) and could also allow effective identification of the major directions of cell progression for murine neural crest cells (Soldatov et al., 2019). Moreover, Kanton et al. (2019) successfully uncovered the differentiation of neural progenitor cells in human development with RNA velocity exploration, but *velocyto* could not efficiently process large datasets and even may run out of memory (e.g., cell number >40,000). More recently, Bergen et al. developed a likelihood-based dynamical model, *scVelo* (can handle >300,000 cells), to infer the RNA velocity of cells by solving the full transcriptional dynamics (Bergen et al., 2019), which is 10 times faster and less memory consuming than that of *velocyto* (La Manno et al., 2018).

RNA velocity inference could predict the direction of cell transition within and between cell clusters/states. By contrast, pseudotime/trajectory analysis aims to identify the paths between cell clusters/subtypes, which does not automatically infer a direction like RNA velocity prediction. However, RNA velocity analysis can benefit trajectory inference or pseudotemporal ordering that aims to deduce the order of cells along developmental paths by overlaying the directionality of velocity to trajectories to better predict cell fate decisions (La Manno et al., 2018). Therefore, integrative analysis of single-cell RNA velocity and trajectory/pseudotime could provide deeper insights into various dynamic cellular processes in development and evolution, such as lineage decisions and gene regulation.

INFERRING CHROMATIN ACCESSIBILITY

Besides transcriptome profiling, scRNA-seq could also provide the potential for decoding the chromatin accessibility of transcribed regions in single cells (**Figure 4B**). Chromatin accessibility is essential for establishing and maintaining cellular identity by governing cell- or context-specific gene expression

A RNA velocity inference based on scRNA-seq data



B Chromatin accessibility prediction from scRNA-seq data

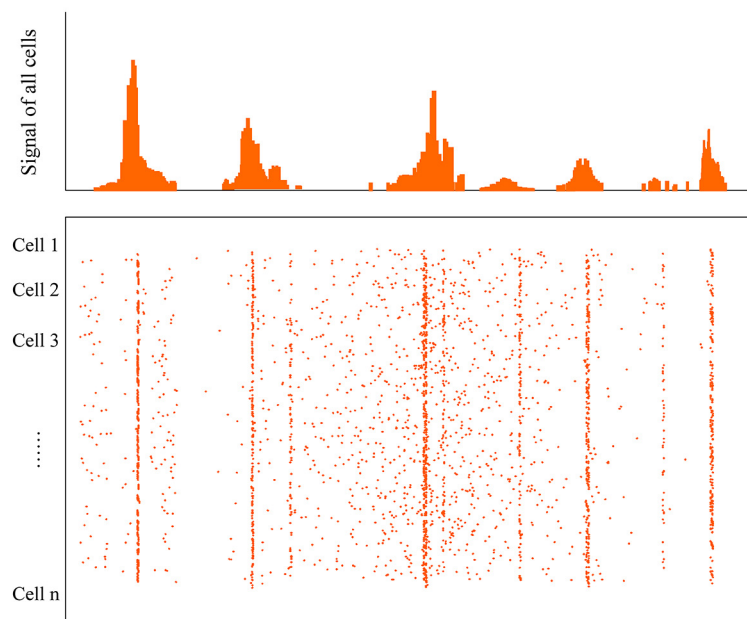


FIGURE 4 | RNA velocity and chromatin accessibility analyses. **(A)** RNA velocity inference of single cells to predict their future transcriptional states. The velocity of gene expression could be represented as the mRNA abundance over time, which enables the prediction of future transcriptional state of cells (the arrows denote the directionality). **(B)** Graphic view of chromatin accessibility prediction with scRNA-seq data. Transcriptome and regulome could have bidirectional interplay because of the feedback, thus scRNA-seq has the potential to predict the chromatin accessibility of transcribed regions using the corresponding computational approach. However, it is worth noting that the chromatin accessibility of non-transcribed regions cannot be predicted with scRNA-seq.

(Pennacchio et al., 2013; Klemm et al., 2019). The landscape of chromatin accessibility broadly reflects the regulatory capacity and is dynamically changing in response to developmental cues and environmental stimulation (Klemm et al., 2019).

Some single-cell technologies are emerging to measure the chromatin accessibility of individual cells including single-cell ATAC-seq (Cusanovich et al., 2015), single-cell DNase-seq (Jin et al., 2015), and single-cell THSseq (Lake et al., 2018).

Moreover, Yu et al. (2020) recently proposed a software, scATAC-pro, for quality estimation and visualization of single-cell chromatin accessibility sequencing data generated by different experimental protocols.

Determining the accessible genome is crucial for understanding the regulatory program of gene expression control. Many studies have demonstrated that the transcriptional activities of genes can be predicted based on the activities of associated regulatory elements (Natarajan et al., 2012; Kumar et al., 2013), but few researches investigated to what extent activities of regulatory elements can be inferred from the RNA-seq data. Gene transcription needs the chromatin to be open and accessible; thus, bidirectional interplay exists between transcriptome and regulome due to the feedback (Neph et al., 2012; Voss and Hager, 2014). Previously, Zhou et al. (2017) demonstrated that their method of BIRD (**Table 1**) can effectively predict the activities of genome-wide regulatory elements measured by DNase I hypersensitivity based on bulk gene expression profiles. Since scRNA-seq technologies enable capturing the gene transcriptional signals in each cell, it may be also possible to predict the regulome of cells based on single-cell transcriptomic data. Recently, Zhou et al. (2019) further suggested that the chromatin accessibility of the genome could be inferred from the scRNA-seq data of a small number of cells. But currently available methods for inferring chromatin accessibility from single-cell transcriptomic data are very few. Both experimental chromatin accessibility profiling technologies and computational methods that predict chromatin accessibility from scRNA-seq data will continue to improve. It remains an open question as to which method will be more accurate. To answer that question, a systematic and independent benchmark study in the future will be required.

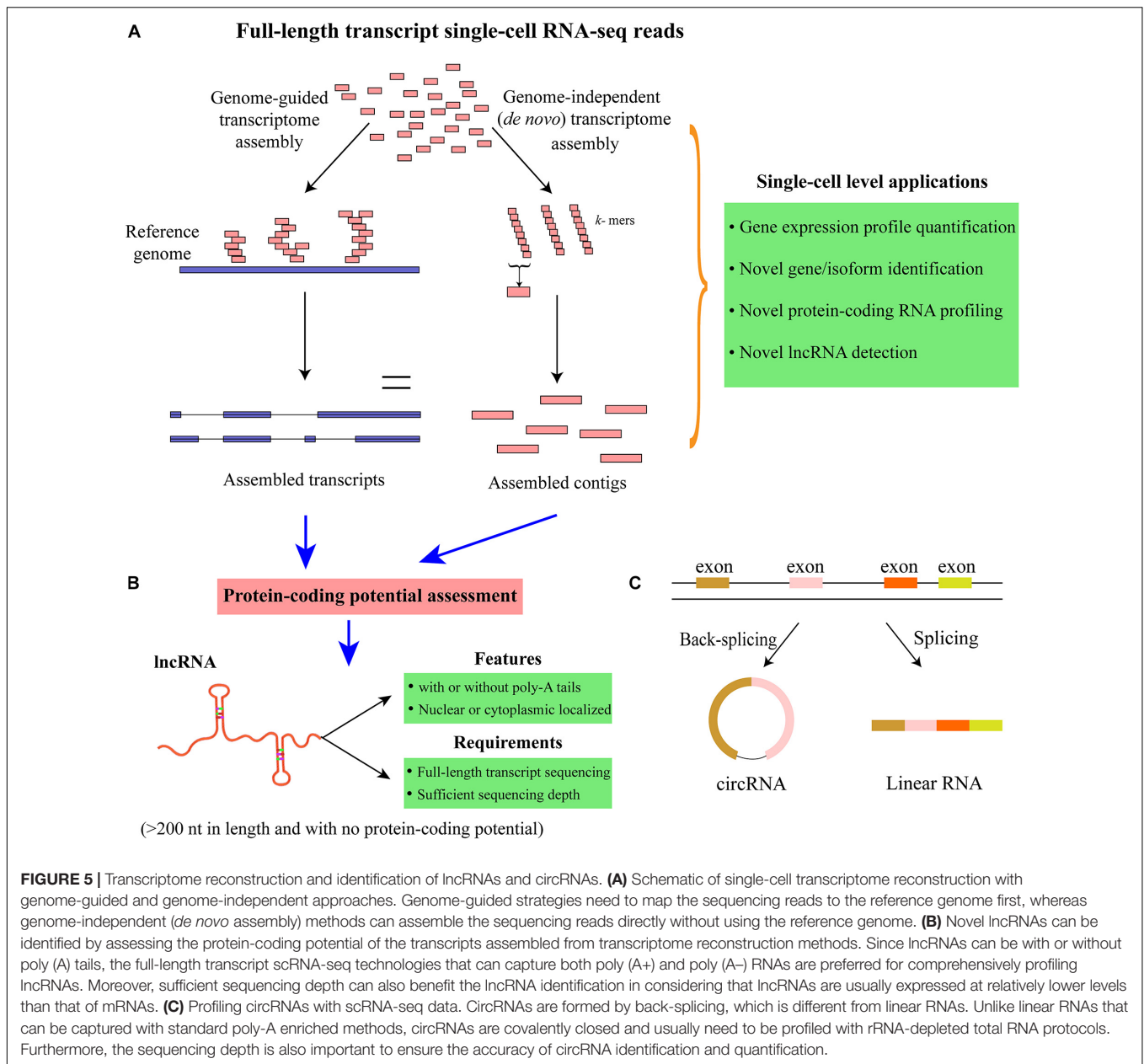
Specifically, the data from single-cell RNA-seq and chromatin profiling technologies can be combined to delineate cellular heterogeneity and elucidate transcriptional regulatory mechanisms. For instance, the computational tool of SOMatic enables the integrative analysis of scATAC-seq and scRNA-seq data for gene regulatory network reconstruction (Jansen et al., 2019). ScAI can deconvolute the cellular heterogeneity based on single-cell transcriptomic and epigenomic profiles (Jin et al., 2020b). Additionally, MAESTRO supports cell clustering and automatic cell-type annotation as well as transcriptional regulator inference for both scRNA-seq and scATAC-seq datasets (Wang et al., 2020). These analyses will help us better elucidate the underlying mechanisms of gene regulation and cellular gene expression heterogeneity.

TRANSCRIPTOME RECONSTRUCTION FOR NOVEL GENE/ISOFORM IDENTIFICATION

For full-length transcript scRNA-seq data, transcriptome reconstruction at the single-cell level is promising to identify cell-type-specific genes/isoforms. Currently, the annotated genes and isoforms for many species including humans are still far

from complete, and a multitude of novel protein-coding and non-coding genes/isoforms remain to be uncovered (Chen et al., 2013). One major reason accounting for this is that gene expression is often spatial and temporal specific; thus, those unannotated genes/isoforms could be only expressed in specific conditions and/or cell types/states. Since gene expression is usually heterogeneous at the single-cell level, different cell subpopulations may express unique and unannotated genes and/or isoforms that could not be identified with bulk RNA-seq data. Thus, scRNA-seq provides great potential for identifying and annotating the novel genes and isoforms.

Transcriptome reconstruction is the most popular strategy for detecting all the expressed genes and isoforms in a particular sample (Garber et al., 2011; Chen et al., 2017). The approaches for transcriptome reconstruction can be mainly grouped into the following two categories: genome-guided and *de novo* (genome independent) transcriptome assembly (Garber et al., 2011) (**Figure 5A** and **Table 1**). Generally, genome-guided strategies [such as TransComb (Liu J. T. et al., 2016), StringTie (Pertea et al., 2015), and Cufflinks (Trapnell et al., 2010)] assemble the overlapping reads aligned to the reference genome into transcripts, which is suitable for the organisms with the available qualified reference genome. By contrast, *de novo* transcriptome assembly methods [e.g., Trinity (Grabherr et al., 2011), *Trans-ABYSS* (Robertson et al., 2010), and rnaSPAdes (Bushmanova et al., 2019)] often utilize de Bruijn graph to directly assemble the reads into transcripts without the need of reference genome. When a qualified reference genome is available, genome-guided approaches are the choice due to their higher sensitivity than *de novo* assembly methods. However, for cancer cells, large-scale rearrangement events may exist in the genome and/or transcriptome; a combination use of these two different strategies may generate a more comprehensive set of transcripts (Garber et al., 2011). After transcriptome reconstruction, the coding potential of those assembled transcripts can be assessed to group them into protein-coding or non-coding RNAs. Although the available transcriptome reconstruction approaches are mainly designed for bulk RNA-seq data, some studies have applied them to scRNA-seq data and successfully identified many novel genes/transcripts (Yan et al., 2013; Fan et al., 2015; Liu S. J. et al., 2016; Wu et al., 2019). For example, Yan et al. (2013) integrated genome-independent and genome-guided assembly methods to predict the new transcripts and detected a set of novel long non-coding RNAs (lncRNAs) that are functionally important in human embryos. Notably, transcriptome assembly is mainly applicable to the scRNA-seq approaches that can sequence the full-length of transcripts [e.g., Smart-Seq2 (Picelli et al., 2014), SUPeR-seq (Fan et al., 2015), and RamDA-seq (Hayashi et al., 2018)] rather than the protocols that only capture the 3'/5'-end of transcripts. Moreover, novel algorithms for reconstructing single-cell transcriptome may be essential to overcome the noise and low coverage of scRNA-seq data. Overall, conducting single-cell transcriptome reconstruction is promising for identifying the novel genes and isoforms (including both protein-coding and non-coding RNAs) expressed in specific cell



types/states, which may transform our understanding of the complexity of single-cell transcriptome.

PROFILING LONG NON-CODING RNAs AND CIRCULAR RNAs

After transcriptome reconstruction, novel lncRNAs could be identified from single cells. lncRNAs are the transcripts with >200 nucleotides in length and have no protein-coding potential. It has been shown that lncRNAs are fundamental regulators and involved in a wide range of biological processes and pathways related to transcriptional and posttranscriptional regulation as well as chromatin remodeling (Mercer et al., 2009;

Slack and Chinnaiyan, 2019). Moreover, lncRNAs can play critical roles in a variety of human diseases, and some of them could be important biomarkers for many cancers (Ransohoff et al., 2018). Additionally, the expression of lncRNAs is more tissue- and cell-type specific than that of mRNAs (Ransohoff et al., 2018); thus, scRNA-seq provides unprecedented opportunities for profiling and annotating the cell-type-specific lncRNAs. To identify lncRNAs with RNA-seq data, the aforementioned transcriptome reconstruction is usually conducted to define the map of all expressed transcription units first (Figure 5B). Then a variety of methods can be applied to discriminate lncRNAs from protein-coding RNAs, such as CPAT (Wang et al., 2013), lncRNA-ID (Achawanantakun et al., 2015), and LGC (Wang G. Y. et al., 2019) (Table 1). CPAT employs a

logistic regression model to discriminate between non-coding and protein-coding transcripts, while lncRNA-ID utilizes the machine learning model of random forest, and LGC is based on the feature relationship between the length of open reading frame (ORF) and GC content. The protein-coding potential assessment tools have been widely used in numerous studies to predict the protein-coding potential of transcripts, which have been reviewed previously (Han et al., 2016; Lorenzi et al., 2019).

An increasing number of studies have explored the lncRNA expression profiles and functions at the single-cell level. For example, Fan et al. (2015) developed SUPeR-seq to sequence both poly (A+) and poly (A-) RNAs and identified hundreds of novel lncRNAs that showed developmental stage-specific expression in mouse. The random (AnchorX-T₁₅N₆) primers were used in SUPeR-seq to enable the simultaneous capture of both polyadenylated and non-polyadenylated RNAs from individual cells. Moreover, novel lncRNAs associated with human early embryonic development were identified (Yan et al., 2013), and cell-type-specific lncRNAs were observed to be abundantly expressed in human neocortex (Liu S. J. et al., 2016). Besides, Wu et al. (2019) detected over 3,000 lncRNAs using the scRNA-seq data of human bone marrow and revealed that a fraction of them could play crucial roles in dysplastic hematopoiesis. It is worth noting that lncRNAs can localize in the nucleus and cytoplasm, and are usually less abundant than mRNAs, and can be expressed simultaneously with relevant protein-coding genes. If cells can be directly lysed without RNA extraction and sequenced with substantial depth, it may allow more comprehensive lncRNA identification. Furthermore, lncRNAs can be with or without poly (A) tails; thus, the full-length transcript scRNA-seq technologies that enable total RNA [including poly (A+) and poly (A-) RNAs] capturing [e.g., SUPeR-seq (Fan et al., 2015), MATQ-seq (Sheng et al., 2017), and RamDA-seq (Hayashi et al., 2018)] are more suitable for comprehensive lncRNA profiling, whereas those single-cell protocols that only sequence poly (A+) RNAs will miss the lncRNAs without poly-A tails. However, currently available scRNA-seq strategies that can provide whole gene body coverage are still suffering certain bias at the 3'/5'-end of transcripts; further improvement of these technologies will greatly benefit single-cell lncRNA profiling.

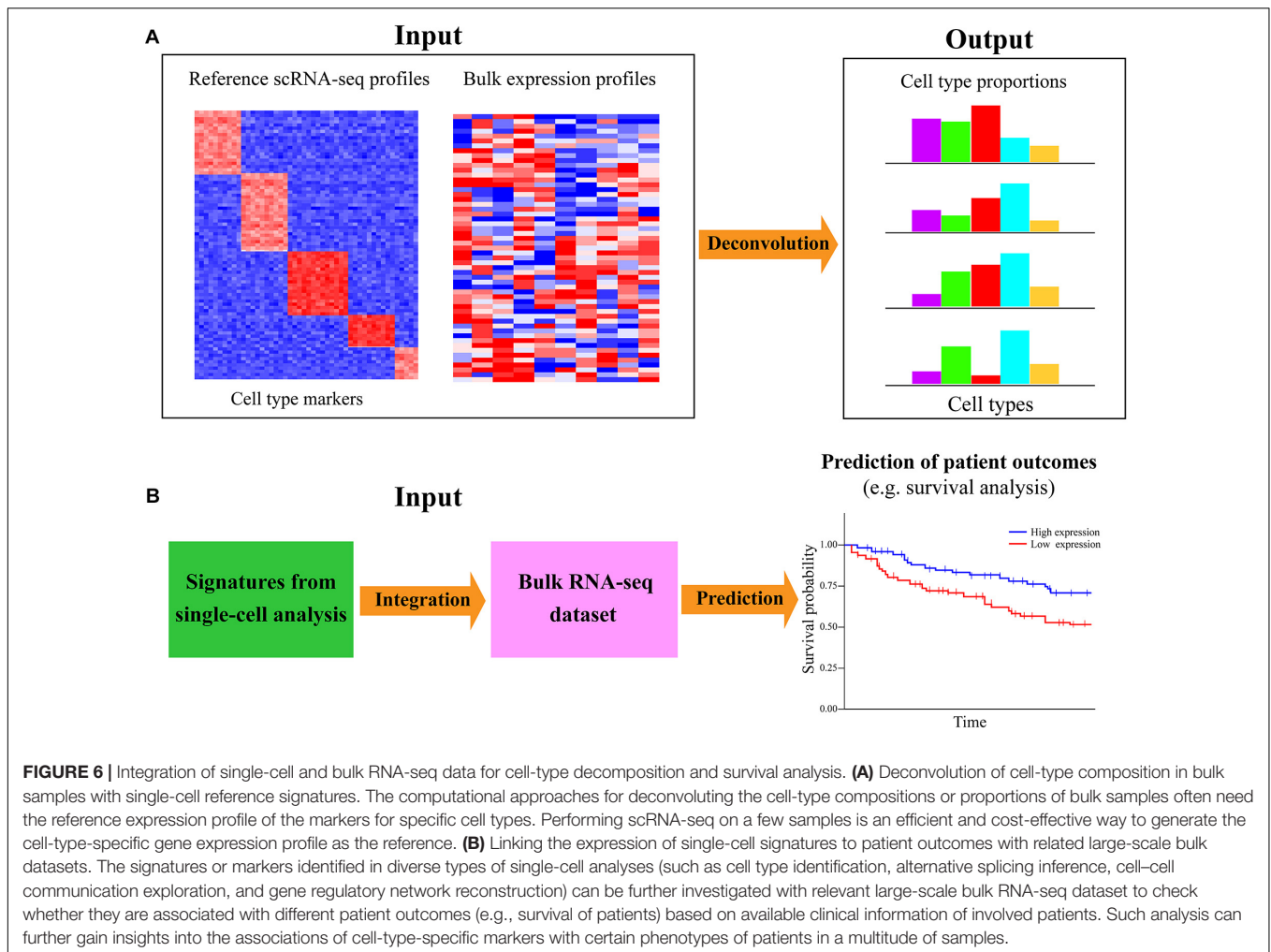
Additionally, circular RNAs (circRNAs) are an essential class of circularized non-coding RNAs, which are formed by back-splicing of linear pre-mRNAs (Figure 5C). CircRNAs can act as sponges for miRNAs or proteins, interfere with pre-mRNA processing, and even produce polypeptides (Lasda and Parker, 2014; Li et al., 2018). Moreover, a multitude of circRNAs have been associated with a variety of human cancers, and some of them could be important biomarkers for cancer diagnosis or prognosis (Greene et al., 2017). However, the specific functions for the great majority of circRNAs in biological systems are still unknown. Multitudinous studies have identified and annotated circRNAs with different bioinformatic pipelines based on bulk RNA-seq data (Memczak et al., 2013; Jakobi and Dieterich, 2019). However, circRNA exploration at the single-cell level is just emerging. Since circRNAs are covalently closed continuous loop and do not have poly (A) tail, they cannot be profiled with standard poly (A) enrichment protocols. Recently, an abundance

of circRNAs involved in the early embryonic development of mice was identified using SUPeR-seq protocol to sequence total RNAs from individual cells (Fan et al., 2015). Furthermore, Verboom et al. (2019) proposed SMARTer technology for conducting single-cell strand-specific total RNA sequencing and detected over 500 circRNAs in neuroblastoma cell lines. A range of computational methods are available for identifying circRNAs with RNA-seq data [such as find_circ2 (Memczak et al., 2013), CircExplorer2 (Zhang et al., 2016), and CIRI2 (Gao et al., 2018)], which have been reviewed recently (Jakobi and Dieterich, 2019) (Table 1). These tools could be applicable to explore the circRNAs in single cells. The commonly used bulk sequencing strategies for circRNA detection are ribosomal RNA (rRNA)-depleted total RNA and poly (A)-depletion methods, but none of them can guarantee that the enriched RNAs are exclusively circular as some other types of ncRNAs would be also captured (Kristensen et al., 2019). By contrast, the scRNA-seq protocols for profiling circRNAs are still in the early phases of development, and the bioinformatic methods specially designed for single-cell circRNA exploration are still lacking. Furthermore, the reliable identification and quantification of circRNAs generally need a substantial sequencing depth to obtain sufficient supporting reads spanning the back-splice junction region of circRNAs.

Currently, available scRNA-seq protocols are still with high technical noise, and the sequencing depth for each cell is relatively low in consideration of the cost, which hinders the identification of lncRNA and circRNA. Additionally, the computational methods specially developed to process single-cell transcriptomic data by taking the data sparsity and noise into account for lncRNA and circRNA investigation are currently lacking. With the development of both single-cell total RNA sequencing methods and related computational approaches, exploring the lncRNAs and circRNAs in individual cells, will be more feasible. These advancements will largely promote the profiling and functional characterization of lncRNAs and circRNAs in different cell types/states under various conditions.

CELL COMPOSITION DECONVOLUTION OF BULK SAMPLES USING SINGLE-CELL DATA

The aforementioned analyses are mainly based on scRNA-seq data alone; single-cell transcriptomic data can also be analyzed with the bulk RNA-seq dataset to infer the cell-type proportions/compositions for a large number of bulk samples (Figure 6A). ScRNA-seq has great advantages in dissecting the heterogeneity of cellular compositions within a given sample; however, such researches were mainly focused on a limited number of samples/individuals in consideration of cost effectiveness and scalability. Bulk RNA-seq is still the primary workhorse for dissecting gene expression for a host of samples in biomedical research due to the low cost and technical simplicity. For investigating the cell-subset specific information in a plethora of samples, an attractive approach is to directly decode the cell-type composition of large-scale heterogeneous bulk samples



via deconvolution algorithms (Shen-Orr and Gaujoux, 2013). Such a strategy is not only cost effective but could also preserve both whole-system level perspective and cell-based view of cell heterogeneity. For example, Li T. et al. (2017) have explored the composition of different tumor-infiltrating immune subsets in 32 cancer types of The Cancer Genome Atlas (TCGA). Moreover, Donovan et al. (2020) deconvoluted the cellular composition of 28 distinct human tissues from Genotype-Tissue Expression (GTEx) project (Aguet et al., 2017), which allowed cell-type-specific functional investigation for the impacts of genetic variation on gene expression.

Currently, a dozen of deconvolution approaches are available for inferring the composition of cell types from bulk RNA-seq data (Cobos et al., 2018), such as CMP (Frishberg et al., 2019), MuSiC (Wang X. R. et al., 2019), DWLS (Tsoucas et al., 2019), and CIBERSORTx (Newman et al., 2019) (Table 1). CMP uses linear regression to estimate the expression abundance of reference cells in the given bulk samples, while MuSiC weights the genes exhibiting cross-subject and cross-cell consistency to transfer cell-type-specific gene expression profile across different datasets. DWLS employs a weighted least squares method to estimate cell-type proportions, and CIBERSORTx is based on the machine

learning method to determine cell type abundance and cell-type-specific gene expression. A systematic comparison of the performance for recently developed deconvolution approaches is very valuable, but such a study is currently lacking. Existing deconvolution tools generally rely on the prior knowledge of reference expression profiles of known cell-type signatures, which can be obtained from the scRNA-seq data of one or a few samples (Figure 6A). At present, it is still highly expensive and time consuming to sequence a multitude of samples using scRNA-seq. Therefore, deconvoluting cell-type compositions from large-scale bulk RNA-seq dataset with a small sample size of single-cell transcriptomic data as the reference is an economically practical and time-saving way. Such analysis is valuable for identifying the cell types vulnerable to disease and detect the cellular targets of disease/cancer.

LINKING SINGLE-CELL SIGNATURE TO PATIENT OUTCOMES WITH BULK DATA

Another important joint analysis of scRNA-seq and bulk RNA-seq data is to associate the signatures identified in single-cell

transcriptomic data exploration to predict patient outcomes. Intratumoral heterogeneity is a pivotal determinant of tumor biology, survival, and treatment response of patients. A major goal of cancer profiling studies is to identify the genetic biomarkers that are predictive for the survival status of cancer patients. The advance in scRNA-seq largely facilitates the biomarker/signature detection at a higher resolution beyond traditional bulk data. Such single-cell signatures can be screened out from different types of single-cell analyses, such as cell clustering, differential expression calling, alternative splicing exploration, and gene regulatory network inference. Specifically, important signatures could be identified from the scRNA-seq data of the tumor ecosystem to potentially predict cancer stage, therapy response, disease-free interval, metastatic probability, or overall patient survival. Although it may not be practical to perform scRNA-seq on an abundance of patients for prognosis prediction, those publicly accessible bulk datasets with available clinical information are valuable resources for such analysis. It is an alternative way to assess whether the single-cell signatures could be useful biomarkers for predicting patient outcomes (**Figure 6B**).

A host of studies have used the bulk datasets from public databases like TCGA (Weinstein et al., 2013) and Gene Expression Omnibus (GEO) to determine the association between the expression level of single-cell signatures and the patient survival of corresponding cancers. For example, signatures from scRNA-seq analysis were successfully applied to predict the overall survival of patients for TCGA melanoma (Nirschl et al., 2017) and hepatocellular carcinoma (Zheng et al., 2018). Furthermore, Li H. P. et al. (2017) identified single-cell biomarkers that can stratify the colorectal tumors from TCGA and GEO databases into subgroups with divergent survival. For survival analysis, Raman et al. (2019) revealed that highly variable results are usually obtained from different methods, and Cox regression (Li, 2003) is superior to other compared approaches based on tests of reliability, accuracy, and robustness. Cox regression is a flexible method that can improve the accuracy of estimation between gene expression level and patient survival by enabling the inclusion of multiple covariates to accommodate explanatory variables. It is worth noting that the single-cell signatures are used to build a model, while the actual data using the model is the bulk RNA-seq data. The continuous decreasing cost and time for scRNA-seq will make single-cell transcriptomic profiling on a large sample size become more affordable and practicable, which will greatly benefit the association analysis between single-cell signatures and patient outcomes. Consequently, the signatures/biomarkers screened out from diverse kinds of single-cell analyses could be further linked to the patient outcomes with related bulk datasets and clinical information to assess their associations and clinical value.

CONCLUSION AND OUTLOOK

ScRNA-seq is widely applied to diverse organisms to dissect a range of biological questions related to developmental biology,

oncology, immunology, neurology, and microbiology at the single-cell resolution. Besides those routine analyses conducted in most studies (e.g., cell type identification, alternative splicing detection, trajectory, and GRN inference), much more other valuable information can be mined from scRNA-seq data. As we summarized in this review, cell-to-cell communications, RNA velocity, and large-scale CNVs and chromatin accessibility could be effectively extracted from single-cell transcriptomic data. Nucleotide sequence changes of SNVs and RNA editing events also could be derived from scRNA-seq experiments to enable multiple modalities. Moreover, transcriptome reconstruction with full-length transcript scRNA-seq data is promising for identifying and annotating the novel genes and isoforms mainly expressed in certain cell types/states. The innovation and optimization of scRNA-seq protocols that can effectively capture both poly (A+) and poly (A-) RNAs with increased throughput will improve the feasibility of profiling and characterizing of lncRNAs and circRNAs at single-cell resolution. Additionally, the results of scRNA-seq analysis can be further explored with traditional bulk RNA-seq data to deconvolute the cell compositions in a multitude of bulk samples or assess the association between single-cell signatures and patient outcomes in a cost-effective way.

Notably, the accuracy of any kind of single-cell analysis largely depends on the quality of single-cell sequencing data (e.g., cell quality, sequencing quality, coverage, and depth) as well as the performance of corresponding bioinformatics algorithms. Special attention needs to be paid to the noise and sparsity of scRNA-seq data, and stringent criteria may be needed to minimize the false positives. Besides, since there is a general lack of studies for benchmarking the computational approaches of the single-cell analyses we summarized in this review, it would be useful to conduct such researches in the future. In consideration of the absence of a gold-standard method, running more than one bioinformatic tools could be an effective way to reduce the number of false positives. Additionally, performance comparison for several commonly used scRNA-seq technologies revealed that if the research goal aims to pursue the highest sensitivity, the low-throughput methods that can produce full-length transcripts (e.g., Smart-seq2) are significantly better than the high-throughput approaches that mainly capture the 3'/5'-end of transcripts (like 10x Chromium) (Ziegenhain et al., 2017; Ding et al., 2020). Future comparative analysis for the newly developed single-cell transcriptome profiling protocols will be very helpful to provide better guidance in experimental designs. The fast evolution of both scRNA-seq approaches and bioinformatics methods will make the single-cell analyses we discussed become more feasible. We anticipate that these useful analyses will add much more value to scRNA-seq data and largely facilitate biomedical and clinical researches.

On the other hand, the states of single cells are determined by the intricate interplay of various molecules from multi-omic levels, such as genomics, transcriptomics, proteomics, and epigenomics. Integrative analysis of multi-omic data will enable a much more comprehensive and systematic view of each cell, which will greatly benefit the study of a variety

of normal development and disease processes. An increasing number of single-cell protocols have been developed to measure different modalities including genome (Vitak et al., 2017), epigenome (Mulqueen et al., 2018), proteome (Darmanis et al., 2016), and chromatin accessibility (Cusanovich et al., 2015), as well as profile spatial (Wang et al., 2018) or lineage (Raj et al., 2018) information. Furthermore, some assays can even simultaneously capture multimodal data from the same cell (Stuart and Satija, 2019). Additionally, the third-generation sequencing technologies like nanopore can sequence RNA and DNA with super long reads (Rand et al., 2017; Garalde et al., 2018); such technological advances and improvements will effectively accelerate the refinement of single-cell multi-omic approaches. As single-cell technology matures (including sensitivity, coverage, and throughput) and the continuous decrease in cost, multi-omic studies will be more feasible and affordable. Collectively, we envision that the advances of multi-omic assays coupled with novel computational approaches will enable a more comprehensive understanding and elucidation of

diverse cellular processes and significantly transform the single-cell biology.

AUTHOR CONTRIBUTIONS

GC conceived the manuscript. GC, YL, QX, and DW wrote the review. YL plotted all the figures. GC provided guidance on the writing and direction of the review. GC and YL revised and finalized the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (31771460, 32070680, and 91629103) and the National Key Research and Development Program of China (2016YFC0902100).

REFERENCES

- Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 545, 446–451. doi: 10.1038/nature22364
- Achawanantakun, R., Chen, J., Sun, Y. N., and Zhang, Y. (2015). LncRNA-ID: long non-coding RNA identification using balanced random forests. *Bioinformatics* 31, 3897–3905. doi: 10.1093/bioinformatics/btv480
- Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277
- Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086. doi: 10.1038/nmeth.4463
- Andrews, T. S., and Hemberg, M. (2018). Identifying cell populations with scRNASeq. *Mol. Aspects Med.* 59, 114–122. doi: 10.1016/j.mam.2017.07.002
- Bergen, V., Lange, M., Peidli, S., Wolf, F. A., and Theis, F. J. (2019). Generalizing RNA velocity to transient cell states through dynamical modeling. *bioRxiv [Preprint]* doi: 10.1101/820936
- Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17, 159–162. doi: 10.1038/s41592-019-0667-5
- Brown, T. I., Carr, V. A., LaRocque, K. F., Favila, S. E., Gordon, A. M., Bowles, B., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* 352, 1323–1326.
- Bryois, J., Buil, A., Evans, D. M., Kemp, J. P., Montgomery, S. B., Conrad, D. F., et al. (2014). Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 10:e1004461. doi: 10.1371/journal.pgen.1004461
- Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8:giz100.
- Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* 48:e55. doi: 10.1093/nar/gkaa183
- Camp, J. G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. *Nature* 546, 533–538. doi: 10.1038/nature22796
- Cang, Z., and Nie, Q. (2020). Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* 11:2084.
- Chen, G., Ning, B. T., and Shi, T. L. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* 10:317. doi: 10.3389/fgene.2019.00317
- Chen, G., Shi, T. L., and Shi, L. M. (2017). Characterizing and annotating the genome using RNA-seq data. *Sci. China Life Sci.* 60, 116–125. doi: 10.1007/s11427-015-0349-4
- Chen, G., Wang, C., Shi, L. M., Qu, X. F., Chen, J. W., Yang, J. M., et al. (2013). Incorporating the human gene annotations in different databases significantly improved transcriptomic and genetic analyses. *RNA* 19, 479–489. doi: 10.1261/rna.037473.112
- Chen, S. N., and Mar, J. C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* 19:232. doi: 10.1186/s12859-018-2217-z
- Cobos, F. A., Vandesompele, J., Mestdagh, P., and De Preter, K. (2018). Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* 34, 1969–1979. doi: 10.1093/bioinformatics/bty019
- Cohen, M., Giladi, A., Gorki, A. D., Solodkin, D. G., Zada, M., Hladik, A., et al. (2018). Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell* 175, 1031.e18–1044.e18.
- Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., et al. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348, 910–914. doi: 10.1126/science.aab1601
- Darmanis, S., Gallant, C. J., Marinescu, V. D., Niklasson, M., Segerman, A., Flamourakis, G., et al. (2016). Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* 14, 380–389. doi: 10.1016/j.celrep.2015.12.021
- Diaz-Papkovich, A., Anderson-Trocme, L., Ben-Eghan, C., and Gravel, S. (2019). UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genet.* 15:e1008432. doi: 10.1371/journal.pgen.1008432
- Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* 38, 737–746.
- Ding, J., Lin, C., and Bar-Joseph, Z. (2019). Cell lineage inference from SNP and scRNA-Seq data. *Nucleic Acids Res.* 47:e56. doi: 10.1093/nar/gkz146
- Donovan, M., D'Antonio-Chronowska, A., D'Antonio, M., and Frazer, K. (2020). Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nat. Commun.* 11:955.
- Duò, A., Robinson, M. D., and Sonesson, C. (2018). A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res* 7:1141. doi: 10.12688/f1000research.15666.1
- Efremova, M., and Teichmann, S. A. (2020). Computational methods for single-cell omics across modalities. *Nat. Methods* 17, 14–17. doi: 10.1038/s41592-019-0692-4

- Efremova, M., Vento-Tormo, M., Teichmann, S. A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15, 1484–1506. doi: 10.1038/s41596-020-0292-x
- Enge, M., Arda, H. E., Mignardi, M., Beusang, 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, 321.e14–330.e14.
- Fan, J., Lee, H. O., Lee, S., Ryu, D. E., Lee, S., Xue, C., et al. (2018). Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* 28, 1217–1227. doi: 10.1101/gr.228080.117
- Fan, X. Y., Zhang, X. N., Wu, X. L., Guo, H. S., Hu, Y. Q., Tang, F. C., et al. (2015). Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* 16:148.
- Ferguson, L. R., Chen, H., Collins, A. R., Connell, M., Damia, G., Dasgupta, S., et al. (2015). Genomic instability in human cancer: molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. *Semin. Cancer Biol.* 35, S5–S24.
- Fiers, M. W. E. J., Minnoye, L., Aibar, S., Gonzalez-Blas, C. B., Atak, Z. K., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics* 17, 246–254. doi: 10.1093/bfpg/ebx046
- Frishberg, A., Peshes-Yaloz, N., Cohn, O., Rosentul, D., Steuerman, Y., Valadarsky, L., et al. (2019). Cell composition analysis of bulk genomics using single-cell data. *Nat. Methods* 16, 327–332. doi: 10.1038/s41592-019-0355-5
- Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M. B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.* 33, 722–729. doi: 10.1038/nbt.3269
- Gao, Y., Zhang, J. Y., and Zhao, F. Q. (2018). Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 19, 803–810. doi: 10.1093/bib/bbx014
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipsos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201–206. doi: 10.1038/nmeth.4577
- Garber, M., Grabherr, M. G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477. doi: 10.1038/nmeth.1613
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *Quant. Biol.*
- Gorelik, L., and Flavell, R. A. (2000). Abrogation of TGFbeta signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* 12, 171–181. doi: 10.1016/s1074-7613(00)80170-3
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Greene, J., Baird, A. M., Brady, L., Lim, M., Gray, S. G., McDermott, R., et al. (2017). Circular RNAs: biogenesis, function and role in human diseases. *Front. Mol. Biosci.* 4:38. doi: 10.3389/fmolb.2017.00038
- Haass, N. K., Smalley, K. S. M., and Herlyn, M. (2004). The role of altered cell-cell communication in melanoma progression. *J. Mol. Histol.* 35, 309–318. doi: 10.1023/b:hijo.0000032362.35354.bb
- Han, S. Y., Liang, Y. C., Li, Y., and Du, W. (2016). Long noncoding RNA identification: comparing machine learning based tools for long noncoding transcripts discrimination. *Biomed. Res. Int.* 2016:8496165.
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013
- 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.* 9:619.
- Hotamisligil, G. S. (2006). Inflammation and metabolic disorders. *Nature* 444, 860–867. doi: 10.1038/nature05485
- Jakobi, T., and Dieterich, C. (2019). Computational approaches for circular RNA analysis. *Wiley Interdiscip. Rev. RNA* 10:e1528. doi: 10.1002/wrna.1528
- Jansen, C., Ramirez, R. N., El-Ali, N. C., Gomez-Cabrero, D., Tegner, J., Merckenschlager, M., et al. (2019). Building gene regulatory networks from scATAC-seq and scRNA-seq using linked self organizing maps. *PLoS Comput. Biol.* 15:e1006555. doi: 10.1371/journal.pcbi.1006555
- Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Myung, P., Plikus, M. V., et al. (2020a). Inference and analysis of cell-cell communication using CellChat. *bioRxiv [Preprint]* doi: 10.1101/2020.07.21.214387
- Jin, S., Zhang, L., and Nie, Q. (2020b). scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. *Genome Biol.* 21:25.
- Jin, W. F., Tang, Q. S., Wan, M. M., Cui, K. R., Zhang, Y., Ren, G., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* 528, 142–146. doi: 10.1038/nature15740
- Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., et al. (2017). Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543, 714–718.
- Kanata, E., Llorens, F., Dafou, D., Dimitriadis, A., Thune, K., Xanthopoulos, K., et al. (2019). RNA editing alterations define manifestation of prion diseases. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19727–19735. doi: 10.1073/pnas.1803521116
- Kanton, S., Boyle, M. J., He, Z. S., Santel, M., Weigert, A., Sanchis-Calleja, F., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* 574, 418–422. doi: 10.1038/s41586-019-1654-9
- Karaïskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., et al. (2017). The drosophila embryo at single-cell transcriptome resolution. *Science* 358, 194–199.
- Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Kallberg, M., et al. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* 15, 591–594. doi: 10.1038/s41592-018-0051-x
- Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20, 310–310. doi: 10.1038/s41576-019-0095-5
- Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220. doi: 10.1038/s41576-018-0089-8
- Kreso, A., and Dick, J. E. (2014). Evolution of the cancer stem cell model. *Cell Stem Cell* 14, 275–291. doi: 10.1016/j.stem.2014.02.006
- Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691.
- Kumar, M. P., Du, J. Y., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D. C., et al. (2018). Analysis of single-cell RNA-seq identifies cell-cell communication associated with tumor characteristics. *Cell Rep.* 25, 1458.e4–1468.e4.
- Kumar, V., Muratani, M., Rayan, N. A., Kraus, P., Lufkin, T., Ng, H. H., et al. (2013). Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* 31, 615–622. doi: 10.1038/nbt.2596
- Kung, C. P., Maggi, L. B., and Weber, J. D. (2018). The role of RNA editing in cancer development and metabolic disorders. *Front. Endocrinol.* 9:762. doi: 10.3389/fendo.2018.00762
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498.
- Lake, B. B., Chen, S., Sos, B. C., Fan, J., Kaeser, G. E., Yung, Y. C., et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36, 70–80. doi: 10.1038/nbt.4038
- Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842. doi: 10.1261/rna.047126.114
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H. P., Courtois, E. T., Sengupta, D., Tan, Y. L., Chen, K. H., Goh, J. J. L., et al. (2017). Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* 49, 708–718. doi: 10.1038/ng.3818
- Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77, e108–e110.
- Li, J. C. A. (2003). Modeling survival data: extending the cox model. *Sociol. Method Res.* 32, 117–120.
- Li, X., Yang, L., and Chen, L. L. (2018). The biogenesis, functions, and challenges of circular RNAs. *Mol. Cell* 71, 428–442. doi: 10.1016/j.molcel.2018.06.034

- Liu, F. L., Zhang, Y. Y., Zhang, L., Li, Z. Y., Fang, Q., Gao, R. R., et al. (2019). Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. *Genome Biol.* 20:242.
- Liu, J. T., Yu, T., Jiang, T., and Li, G. J. (2016). TransComb: genome-guided transcriptome assembly via combing junctions in splicing graphs. *Genome Biol.* 17:213.
- Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., et al. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* 17:67.
- Lorenzi, L., Cobos, F. A., Decock, A., Everaert, C., Helmsmoortel, H., Lefever, S., et al. (2019). Long noncoding RNA expression profiling in cancer: challenges and opportunities. *Gene Chromosome Canc.* 58, 191–199. doi: 10.1002/gcc.22709
- Lueken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15:e8746.
- Lun, A. T. L., Bach, K., and Marioni, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17:75.
- Maas, S., Kawahara, Y., Tamburro, K. M., and Nishikura, K. (2006). A-to-I RNA editing and human disease. *RNA Biol.* 3, 1–9. doi: 10.4161/rna.3.1.2495
- 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, 1202–1214. doi: 10.1016/j.cell.2015.05.002
- Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., et al. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell* 171, 1029.e21–1041.e21.
- Mayr, U., Serra, D., and Liberali, P. (2019). Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development* 146:dev176727. doi: 10.1242/dev.176727
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928
- Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159.
- Moon, K. R., Stanley, J. S., Burkhardt, D., van Dijk, D., Wolf, G., and Krishnaswamy, S. (2018). Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr. Opin. Syst. Biol.* 7, 36–46. doi: 10.1016/j.coisb.2017.12.008
- Mulqueen, R. M., Pokholok, D., Norberg, S. J., Torkenczy, K. A., Fields, A. J., Sun, D. C., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* 36, 428–431. doi: 10.1038/nbt.4112
- Natarajan, A., Yardimci, G. G., Sheffield, N. C., Crawford, G. E., and Ohler, U. (2012). Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* 22, 1711–1722. doi: 10.1101/gr.135129.111
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94. doi: 10.1038/nature09807
- Navin, N., Krasnitz, A., Rodgers, L., Cook, K., Meth, J., Kendall, J., et al. (2010). Inferring tumor progression from genomic heterogeneity. *Genome Res.* 20, 68–80. doi: 10.1101/gr.099622.109
- Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic instability - an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Bio.* 11, 220–228. doi: 10.1038/nrm2858
- Neph, S., Stergachis, A. B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J. A. (2012). Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 150, 1274–1286. doi: 10.1016/j.cell.2012.04.040
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782. doi: 10.1038/s41587-019-0114-2
- Nirschl, C. J., Suárez-Fariñas, M., Izar, B., Prakadan, S., Dannenfels, R., Tirosh, I., et al. (2017). IFN γ -dependent tissue-immune homeostasis is co-opted in the tumor microenvironment. *Cell* 170, 127.e15–141.e15.
- Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79, 321–349. doi: 10.1146/annurev-biochem-060208-105251
- Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Bio.* 17, 83–96. doi: 10.1038/nrm.2015.4
- Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. *Nature* 576, 132–137. doi: 10.1038/s41586-019-1773-3
- Park, E., Williams, B., Wold, B. J., and Mortazavi, A. (2012). RNA editing in the human ENCODE RNA-seq data. *Genome Res.* 22, 1626–1633. doi: 10.1101/gr.134957.111
- 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, 1396–1401. doi: 10.1126/science.1254257
- Pavlicev, M., Wagner, G. P., Chavan, A. R., Owens, K., Maziarz, J., Dunn-Fletcher, C., et al. (2017). Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* 27, 349–361. doi: 10.1101/gr.207597.116
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., and Bejerano, G. (2013). Enhancers: five essential questions. *Nat. Rev. Genet.* 14, 288–295. doi: 10.1038/nrg3458
- Perlea, M., Perlea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Picardi, E., and Pesole, G. (2013). REDIttools: high-throughput RNA editing detection made easy. *Bioinformatics* 29, 1813–1814. doi: 10.1093/bioinformatics/btt287
- Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181. doi: 10.1038/nprot.2014.006
- Poirion, O., Zhu, X., Ching, T., and Garmire, L. X. (2018). Using single nucleotide variations in single-cell RNA-seq to identify subpopulations and genotype-phenotype linkage. *Nat. Commun.* 9:4892.
- Pratapa, A., Jaliyal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17, 147–154. doi: 10.1038/s41592-019-0690-6
- Puram, S. V., Tirosh, I., Park, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171, 1611.e24–1624.e24.
- Quadrato, G., Nguyen, T., Macosko, E. Z., Sherwood, J. L., Yang, S. M., Berger, D. R., et al. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545, 48–53. doi: 10.1038/nature22047
- 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, 442–450. doi: 10.1038/nbt.4103
- Raman, P., Zimmerman, S., Rathi, K. S., de Torrenté, L., Sarmady, M., Wu, C., et al. (2019). A comparison of survival analysis methods for cancer gene expression RNA-Sequencing data. *Cancer Genet.* 235–236, 1–12. doi: 10.1016/j.cancergen.2019.04.004
- Ramaswami, G., Zhang, R., Piskol, R., Keegan, L. P., Deng, P., O'Connell, M. A., et al. (2013). Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* 10, 128–132. doi: 10.1038/nmeth.2330
- Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., et al. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* 14, 411–413. doi: 10.1038/nmeth.4189
- Ransohoff, J. D., Wei, Y. N., and Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Bio.* 19, 143–157. doi: 10.1038/nrm.2017.104
- Ren, X., Zhong, G., Zhang, Q., Zhang, L., Sun, Y., and Zhang, Z. (2020). Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* 30, 763–778. doi: 10.1038/s41422-020-0353-2
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., et al. (2010). De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912.
- Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37, 547–554. doi: 10.1038/s41587-019-0071-9
- Sheng, K., Cao, W., Niu, Y., Deng, Q., and Zong, C. (2017). Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* 14, 267–270. doi: 10.1038/nmeth.4145
- Shen-Orr, S. S., and Gaujoux, R. (2013). Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr. Opin. Immunol.* 25, 571–578. doi: 10.1016/j.coi.2013.09.015
- Shlien, A., and Malkin, D. (2009). Copy number variations and cancer. *Genome Med.* 1:62.

- Skelly, D. A., Squiers, G. T., McLellan, M. A., Bolisetty, M. T., Robson, P., Rosenthal, N. A., et al. (2018). Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Rep.* 22, 600–610. doi: 10.1016/j.celrep.2017.12.072
- Slack, F. J., and Chinnaiyan, A. M. (2019). The role of non-coding RNAs in oncology. *Cell* 179, 1033–1055. doi: 10.1016/j.cell.2019.10.017
- Slotkin, W., and Nishikura, K. (2013). Adenosine-to-inosine RNA editing and human disease. *Genome Med.* 5:105. doi: 10.1186/gm508
- Soldatov, R., Kaucska, M., Kastriti, M. E., Petersen, J., Chontorotzea, T., Englmaier, L., et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* 364:eaas9536.
- Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272. doi: 10.1038/s41576-019-0093-7
- Svensson, V., and Pachter, L. (2018). RNA velocity: molecular kinetics from single-cell RNA-seq. *Mol. Cell* 72, 7–9. doi: 10.1016/j.molcel.2018.09.026
- Tang, L. (2020). Integrating lineage tracing and single-cell analysis. *Nat. Methods* 17:359. doi: 10.1038/s41592-020-0802-3
- 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, 189–196.
- 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 oligodendrogloma. *Nature* 539, 309–313.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tsoucas, D., Dong, R., Chen, H. D., Zhu, Q., Guo, G. J., and Yuan, G. C. (2019). Accurate estimation of cell-type composition from gene expression data. *Nat. Commun.* 10:2975.
- Tsuyuzaki, K., Ishii, M., and Nikaido, I. (2019). Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. *bioRxiv [Preprint]* doi: 10.1101/566182
- Tyler, S. R., Rotti, P. G., Sun, X. S., Yi, Y. L., Xie, W. L., Winter, M. C., et al. (2019). PyMINER Finds Gene and Autocrine-Paracrine Networks from Human Islet scRNA-Seq. *Cell Rep.* 26, 1951.e8–1964.e8.
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Verboom, K., Everaert, C., Bolduc, N., Livak, K. J., Yigit, N., Rombaut, D., et al. (2019). SMARTer single cell total RNA sequencing. *Nucleic Acids Res.* 47:e93.
- Vitak, S. A., Torkency, K. A., Rosenkrantz, J. L., Fields, A. J., Christiansen, L., Wong, M. H., et al. (2017). Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* 14, 302–308. doi: 10.1038/nmeth.4154
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., and Kinzler, K. W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558. doi: 10.1126/science.1235122
- Voss, T. C., and Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* 15, 69–81. doi: 10.1038/nrg3623
- Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981–987. doi: 10.1126/science.aar4362
- Wang, C., Sun, D., Huang, X., Wan, C., Li, Z., Han, Y., et al. (2020). Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.* 21:198.
- Wang, G. Y., Yin, H. Y., Li, B. Y., Yu, C. L., Wang, F., Xu, X. J., et al. (2019). Characterization and identification of long non-coding RNAs based on feature relationship. *Bioinformatics* 35, 2949–2956. doi: 10.1093/bioinformatics/bt z008
- Wang, X. R., Park, J., Susztak, K., Zhang, N. R., and Li, M. Y. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* 10:380.
- Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., et al. (2019). iTALK: an R package to characterize and illustrate intercellular communication. *bioRxiv [Preprint]* doi: 10.1101/507871
- Wang, L., Park, H. J., Dasari, S., Wang, S. Q., Kocher, J. P., and Li, W. (2013). CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41:e74. doi: 10.1093/nar/gk t006
- Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361:eaat5691. doi: 10.1126/science.aat5691
- Watabe, T., and Miyazono, K. (2009). Roles of TGF-beta family signaling in stem cell renewal and differentiation. *Cell Res.* 19, 103–115. doi: 10.1038/cr.2008.323
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764
- Wu, Z. J., Gao, S. G., Zhao, X., Chen, J. G., Keyvanfar, K., Feng, X. M., et al. (2019). Long noncoding RNAs of single hematopoietic stem and progenitor cells in healthy and dysplastic human bone marrow. *Haematologica* 104, 894–906. doi: 10.3324/haematol.2018.208926
- Xu, X., Hou, Y., Yin, X. Y., Bao, L., Tang, A. F., Song, L. T., et al. (2012). Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148, 886–895. doi: 10.1016/j.cell.2012.02.025
- Yan, L. Y., Yang, M. Y., Guo, H. S., Yang, L., Wu, J., Li, R., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139. doi: 10.1038/nsmb.2660
- Yip, S. H., Sham, P. C., and Wang, J. W. (2019). Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. *Brief. Bioinform.* 20, 1583–1589. doi: 10.1093/bib/bby011
- Yu, W., Uzun, Y., Zhu, Q., Chen, C., and Tan, K. (2020). scATAC-pro: a comprehensive workflow for single-cell chromatin accessibility sequencing data. *Genome Biol.* 21:94. doi: 10.1186/s13059-020-02008-0
- Yuan, D. S., Tao, Y. R., Chen, G., and Shi, T. L. (2019). Systematic expression analysis of ligand-receptor pairs reveals important cell-to-cell interactions inside glioma. *Cell Commun. Signal* 17:48.
- Zafar, H., Wang, Y., Nakhleh, L., Navin, N., and Chen, K. (2016). Monovar: single-nucleotide variant detection in single cells. *Nat. Methods* 13, 505–507. doi: 10.1038/nmeth.3835
- Zeisel, A., Kostler, W. J., Molotski, N., Tsai, J. M., Krauthgamer, R., Jacob-Hirsch, J., et al. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.* 7:529. doi: 10.1038/msb.2011.62
- Zhang, Q., and Xiao, X. (2015). Genome sequence-independent identification of RNA editing sites. *Nat. Methods* 12, 347–350. doi: 10.1038/nmeth.3314
- Zhang, Q. M., He, Y., Luo, N., Patel, S. J., Han, Y. J., Gao, R. R., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 179, 829.e20–845.e20.
- Zhang, X. O., Dong, R., Zhang, Y., Zhang, J. L., Luo, Z., Zhang, J., et al. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287. doi: 10.1101/gr.202895.115
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049.
- Zheng, H. P., Pomyen, Y., Hernandez, M. O., Li, C. Y., Livak, F., Tang, W., et al. (2018). Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. *Hepatology* 68, 127–140. doi: 10.1002/hep.29778
- Zhou, W. Q., Ji, Z. C., Fang, W. X., and Ji, H. K. (2019). Global prediction of chromatin accessibility using small-cell-number and single-cell RNA-seq. *Nucleic Acids Res.* 47:e121. doi: 10.1093/nar/gkz716
- Zhou, W. Q., Sherwood, B., Ji, Z. C., Xue, Y. C., Du, F., Bai, J. W., et al. (2017). Genome-wide prediction of DNase I hypersensitivity using gene expression. *Nat. Commun.* 8:1038.
- 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, 631.e4–643.e4. doi: 10.1016/j.molcel.2017.01.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Li, Xu, Wu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.