



REVIEW

Integration of Computational Analysis and Spatial Transcriptomics in Single-cell Studies



Ran Wang¹, Guangdun Peng^{2,3}, Patrick P.L. Tam^{4,5}, Naihe Jing^{1,2,3,6,*}

¹ State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai 200031, China

² CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

³ Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

⁴ Embryology Research Unit, Children's Medical Research Institute, University of Sydney, Sydney, NSW 2145, Australia

⁵ School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2145, Australia

⁶ Guangzhou Laboratory, Guangzhou 510005, China

Received 17 March 2022; revised 8 June 2022; accepted 21 June 2022

Available online 25 July 2022

Handled by Wei Lin

KEYWORDS

scRNA-seq;
Computational
methodology;
Spatial transcriptome;
Data integration;
Mathematical model

Abstract Recent advances of single-cell transcriptomics technologies and allied computational methodologies have revolutionized molecular cell biology. Meanwhile, pioneering explorations in spatial transcriptomics have opened up avenues to address fundamental biological questions in health and diseases. Here, we review the technical attributes of single-cell RNA sequencing and spatial transcriptomics, and the core concepts of computational data analysis. We further highlight the challenges in the application of **data integration** methodologies and the interpretation of the biological context of the findings.

Introduction

High-throughput sequencing techniques, in particular single-cell omics analyses, have revolutionized molecular cell biology research. Single-cell RNA sequencing (scRNA-seq) is one of the most widely used single-cell analytical approaches. Since the first scRNA-seq profiling of only eight cells [1], the analysis

has now been expanded to profile the transcriptome of around 2 million single cells [2]. The huge amounts of scRNA-seq data are of great potential to provide comprehensive cell atlas and cell connectivity for each biological system.

In order to make the most of these rich datasets, an effective computational analysis of single-cell data, including data quality control, read mapping, normalization, dimensionality reduction, clustering, heterogeneous cell population identification, differential gene expression, and trajectory inference, is an essential requisite to achieve a biologically meaning outcome of the study. To this end, a multitude of analytical tools are available, such as principal component analysis (PCA) [3], t-distributed stochastic neighbor embedding (t-SNE) [4],

* Corresponding author.

E-mail: njing@sibcb.ac.cn (Jing N).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2022.06.006>

1672-0229 © 2023 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

uniform manifold approximation and projection (UMAP) [5], and pseudotime inference [6]. Additionally, there are applications that integrate the analytical workflow for single-cell data, *e.g.*, Scran [7] and Seurat [8]. As each method has unique kernel algorithm and functionality, it is imperative to have a good understanding of computational attribute and biological implication of the methodology.

Further than identifying the signature features of major cell types in a cell population and characterizing the differentiation process, there is the aspiration to characterize the heterogeneity of cell types and the cell membership in a population [6]. The conventional single-cell assays necessitate the dissociation of tissue samples and does not register the tissue source and whereabouts of the single cells [9]. Since biological processes and cellular functions may be influenced by the immediate environment of cells, missing the spatial information during the isolation of cells is a major drawback of many single-cell studies [10]. Here entered the spatial transcriptome analysis that enables the elucidation of the cell type composition of a population in a defined tissue domain, organ, or body part of an organism. From these findings, inferences can be made on cell states and lineage fate, and together with the additional temporal information, the activation of specific gene regulatory networks (GRNs) along a developmental trajectory in time and space of specific cell types, and emergence of cellular complexity in an organism [8]. To accomplish this, analytical methods to achieve spatial resolution are developed, such as sequential fluorescence *in situ* hybridization (seqFISH) [11], multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) [12], and slide sequencing (Slide-seq) [13]. Harnessing the power of high-resolution gene expression profiles and spatial reference systems by integrating the spatial coordinates gleaned from these studies with single-cell omics data would potentially empower the study of single-cell biology [14,15].

This review focuses on bioinformatics methodology and spatial transcriptomics, which encompass most single-cell studies published to date. We outline the advances of single-cell analytical techniques, compare the technical attributes of various spatial omics technologies, and highlight the challenges and future trend of single-cell biology research.

State-of-the-art analysis techniques

Recent advances in scRNA-seq technology provide unprecedented opportunities to investigate cell-to-cell heterogeneities, to identify new cell subtypes, and to model differentiation processes [16]. The analytical methodologies for scRNA-seq data differ dramatically from those for bulk RNA sequencing (RNA-seq). We summarize the core concepts of computational methodologies of scRNA-seq data analysis.

Dimensionality reduction analysis for scRNA-seq data

In contrast to bulk RNA-seq, scRNA-seq generates complex datasets of high cell numbers and high-dimensional expression data [17], which require stringent quality control tests [18,19] (Figure 1A, i). Dimensionality reduction is the primary step for scRNA-seq data analysis (Figure 1A, ii). PCA is one widely used dimensionality reduction method to extract variations and bring out significant patterns in a dataset [3]. PCA reduces

the data dimension by geometrically projecting them onto fewer principal components (PCs), which act as summarizing the keynote features of data structure. Regarding the linear combination of the data's primary variables, the top PCs, generally the top 1–3, reflect the foremost biological differences among cell clusters, since these PCs explain the major variances. Top PC loading genes, which can be interpreted as the signature genes of biological significance, account for the major variances for delineation of cell clustering, and can be used for further iterative clustering analysis [9,20].

With the rapid advances of experimental techniques that could be applied to analyze an exponentially increasing number of cells per study, conventional methods may not be ideal for analyzing the complex multivariate data [21]. Among the computational tools and methodologies, the non-linear dimensionality reduction method [22], t-SNE, enables creating a single map that displays fine structure at many different levels (Figure 1A, ii). This is especially useful for high-dimensional data that comprise several distinct, but related, low-dimensional manifolds, such as graphics of objects from different classes seen from different viewpoints. For visualizing the structure of extremely large datasets, t-SNE uses random walk algorithm on neighborhood graphs to allow the potential and implicit structure of all of the data to be deduced from a concise sub-dataset.

Similar to t-SNE, UMAP is a latest non-linear dimensionality reduction method with great visualization capabilities [5]. UMAP is one of the fastest manifold learning implementations, and has a significant improvement over t-SNE in runtime and workload, especially for extremely large datasets [23–25], such as those in cell atlas studies (Figure 1A, ii). Of note, t-SNE and UMAP present similar arrangements of cell clusters, but display dissimilar morphological variations in visualization [25]. UMAP provides more specific and concrete visualization as well as preserving the global structures of the data. Additionally, UMAP's branching clusters are able to highlight biological significances. In our own experience of single-cell data analysis, UMAP generates better results compared to t-SNE when the number of cells in scRNA-seq exceeds 30,000 (Figure 1B). To cope with high cellular throughput in recent single-cell studies, UMAP has taken the prior place of t-SNE to reduce dimensionality and to visualize the data structure [26,27].

Cell-cell clustering and cell type identification

Among the analytical methodologies of scRNA-seq data, clustering is equally important as dimensionality reduction. Within the lower dimensional space, most informative genes, *e.g.*, highly variable genes, were selected for further analysis [28]. Subsequently, clustering is the key step in categorizing cells and defining cell types (Figure 1A, iii), and this step requires thorough considerations of both biological and computational aspects. Many clustering methods are generic that can be applied to any dataset that possesses the measurement of distance/similarity between data nodes [29]. There are three major measurements for cell-cell distance: Euclidean distance, Pearson correlation coefficient (PCC), and Spearman's rank correlation coefficient (SRCC) [30]. Euclidean distance represents the geometric similarities and amplifies the influence of highly variable genes, while the two latter measurements

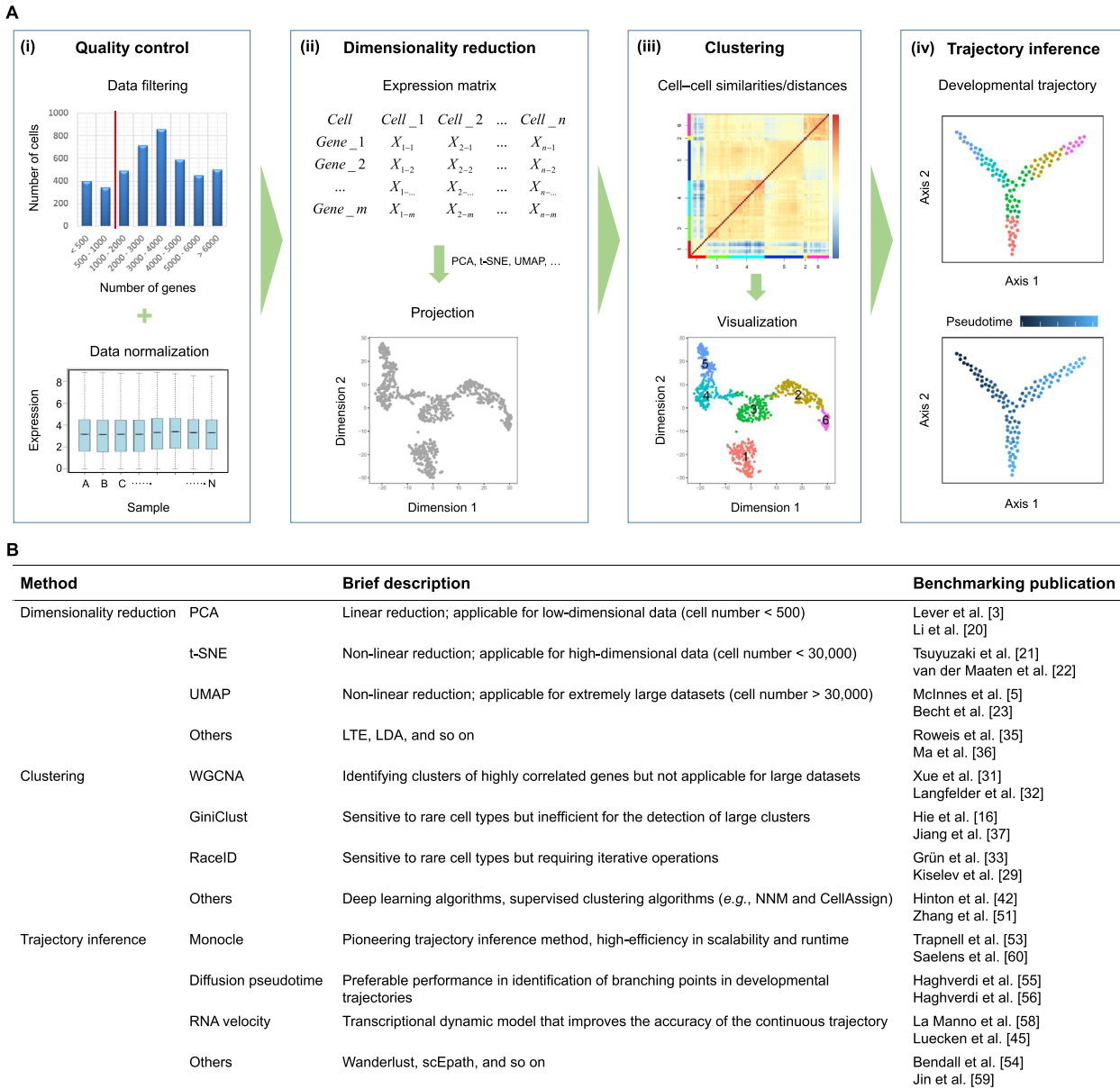


Figure 1 State-of-the-art analysis techniques

A. Overview of the workflow for the computational analysis of scRNA-seq data. (i) Quality control; (ii) dimensionality reduction; (iii) cell-cell clustering; (iv) trajectory inference. **B.** Characterization of the state-of-the-art analysis techniques. PCA, principal component analysis; t-SNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; WGCNA, weighted gene co-expression network analysis; RaceID, rare cell type identification; NNM, neural network model; LTE, locally linear embedding; LDA, linear discriminant analysis; scEpath, single-cell energy path.

consider the scale invariance, in other words, they reflect the relative differences in values. For example, along with scRNA-seq technology, a popular clustering algorithm is weighted gene co-expression network analysis (WGCNA) [31], which generates weighted networks based on the PCC values and combines similar cells into larger clusters or divides heterogeneous clusters into smaller groups. Although designed for bulk RNA-seq analysis, this approach can be used for finding modules (clusters) of highly correlated genes in single cells or pseudocells generated by aggregating single-cell clusters, characterizing such modules using the module eigengene or GRNs, and identifying candidate biomarkers or therapeutic

targets [32]. Grün et al. [33] sequenced the transcriptome of thousands of single cells isolated from mouse intestinal organoids, and by applying t-SNE analysis, they characterized major groups of cells. The unique element is that they incorporated an iterative algorithm to increase the detection sensitivity for rare cell populations. The optimized algorithm for rare cell type identification (RaceID) measures the Euclidean distance between single cells and enhances the sensitivity for detecting signature genes. In recent studies, RaceID has shown high accuracy in discovering rare and novel cell subtypes.

Single-cell analytical methodology is a fast-developing research field. Currently, a variety of integrative software

packages for single-cell data analysis have been developed, with the most representative ones, including Scran [7], Seurat [8], and Scanpy [34]. These packages involve a step-by-step workflow for analyzing scRNA-seq data, aiming to identify cell–cell heterogeneities from transcriptomic measurements. Besides the computational approaches we summarized above, new approaches are emerging with unique kernel algorithms for data analyzing [16,35–37] (Figure 1B). For example, single-cell regulatory network inference and clustering (SCE-NIC) [38] identifies co-expression modules, referring to as regulons, to construct the gene regulatory network [39,40]. Census is a regression-based method to normalize scRNA-seq expression levels and detect gene regulatory changes [41]. Deep learning approaches have also been introduced into single-cell study, and neural network model (NNM) is a powerful mathematical algorithm that is able to capture and represent complex variable relationships [42,43].

Although methodologies for single-cell analysis progress significantly, it is evident that the major challenge in analyzing scRNA-seq data is to understand the biological implication rather than computational outcome. The inference of cell states and functions, and the annotation of cell clusters require an in-depth understanding of the biological significance of each cluster [44]. In previous studies, researchers annotated cell types through unsupervised clustering with marker genes or projecting cells to existing cell atlases [24,45,46]. However, manual annotation is laborious, time-consuming, and error-prone due to lack of standard ontologies, especially for large datasets. To meet these challenges, supervised and semi-supervised clustering algorithms were developed [47]. In order to efficiently characterize cell types at the user end, SuperCT [48] and scPred [49] employ supervised training models to predict cell types. In comparison with unsupervised clustering methods, the supervised models showed notable performance in the prediction of cell subtypes. Representative semi-supervised clustering algorithms, Garnett [50] and CellAssign [51], are based on a paradigm of machine learning. Unlike unsupervised clustering, Garnett and CellAssign operate in an intuitive manner by incorporating prior knowledge of marker genes for interpretation and annotation of cell clusters. An additional advantage is that the analysis also expands the marker gene database for future application. With the increasing number of cell types being profiled, it would be useful to perform supervised and unsupervised clustering analyses simultaneously when analyzing new data to identify rare cell populations as well as transitional cell types during lineage differentiation.

Developmental trajectory and pseudotime

In many biological systems, cells display a changing spectrum of states, which may mirror the differentiation process especially in development [52]. To infer the lineage history of single cells, it necessitates interrogating the continuity of cell states in scRNA-seq data. A unique advantage of scRNA-seq is that researchers can recapitulate developmental stages in a single experiment, given enough single cells and cell statuses sampled. Analytical algorithms underpinned by the concept of pseudotime have been applied to infer the developmental trajectory signposted by the transition of cell states during differentiation (Figure 1A, iv).

Pioneering trajectory inference methods such as Monocle [53] and Wanderlust [54] have been instrumental in the devising of pseudotime methodology to mimic the developmental process. Utilizing diffusion-like random walks to evaluate transitions across cell states, diffusion pseudotime (DPT) shows favorable performance in identification of branching points in developmental trajectories [55,56]. RNA velocity employs a transcriptional dynamic model [57] that counts the unspliced and spliced mRNAs in scRNA-seq datasets [58] to enhance the fidelity of mapping the biological trajectory. Single-cell energy path (scEpath) calculates the entropy of dynamical process through statistical physics modeling, to predict the transition cell states and infer the lineage trajectories [59]. In essence, pseudotime displays an ordering of cell states by extracting the developmental information in the dataset, thereby allowing the identification of the cell types at early, intermediate, and end states of the pseudotime trajectory. From the cell orders, the gene regulatory networks that accompany the trajectory can be inferred [45]. While there are various methodologies to infer the lineage trajectory, no single method works best for all types of scRNA-seq data [6]. Different approaches have advantages and disadvantages in attributes, such as scalability, runtime, and resolution [60] (Figure 1B). Thus, the pseudotime methods for inferring cell trajectories should be employed iteratively with consideration of the alignment of biological context and computational output. Biological questions vary from one biological system to the others, and the choice of mathematical models is driven by the underlying assumptions for reconstructing lineage trajectories.

Spatial transcriptomics

Cell type heterogeneity and cell fate acquisition are influenced by the environmental conditions and the interactive signals perceived by the cells at a specific location in the biological entity. However, current scRNA-seq technologies require tissue dissociation, resulting inevitably in the loss of spatial information on tissue architecture and organization. The methodology for collating the spatially resolved transcriptome involves bringing together the imaging technology and the profiling of gene expression, starting from classic low-plex RNA or protein expression assays to multimodalities.

In situ spatial transcriptome analysis

In situ hybridization is a road-tested technique that is used for localization and detection of specific DNA or RNA sequences in cells in segmented tissue domains, or in defined positions in the whole tissue [61,62]. Leveraging the *in situ* patterns of spatially restricted genes as “landmarks”, individual cells with cell type-specific expression profiles identified by scRNA-seq analysis can be mapped to the inferred locations in the tissues [63] (Figure 2A, i), e.g., the mapping of single cells in the brain of the marine annelid *Platynereis dumerilii* [64] (the spatial mapping will be discussed in more detail later).

The *in situ*-based reconstruction method is most effective where the tissue structure is stably stereotypic for the biological entity to assure universal consistency of position mapping among samples. These methods are, however, constrained by the scope of prior knowledge of cell markers, and only a

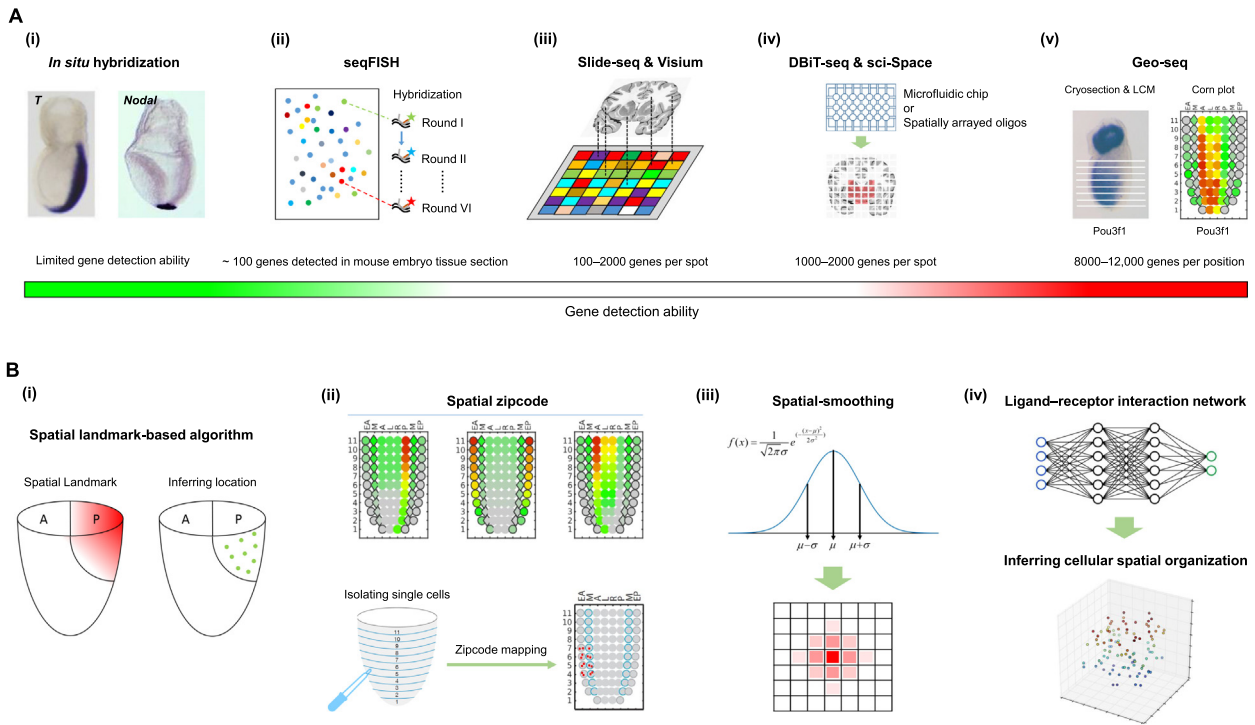


Figure 2 The advances of spatial transcriptomics

A. Single-cell resolved spatial transcriptome technologies. In addition to the classic *in situ* hybridization (i), pioneering work includes seqFISH (ii), Slide-seq/Visium (iii), DBiT-seq/sci-Space (iv), and Geo-seq (v) greatly improve the spatial sensitivity and data depth. Whole-mount *in situ* hybridization images in (i) were obtained from the eMouseAtlas Project (www.emouseatlas.org). **B.** Combining scRNA-seq with spatial transcriptomics. (i) Inferring cellular localization based on spatial landmarks; (ii) zipcode mapping algorithm; (iii) spatial-smoothing algorithm; (iv) ligand-receptor interaction encoding cellular spatial organization. seqFISH, sequential fluorescence *in situ* hybridization; Slide-seq, slide sequencing; DBiT-seq, deterministic barcoding in tissue for spatial omics sequencing; Geo-seq, geographical positional sequencing; LCM, laser capture microdissection; A, anterior epiblast; P, posterior epiblast; L, left lateral epiblast; R, right lateral epiblast; M, mesoderm; EA, anterior endoderm; EP, posterior endoderm.

customized set of markers can be assayed concurrently to establish high-confidence landmarks [9], which therefore greatly reduce the robustness of position mapping. To improve the detection throughput, recently developed methods of seqFISH [11] and MERFISH [12] use sequencing-by-hybridization techniques for transcriptomic quantification. For these techniques, cells fixed in space are subject to repeated hybridizations of fluorescently labeled DNA probes and rounds of imaging for registering the position of the *in situ* hybridization signals of up to hundreds of genes (Figure 2A, ii). While achieving an enhanced throughput compared to conventional *in situ* hybridization, the sensitivity of marker detection remains less than scRNA-seq.

Recently, the seqFISH technique is further enhanced by introducing pseudocolor and spectral overlapping [65], and this study reported the quantification of over 10,000 transcripts in individual cells. Multiplexed fluorescence *in situ* hybridization could image hundreds to thousands of genes simultaneously; however, this technology is constrained by the coverage of the marker gene sets, the availability of sensitive probes, and the requirement of high-resolution microscopy. Additionally, with increasing level of multiplexing, molecular markers starting to overlap and tight associations between cells may result in signal crowding, both leading to false-positive signals [66].

Array-based spatial transcriptome analysis

To obtain a high-throughput, genome-wide profile of gene expression with spatial locations at cellular resolution, the latest spatial transcriptomic technology combines traditional histology technologies with massive throughput RNA-seq. For example, Slide-seq is a methodology which involves transferring mRNA from a tissue section onto a surface covered with DNA barcoded beads anchored at pre-defined positions, where the barcode registers the location of the captured mRNA [13] (Figure 2A, iii). Applying Slide-seq, cell types with characteristic gene expression profiles were identified at specific positions in the mouse cerebellum and hippocampus, thereby revealing the spatial pattern of gene expression of cells in the Purkinje layer and the temporal progression of cell type-specific responses to traumatic brain injury. Compared to the imaging-based methodologies, the density of sampling for Slide-seq analysis significantly improves the spatial resolution. However, the on-slide reaction is not uniformly efficient due to variations of cellular structures. Slide-seq detects 100–300 unique molecular identifiers (UMIs) per cell, which are equivalent to ~100 genes [67]. The impact of sparsity of data signals over noise is yet unsolved, and the number of genes that can be simultaneously assayed is still limited. Slide-seq V2 has

improved library generation and bead synthesis to increase RNA capture efficiency [68]. While the gene detection ability has been enhanced (5–10-fold of Slide-seq), the low-throughput has confounded the efficiency of cell type clustering.

Originated from a similar platform as Slide-seq, 10X Genomics Visium Spatial Gene Expression Solution (Visium) captures the transcriptome from cells at their positions in an intact tissue section to generate a spatial display of the gene activity (Figure 2A, iii). Using the Visium technology, the spatial topography of gene expression of the human dorsolateral prefrontal cortex (DLPFC) has identified laminae-enriched expression signatures of the six-layered DLPFC, and refined the molecular connectivity to previous laminar markers [69]. Compared with Slide-seq, Visium has an improved throughput and enables gene expression analysis at the transcriptome level on intact sections of tissues. Nevertheless, this technology is still discommoded by unsatisfactory gene detection ability (500–2000 genes per spatial spot) and incompatibility to all tissue types, meanwhile, the limited spatial resolution also makes it insufficient to elucidate cell-to-cell heterogeneity [70].

In recent spatial transcriptomics studies, a number of outstanding methodologies and improvements have been developed that enable high-throughput *in situ* profiling of gene expression. Additional array-based platforms, such as high-definition spatial transcriptomics (HDST) [71] and spatial enhanced resolution omics-sequencing (Stereo-seq) [72] offer high-resolution and high-density coverage, by refining the spatial resolution to 2 μm subcellular range, though the limited gene detection ability related to the sequencing depth remains unresolved.

Direct spatial barcoding captures tissue coordinates

Pioneering attempts toward spatial transcriptome analysis were based on either multiplexed fluorescence *in situ* hybridization or barcoded bead array [73,74]. These methodologies are still technically demanding, requiring complicated processing steps and hard pressed by the low mRNA capture efficiency. A nascent technology takes the approach of deterministic barcoding in tissue for spatial omics sequencing (DBiT-seq) [75], instead of transferring mRNAs onto a solid-phase substrate. DBiT-seq places microfluidic chips with perpendicular channels against a fixed tissue section to directly barcode the biomolecules (Figure 2A, iv). In this regard, DBiT-seq obviates the need for sophisticated processes of sequential hybridization or bead decoding by directly capturing the approximate coordinates of the cells upon sequencing. Conceptually similar to DBiT-seq, sci-Space method [76] employs the spatially arrayed unique combination of oligonucleotides (oligos) that are transferred to nuclei of cells by diffusion. The nuclei are then extracted from the tissue section and prepared for scRNA-seq (Figure 2A, iv). The spatial coordinates of large-scale regions can be retrieved by sci-Space, which meets a need of extended coverage that cannot be fulfilled by other methodologies [13].

Both DBiT-seq and sci-Space have strengths and limitations. DBiT-seq greatly improves spatial resolution close to single-cell level, but does not resolve single cells directly. High-resolution immunofluorescence on the same tissue section simultaneously can facilitate cell segmentation. sci-Space

focuses on high-throughput screens and can cover large tissue domains, but the spatial resolution ($\sim 200 \mu\text{m}$) is limiting. Decreasing spot size and increasing spot density may refine resolution, but the gene detection ability would be compromised. The capturing rates (1000–2000 genes per spot) of both technologies would benefit from further optimization.

Laser capture microdissection records geographical location

Spatial locations of cells in tissues strongly influence cellular functions. A key prerequisite for single cell research is the efficient harvesting of single cells of interest from a tissue or cell group. Laser capture microdissection (LCM) enables capture of target cells with structural and spatial information retained [77,78], but the resolution remains relatively low. The limitation lies in the number of cells captured (ranging from hundreds to thousands for library preparation) and the coverage of profiling limiting to a smaller gene set, making it difficult to discern the heterogeneity and cell-to-cell interactions within a sampled population.

Geographical positional sequencing (Geo-seq), a technique combining scRNA-seq and LCM, was developed [14]. This technique features optimized preservation tissue morphology and enhanced RNA integrity to enable low input RNA-seq analysis, making it possible to analyze the transcriptome of limited materials of as few as ten cells sampled from defined locations in the tissue. Using Smart-seq2 for full-length scRNA-seq [79], the gene detection efficiency has significantly increased to detect up to 8000–12,000 genes in each LCM sample. The depth of sequencing data enables the elucidation of gene structure based on splicing information. Incorporating the spatial information enables the construction of the 3D spatial and quantitative transcriptome atlas [15] (Figure 2A, v). Through bioinformatics analysis, Geo-seq data can be mined to infer the gene regulatory networks that underpin the molecular architectures of embryonic development from preimplantation to gastrulation in mice [39].

Geo-seq is potentially applicable to investigate a variety of biological questions, such as cell fate determination, cell type heterogeneity, and gene regulatory network. By applying Geo-seq, a spatial transcriptome map of early post-implantation mouse embryos utilizing data collated from sections along their proximal–distal axis has revealed that podocalyxin exocytosis is associated with the polarization of extra-embryonic tissue and the formation of pro-amniotic cavity [80]. Geo-seq analysis of the caudal hematopoietic tissue (CHT) of zebrafish has unveiled the spatio-temporal dynamics of the expansion of hematopoietic stem cell and progenitor cell in CHT and the accompanying gene regulatory networks, which points to a strong association between hematopoietic stem cell differentiation and cell cycle status [81].

Combining scRNA-seq with spatial transcriptomics

Other than the “direct” *in situ* single-cell analytical techniques (seqFISH, MERFISH, DBiT-seq, *etc.*), leveraging the attributes of scRNA-seq and spatial transcriptomics to retrieve from the integrated datasets the spatial positions of single cells offers another avenue for single-cell spatial transcriptomics study [9]. For example, a core feature of Seurat package is inferring cell location by integrating scRNA-seq data with

the spatial pattern of *in situ* RNA expression [63]. Seurat first divides the embryo or tissue structure into high-resolution bins followed by binarization of the expression of *in situ* hybridization signals of specific landmark genes. As a result, the landmarks are normalized as “on” or “off” within each spatial bin and the reference positioning system is thereby constructed. By aligning the expression of landmark genes of single cells to the reference gene set, the locations of single cells can be determined (Figure 2B, i). However, the limited number of spatial landmark genes and the different platforms for generating the references and the single-cell data may impact the precision of spatial mapping. In the latest version of Seurat, through identification of pairwise correspondences between single cells, termed “anchors”, across diverse datasets, a shared space that improves the precision of inferring single-cell locations, can be constructed [8]. Similarly, the distributed mapping (DistMap) scores [82], by binning the entire *Drosophila* embryo into refined expression regions and computing the Matthews correlation coefficients (MCCs) between the expression values of spatial landmarks for every cell–region combination, can predict multiple likely positions for each cell, instead of assigning to a single location. The higher the resolution of the reference positioning system, the higher is the precision of inferred localization. Employing high content seqFISH data as a reference system increases the number of spatial-specific markers compared to conventional *in situ* hybridization. For instance, by integrating scRNA-seq data of “Gastrulation Atlas” [24] and quantified seqFISH data, a spatially resolved map of gene expression at quasi single-cell resolution can be constructed for the mouse embryo [83].

One drawback of the landmark-based algorithms is the limited availability of known landmarks. To mitigate this shortcoming, Geo-seq incorporates a spatial signature gene (zipcode) mapping protocol that computationally allocate single cells to spatial positions (Figure 2B, ii). The unbiased identification of “zipcodes” embedded in Geo-seq data enhances the imputation of the positions of single cells. Over 90% of the single cells isolated from known positions can be mapped correctly with high degrees of precision to the original position in the embryo [15]. By applying zipcode mapping protocol, stem cells in culture or cells from germ layers of a gastrulating embryo can be mapped to their best-fit positions in the embryo [84,85]. By incorporating a spatial-smoothing algorithm, the mapping efficiency and spatial confidence intervals can be further enhanced. Gaussian distribution-based spatial simulation that mirrors the spatial dispersity of single-cell mapping (Figure 2B, iii) showed that the simulation of diffused domain has a significantly improved fidelity compared with other spatial mapping algorithms (Wang and Jing, unpublished). The methodologies for combining scRNA-seq data and spatial transcriptomics are still in its infancy. Efficient extraction of both spatial coordinates and spatial signature/zipcode genes continues to be a challenging task for computational biology [86]. Mathematical modeling may also be introduced to gain a deeper level of understanding of the spatial information.

Mathematical models for spatial organization

Besides experimentally recording the spatial information of individual cells, computational biologists have attempted to reconstruct spatial map of single cells mathematically. The

core idea is to use a reference atlas of signature genes as an “anchor” to assign spatial coordinates to each cell [63,64,82]. However, such methodologies rely heavily on the coverage of spatial marker genes, even with a specific combination of reference genes which may not be adequate to tag every spatial position. To resolve the cell positions, a *de novo* spatial reconstruction (novoSpaRc) algorithm was applied to *de novo* construct the cartography of single-cell gene expression, without any reliance on prior knowledge [87]. The assumption for novoSpaRc is that neighboring cells display more similar transcriptional identities than cells farther apart and the physical distance between cells increases with their biological distance in gene expression space. Applying novoSpaRc on published single-cell datasets, cells of mammalian intestinal epithelia [88] and liver lobules [89] can be re-positioned. Consistent with the notion of structural correspondence, classified cell types showed distinct pairwise distances.

Similar to the spatial–distance approach of novoSpaRc, the cellular spatial organization mapper (CSOmap) algorithm assumes that ligand–receptor interaction information encodes the cellular spatial organization [90], for inferring *de novo* cellular interactions from scRNA-seq data. CSOmap is built on the hypothesis that ligand–receptor interactions mediate cellular self-assembly rather than the transcriptome profiles (Figure 2B, iv). CSOmap suggests that transcriptome-similar cells often, but not always, have similar spatial locations [91,92]. Decoding intercellular communication networks in the perspective of ligand–receptor interactions have important biological implication [93]. Recently, a growing number of methodologies that model how gene expression of a cell is influenced by interacting cells are developed. Based on cellular communication models, methods such as CellPhoneDB, NicheNet, SingleCellSignalR, and iTALK [94–97] can predict ligand–receptor links between interacting cells by integrating their expression profiles with prior knowledge on gene regulatory networks and signaling activities. Although these methodologies did not implement the analysis of the space-resolved reorganization of single cells, they open new opportunities to dissect the complexity, heterogeneity, and molecular dynamics of cell–cell communication. Further development of the spatial algorithms of cellular communication models is anticipated in future studies.

Mathematical modeling provides important insights into spatial transcriptomics, but the pseudo-space may not depict the complexity of biological contexts. For example, different cell types may require higher-order regulatory networks or exhibit distinct principles of spatial organization. Incorporating biological insights into mathematical models, and in combination with algorithms of machine learning [98], will be a priority of methodology development.

Outlook and challenges

Single-cell analytical techniques that enable high-throughput profiling of multi-omics have revolutionized molecular cell biology. There is a compelling need to efficiently analyze the large-scale single-cell omics data for thousands to millions of cells. On one hand, the ability to unravel biological complexity is empowered by the accessibility to big data. On the other hand, higher computing power is needed to raise the ability to cluster cell populations comprehensively and identify rare

cell types [29]. Visualizing and deciphering the results continues to be testing. Classic linear methods, such as hierarchical clustering and PCA, are insufficient to capture relationships between cells in large datasets accurately due to the high levels of data variation. Deep learning and non-linear techniques are more flexible (t-SNE, UMAP, *etc.*), as they can identify all variations, extract significant features, and visualize the structure of very large datasets aesthetically.

With the advances of single-cell analytical methodologies, another challenge is the choosing of suitable analytic methodology for the single-cell data. The most commonly used methods, such as Scraper [7], Seurat [8], and Scanpy [34], each with the unique kernel algorithm may have an impact on the outcome. As these analytical methods offer many parameters that can be customized to define the criteria of normalization and resolution of clustering, the selection of parameters could have an unintended effect on the outcome [99]. Thus, it is important to inform computational analysis and data interpretation with biological guidance, particularly in the clustering and definition of cell types.

Time-series scRNA-seq experiments provide voluminous information on the gene sets that are dynamically activated, and their role in cellular interaction and development. Given the complexity of dynamic biological processes, a discrete clustering or classification of cells may not fully mirror the full spectrum of cell states. Pseudotime study of time-series data will be critical in future studies. Current pseudotime methods, such as Monocle [53] and DPT [55], simulate a biological process such as cell differentiation by ordering cells along a model of developmental trajectory [100]. This allows the identification of intermediate cell states and the branching points of the trajectory, besides inferring the beginning and end points. Innovation in pseudotime methodology, for revealing time-series activity of transcription factors (TFs) [58] and the gene regulatory networks anchored by these TFs, as well as the modulation of signaling activity, is crucial for understanding the dynamics of the molecular function [101].

A pressing question in biological research is “what defines a cell type and cellular functionality?” [102]. Previous single-cell analyses often try to define a cell type using marker genes or transcriptome identity [103,104]. However, cell fate decision and cellular heterogeneity can be influenced by the environmental factors and the signals in the neighborhood of the cell [39]. Most of the single-cell studies lack the spatial information of cells. To answer the question “what is a cell type?”, knowledge of the environmental input to guide cell differentiation and lineage development is a vital prerequisite in single-cell research [105]. How to computationally retrieve the original spatial coordinates of individual cells requires an integrated approach of spatial transcriptome and precise gene detection at single-cell resolution [39]. Once generated, the spatially resolved single-cell maps will facilitate the computational inference of gene regulation networks and signaling networks. To address the question of “cellular functionality”, single-cell multi-omics data beyond the transcriptome have to be collated and be integrated with the spatio-temporal information into a biomolecular cell atlas that would inform the functional output of the cells in development and homeostasis. Additionally, when the mathematical models can be complemented by experimental validation, single-cell technologies would elevate the dimensionality of our fundamental understanding of cell and developmental biology [10].

Competing interests

The authors have declared no competing interests.

CRediT authorship contribution statement

Ran Wang: Conceptualization, Investigation, Writing – original draft. **Guangdun Peng:** Writing – review & editing. **Patrick P.L. Tam:** Writing – review & editing. **Naihe Jing:** Supervision, Writing – review & editing. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported in part by the National Key Basic Research and Development Program of China (Grant Nos. 2019YFA0801402, 2018YFA0107200, 2018YFA0801402, 2018YFA0800100, 2018YFA0108000, and 2017YFA0102700), the “Strategic Priority Research Program” of the Chinese Academy of Sciences (Grant Nos. XDA16020501 and XDA16020404), the National Natural Science Foundation of China (Grant Nos. 31630043 and 31900573), and the China Postdoctoral Science Foundation Grant (Grant No. 2018M642106).

ORCID

ORCID 0000-0003-2644-6337 (Ran Wang)
ORCID 0000-0002-8586-0637 (Guangdun Peng)
ORCID 0000-0001-6950-8388 (Patrick P.L. Tam)
ORCID 0000-0003-1509-6378 (Naihe Jing)

References

- [1] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;6:377–82.
- [2] Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;566:496–502.
- [3] Lever J, Krzywinski M, Altman N. Principal component analysis. *Nat Methods* 2017;14:641–2.
- [4] Bushati N, Smith J, Briscoe J, Watkins C. An intuitive graphical visualization technique for the interrogation of transcriptome data. *Nucleic Acids Res* 2011;39:7380–9.
- [5] McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. *J Open Source Softw* 2018;3:861.
- [6] Tritschler S, Büttner M, Fischer DS, Lange M, Bergen V, Lickert H, et al. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* 2019;146:dev170506.
- [7] Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 2016;5:2122.
- [8] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck 3rd WM, et al. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–902.
- [9] Peng G, Cui G, Ke J, Jing N. Using single-cell and spatial transcriptomes to understand stem cell lineage specification during early embryo development. *Annu Rev Genomics Hum Genet* 2020;21:163–81.

- [10] Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* 2017;541:331–8.
- [11] Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell *in situ* RNA profiling by sequential hybridization. *Nat Methods* 2014;11:360–1.
- [12] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;348:aaa6090.
- [13] Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;363:1463–7.
- [14] Chen J, Suo S, Tam PP, Han JDJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat Protoc* 2017;12:566–80.
- [15] Peng G, Suo S, Chen J, Chen W, Liu C, Yu F, et al. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev Cell* 2016;36:681–97.
- [16] Hie B, Peters J, Nyquist SK, Shalek AK, Berger B, Bryson B. Computational methods for single-cell RNA sequencing. *Annu Rev Biomed Data Sci* 2020;3:339–64.
- [17] Kulkarni A, Anderson AG, Merullo DP, Konopka G. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr Opin Biotechnol* 2019;58:129–36.
- [18] Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 2016;17:29.
- [19] Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 2020;21:12.
- [20] Li Y, Wang R, Qiao N, Peng G, Zhang K, Tang K, et al. Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. *J Biol Chem* 2017;292:19590–604.
- [21] Tsuyuzaki K, Sato H, Sato K, Nikaido I. Benchmarking principal component analysis for large-scale single-cell RNA-sequencing. *Genome Biol* 2020;21:9.
- [22] van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9:2579–605.
- [23] Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2019;37:38–44.
- [24] Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 2019;566:490–5.
- [25] Wu D, Sheng JYP, Su-En GT, Chevrier M, Hua JJJ, Hon TLK, et al. Comparison between UMAP and t-SNE for multiplex-immunofluorescence derived single-cell data from tissue sections. *bioRxiv* 2019;549659.
- [26] Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. *Science* 2020;370:aba7721.
- [27] Park JE, Botting RA, Conde CD, Popescu DM, Lavaert M, Kunz DJ, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* 2020;367:eaay3224.
- [28] Zhang Z, Cui F, Wang C, Zhao L, Zou Q. Goals and approaches for each processing step for single-cell RNA sequencing data. *Brief Bioinform* 2021;22:bbaa314.
- [29] Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet* 2019;20:273–82.
- [30] Kim T, Chen IR, Lin Y, Wang AYY, Yang JYH, Yang P. Impact of similarity metrics on single-cell RNA-seq data clustering. *Brief Bioinform* 2019;20:2316–26.
- [31] Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 2013;500:593–7.
- [32] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- [33] Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 2015;525:251–5.
- [34] Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;19:15.
- [35] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* 2000;290:2323–6.
- [36] Ma C, Zhang L, He T, Cao H, Ren X, Ma C, et al. Single cell Raman spectroscopy to identify different stages of proliferating human hepatocytes for cell therapy. *Stem Cell Res Ther* 2021;12:555.
- [37] Jiang L, Chen H, Pinello L, Yuan GC. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol* 2016;17:144.
- [38] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083–6.
- [39] Peng G, Suo S, Cui G, Yu F, Wang R, Chen J, et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* 2019;572:528–32.
- [40] Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft L, et al. A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell* 2018;174:982–98.
- [41] Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* 2017;14:309–15.
- [42] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313:504–7.
- [43] Lin C, Jain S, Kim H, Bar-Joseph Z. Using neural networks for reducing the dimensions of single-cell RNA-seq data. *Nucleic Acids Res* 2017;45:e156.
- [44] Lee JTH, Hemberg M. Supervised clustering for single-cell analysis. *Nat Methods* 2019;16:965–6.
- [45] Lueken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 2019;15:e8746.
- [46] Pasquini G, Arias JER, Schäfer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J* 2021;19:961–9.
- [47] Huang Q, Liu Y, Du Y, Garmire LX. Evaluation of cell type annotation R packages on single-cell RNA-seq data. *Genomics Proteomics Bioinformatics* 2021;19:267–81.
- [48] Xie P, Gao M, Wang C, Zhang J, Noel P, Yang C, et al. SuperCT: a supervised-learning framework for enhanced characterization of single-cell transcriptomic profiles. *Nucleic Acids Res* 2019;47:e48.
- [49] Alquicira-Hernandez J, Sathe A, Ji HP, Nguyen Q, Powell JE. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol* 2019;20:264.
- [50] Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods* 2019;16:983–6.
- [51] Zhang AW, O'Flanagan C, Chavez EA, Lim JLP, Ceglia N, McPherson A, et al. Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat Methods* 2019;16:1007–15.
- [52] Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol Syst Biol* 2018;14:e8046.
- [53] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
- [54] Bendall SC, Davis KL, Amir EAD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 2014;157:714–25.

- [55] Haghverdi L, Büttner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 2015;31:2989–98.
- [56] Haghverdi L, Büttner M, Wolf FA, Büttner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* 2016;13:845–8.
- [57] Zeisel A, Köstler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, et al. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol Syst Biol* 2011;7:529.
- [58] Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature* 2018;560:494–8.
- [59] Jin S, MacLean AL, Peng T, Nie Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 2018;34:2077–86.
- [60] Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019;37:547–54.
- [61] Jensen E. Technical review: *in situ* hybridization. *Anat Rec (Hoboken)* 2014;297:1349–53.
- [62] Armit C, Richardson L, Venkataraman S, Graham L, Burton N, Hill B, et al. eMouseAtlas: an atlas-based resource for understanding mammalian embryogenesis. *Dev Biol* 2017;423:1–11.
- [63] Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502.
- [64] Achim K, Pettit JB, Saraiva LR, Gavriouchkina D, Larsson T, Arendt D, et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* 2015;33:503–9.
- [65] Eng CHL, Shah S, Thomassie J, Cai L. Profiling the transcriptome with RNA SPOTS. *Nat Methods* 2017;14:1153–5.
- [66] Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* 2017;358:64–9.
- [67] Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat Methods* 2018;15:539–42.
- [68] Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Bella DJD, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2021;39:313–9.
- [69] Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 2021;24:425–36.
- [70] Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol* 2022;40:517–26.
- [71] Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat Methods* 2019;16:987–90.
- [72] Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;185:1777–92.
- [73] Pichon X, Lagha M, Mueller F, Bertrand E. A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol Cell* 2018;71:468–80.
- [74] Trcek T, Lionnet T, Shroff H, Lehmann R. mRNA quantification using single-molecule FISH in *Drosophila* embryos. *Nat Protoc* 2017;12:1326–48.
- [75] Liu Y, Yang M, Deng Y, Su G, Enniful A, Guo CC, et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 2020;183:1665–81.
- [76] Srivatsan SR, Regier MC, Barkan E, Franks JM, Packer JS, Grosjean P, et al. Embryo-scale, single-cell spatial transcriptomics. *Science* 2021;373:111–7.
- [77] Morrison JA, Bailey CM, Kulesa PM. Gene profiling in the avian embryo using laser capture microdissection and RT-qPCR. *Cold Spring Harb Protoc* 2012;2012:1249–62.
- [78] Cañas RA, Canales J, Gómez-Maldonado J, Ávila C, Cánovas FM. Transcriptome analysis in maritime pine using laser capture microdissection and 454 pyrosequencing. *Tree Physiol* 2014;34:1278–88.
- [79] Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 2013;10:1096–8.
- [80] Christodoulou N, Kyprianou C, Weberling A, Wang R, Cui G, Peng G, et al. Sequential formation and resolution of multiple rosettes drive embryo remodelling after implantation. *Nat Cell Biol* 2018;20:1278–89.
- [81] Xue Y, Liu D, Cui G, Ding Y, Ai D, Gao S, et al. A 3D atlas of hematopoietic stem and progenitor cell expansion by multi-dimensional RNA-seq analysis. *Cell Rep* 2019;27:1567–78.
- [82] Karaïskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, et al. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 2017;358:194–9.
- [83] Lohoff T, Ghazanfar S, Missarova A, Koulina N, Pierson N, Griffiths JA, et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat Biotechnol* 2022;40:74–85.
- [84] Sozen B, Amadei G, Cox A, Wang R, Na E, Czukiewska S, et al. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat Cell Biol* 2018;20:979–89.
- [85] Osteil P, Studdert JB, Goh HN, Wilkie EE, Fan X, Khoo PL, et al. Dynamics of Wnt activity on the acquisition of ectoderm potency in epiblast stem cells. *Development* 2019;146:dev172858.
- [86] Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 2021;22:627–44.
- [87] Nitzan M, Karaïskos N, Friedman N, Rajewsky N. Gene expression cartography. *Nature* 2019;576:132–7.
- [88] Moor AE, Harnik Y, Ben-Moshe S, Massasa EE, Rozenberg M, Eilam R, et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* 2018;175:1156–67.
- [89] Halpern KB, Shenhav R, Matcovitch-Natan O, Toth B, Lemze D, Golan M, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 2017;542:352–6.
- [90] Ren X, Zhong G, Zhang Q, Zhang L, Sun Y, Zhang Z. Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res* 2020;30:763–78.
- [91] Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* 2016;3:346–60.
- [92] Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 2017;171:1611–24.
- [93] Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cell PhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* 2020;15:1484–506.
- [94] Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* 2020;48:e55.
- [95] Wang Y, Wang R, Zhang S, Song S, Jiang C, Han G, et al. iTALK: an R package to characterize and illustrate intercellular communication. *bioRxiv* 2019;507871.

- [96] Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* 2020;17:159–62.
- [97] Toda S, Blauch LR, Tang SKY, Morsut L, Lim WA. Programming self-organizing multicellular structures with synthetic cell–cell signaling. *Science* 2018;361:156–62.
- [98] Petegrosso R, Li Z, Kuang R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. *Brief Bioinform* 2020;21:1209–23.
- [99] Vieth B, Parekh S, Ziegenhain C, Enard W, Hellmann I. A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat Commun* 2019;10:4667.
- [100] Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, et al. A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* 2019;365:eaax1971.
- [101] Yeger-Lotem E, Riva L, Su LJ, Gitler AD, Cashikar AG, King OD, et al. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat Genet* 2009;41:316–23.
- [102] Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet* 2019;20:257–72.
- [103] Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* 2019;364:685–9.
- [104] Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, et al. Construction of a human cell landscape at single-cell level. *Nature* 2020;581:303–9.
- [105] Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* 2013;41:D996–D1008.