8

Computational Approaches in Spatial Transcriptomics for the Study of Mammalian Spermatogenesis

Deina Bossa, Melanie Evans, Shreya Rajachandran, Xin Zhang, Qiqi Cao, and Haiqi Chen

Abstract

Spermatogenesis is a complex and dynamic cellular differentiation process critical to male fertility. Although the full continuum of gene expression patterns from spermatogonial stem cells (SSCs) to spermatozoa in steady state was characterized using single-cell RNA sequencing technologies, the transcriptional dynamics of spermatogenesis within its native tissue context was largely unexplored. The recent development of spatial transcriptomics (ST) technologies has transformed male fertility research from a single-cell level to a two-dimensional spatial coordinate system and facilitated the study of spermatogenesis in the native environment of both the rodent and human testes. The spatial gene expression information generated by these ST technologies requires new computational approaches to extract novel biological insights. These requirements include, but are not limited to, spatial mapping of testicular cell types, identifying spatially variable genes, and understanding the molecular cross-talk between testicular cell types. Here, we review computational approaches that have been used to dissect mammalian spermatogenesis in the context of ST. We also highlight new computational approaches that can be leveraged to reveal novel insights into male fertility.

Keywords

Testis · Spermatogenesis · Spatial Transcriptomics · Spatially variable genes · Cell-cell communication

1 Introduction

Male fertility relies upon proper germ cell proliferation and differentiation within the seminiferous tubules of the testis to facilitate the constant production of sperm. Spermatogenesis begins with the self-renewal and differentiation of spermatogonial stem cells (SSCs) and ends with the generation of spermatozoa [1, 2]. This process is driven by the dynamic coordination between germ cells and their surrounding somatic cells [3, 4]. Such coordination is difficult to investigate using current single-cell sequencing technologies such as single-cell RNA sequencing (scRNA-

D. Bossa · M. Evans
Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

S. Rajachandran · X. Zhang · Q. Cao · H. Chen (✉)
Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA

Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA
e-mail: haiqi.chen@utsouthwestern.edu

C. Y. Cheng, F. Sun (eds.), *Molecular Male Reproductive Medicine*, Advances in Experimental Medicine and Biology 1469, https://doi.org/10.1007/978-3-031-82990-1_8

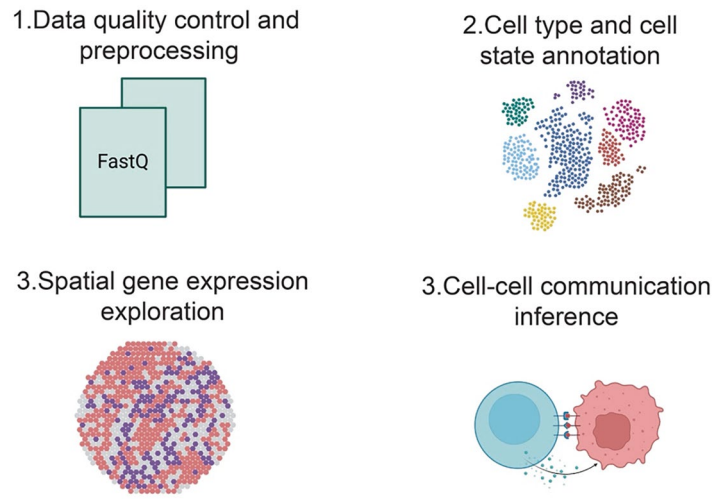

Fig. 8.1 Four critical topics related to ST data analysis and interpretation

seq). This is because these technologies require tissue disassociation, therefore disrupting the native tissue context. In contrast, recent breakthroughs in spatial transcriptomics (ST) technologies have made it possible to examine reproductive processes, including spermatogenesis, within the native tissue context (see a recent review [5] for details). The data generated by ST technologies is an entirely new type which requires specialized data analytical solutions. Thus, ST-specific algorithms that utilize spatial information must be developed to investigate and interpret gene expression data from a spatial perspective. This book chapter focuses on four critical topics related to ST data analysis and interpretation in the context of mammalian spermatogenesis (Fig. 8.1): (1) ST data quality control and preprocessing, including data quality assessment, filtering, and quality improvement; (2) single-cell- and tissue-level annotations in ST data; (3) tissue-wide gene expression exploration using ST data; and (4) inference of cell-cell communications from ST data.

2 ST Data Quality Control and Preprocessing

Current ST technologies can be primarily categorized into two classes based on their design principles (for reviews, see [5–7]). The first class relies on the imaging of predetermined mRNA targets such as in situ hybridization (ISH)- and in situ sequencing (ISS)-based methods. For example, an in situ RNA sequencing protocol has recently been established specifically for testicular samples [8]. This protocol can be divided into three parts (Fig. 8.2): (i) sample pretreatment & library preparation. A set of testicular genes are targeted with a custom-designed barcoded padlock probe library. Probes that have specifically interacted with the targeted transcripts are amplified by rolling circle amplification (RCA) reaction using phi29 polymerases, and the specific amplification generates signals with a high signal-to-noise ratio in testis cross sections; (ii) in situ barcode sequencing reaction & imaging. Specific barcode sequence on the padlock probes for each targeted gene is decoded by sequential hybridization/ligation reaction and imaging cycles. The signals are detected by fluorescently labeled oligonucleotide libraries. Automatic confocal imaging through multiple fields of view enables high throughput; and (iii) image analysis. After image acquisition, images are processed with an analysis pipeline tailored for testicular samples. The pipeline includes signal deconvolution, image registration, and image segmentation. A x-y coordinates map of each individual mRNA transcript and the gene count matrix for each cell in the tissue section are generated, allowing accurate reconstruction of the spatial distribution of

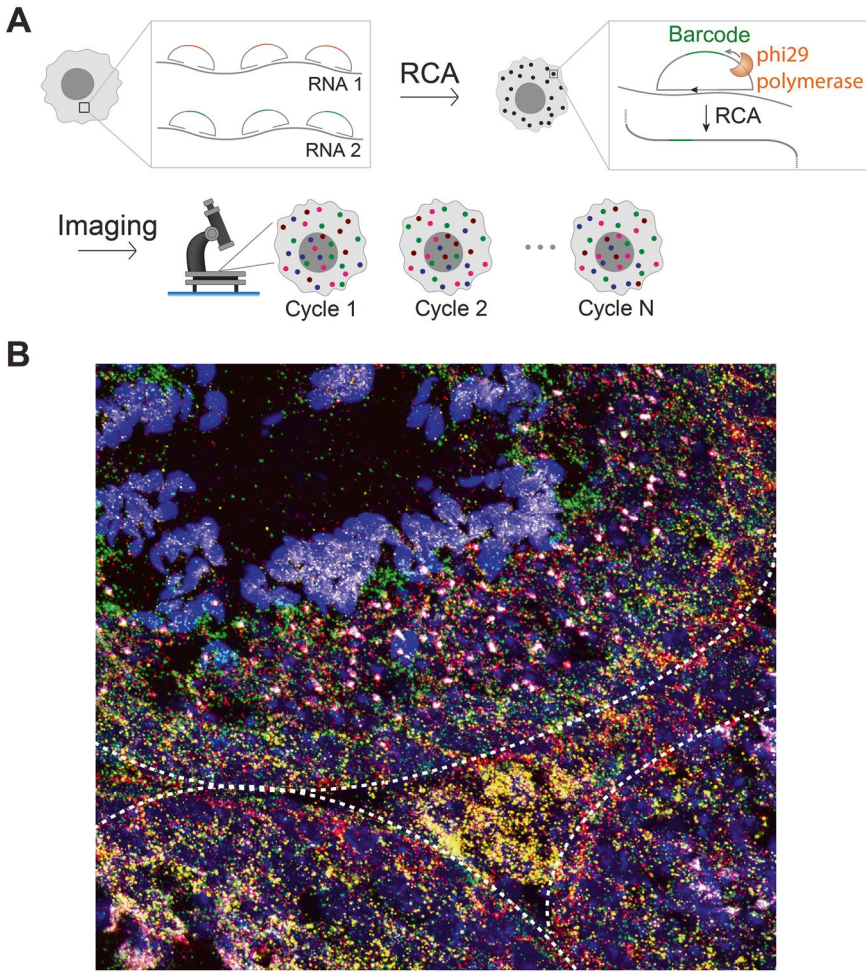

Fig. 8.2 In situ RNA sequencing. (**a**) Schematic of the in situ RNA sequencing workflow. (**b**) An example in situ RNA sequencing image of a mouse testis cross section. White dashed lines mark the basement membranes of the seminiferous tubules.

mRNA transcripts within the testicular tissue section.

The second class of ST technologies includes unbiased approaches that build on spatial isolation/capture of RNA molecules using spatially indexed arrays followed by next-generation sequencing (NGS) (Fig. 8.3). For example, a high-resolution ST approach called Slide-seqV2 was employed to generate the first spatial transcriptome atlas of the mouse and human testis [8]. The Slide-seqV2 technology relies on the in situ capture of testicular mRNA transcripts by an array of poly(dT)-containing, spatially barcoded beads [9]. Each bead is also called a "spot" and is 10 μm in diameter. Through a streamlined computational workflow, the identity of each mRNA species and their counts and spatial distributions can be inferred from the sequencing data with high accuracy.

Despite a variety of designs of the ST technologies, the resulting ST data all take the form of two matrices: (1) a gene expression matrix that records the mRNA abundance for each cell or "spot" and (2) a spatial coordinate matrix that records the spatial location (x-y) of each cell or spot. In imaging-based ST technologies as well

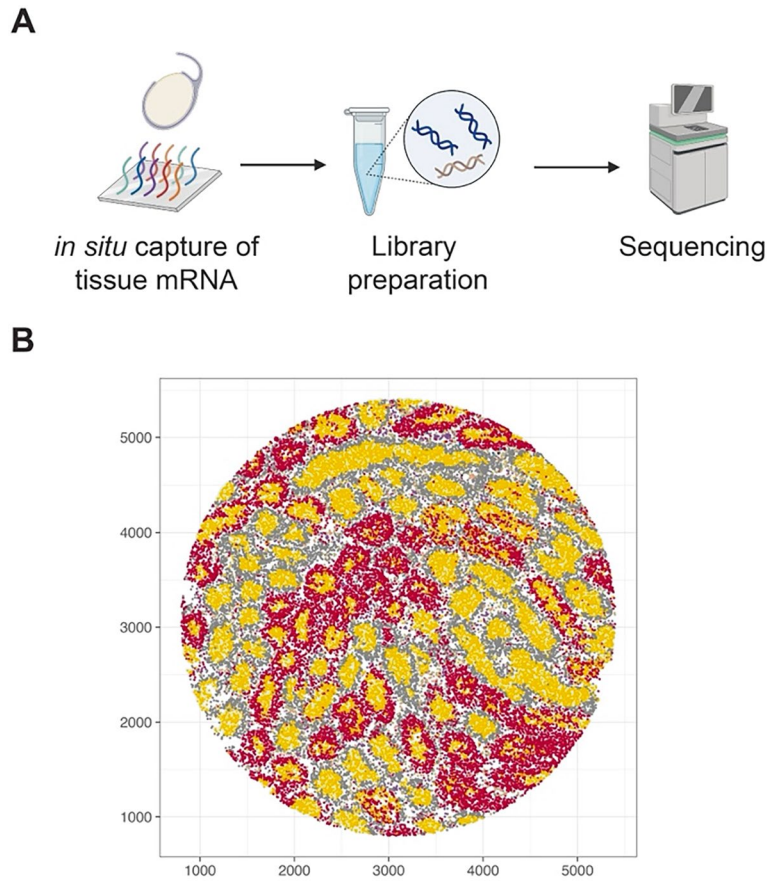

Fig. 8.3 Array-based ST technologies. (**a**) Schematic of the Slide-seqV2 workflow. (**b**) An example of a mouse testis cross-section image digitally reconstructed using the Slide-seqV2 data

as the array-based approaches such as 10X Genomics Visium, the morphology of the profiled tissue section is also recorded. Similar to scRNA-seq data, the amount of mRNA transcripts captured by the ST technologies only accounts for a small portion of the total mRNA in the tissue due to a low mRNA capture efficiency. Moreover, some genes known to be present in the sample may be missing entirely (i.e., gene dropouts). In addition, gene diffusion caused by the experiment and the inner batch effect is a common problem in ST data generated by array-based ST technologies. Therefore, it is necessary to perform data filtering, normalization, and data imputation to improve the data quality for subsequent analysis.

2.1 Data Filtering

For the data generated by array-based ST technologies, it is necessary to determine which spots are covered by tissues and contain real transcripts. Some arbitrary screening can be performed to retain these spots, such as by screening the number of unique molecular identifiers (UMIs) to filter out spots with a low mRNA capture rate. Low-quality cells or spots are further filtered based on the number of genes captured in each cell or spot (low-quality cells or spots contain very few genes) and the percentage of mitochondrial genes (low-quality or dying cells often exhibit extensive mitochondrial contamination). Genes detected in a few cells or spots (usually less than three) are also removed.

2.2 Data Normalization

Normalization helps to reduce the negative impact of technical noises and bias in sequencing depth on downstream analyses. Currently, most ST data are normalized using tools that are made for scRNA-seq data. For example, the testis Slide-seqV2 data was normalized using an approach called SCTransform. SCTransform [10] fits a regularized negative binomial model to raw gene count data. The residuals of this model can be used as normalized and variance-stable values. However, depending on the ST technology used, the distribution and sparsity of the resulting ST data may differ from those of scRNA-seq data. Therefore, various scRNA-seq data normalization algorithms may need to be tested on the ST data. The spatial information of the ST data may be leveraged to enable the normalization of spatial expression. For example, the stSME algorithm from stLearn incorporates spatial location, gene expression, and morphological similarity to normalize gene expression data [11].

2.3 Data Imputation

Imputation aims to leverage the existing gene expression data to predict the expression of unmeasured genes. It is usually performed on data generated by array-based ST technologies because the data are relatively sparse compared to the data generated by other ST technologies. For instance, adaptively thresholded low-rank approximation (ALRA), a method for imputation of scRNA-seq data, has recently been employed to impute testis Slide-seqV2 data [12]. ALRA takes advantage of the nonnegativity and low-rank structure of the gene expression matrix to selectively impute technical zero values while preserving biological zeros [13]. Other approaches, such as linked inference of genomic experimental relationships (LIGER) [14] and Seurat [15] perform imputation by joint embedding a scRNA-seq reference dataset of the same tissue with the ST data. They impute the unmeasured genes of ST data based on the connection between cells in the two datasets. Recent adaptations of machine learning techniques such as stPLUS [16] impute missing genes in ST data by learning the gene expression information from a reference scRNA-seq dataset of the tissue in question. However, none of these methods take advantage of the spatial and histology information of the tissue which may provide additional assistance for gene imputation.

3 Cell-Type Annotations

Similar to the scRNA-seq data analysis, cell-type annotation provides the foundation for downstream ST data interpretation, including the study of tissue organization and cell-cell communications (CCCs).

For imaging-based ST technologies, cell-type annotation can be achieved by clustering analysis on gene expression features followed by identifying marker genes that match prior biological knowledge in digitally segmented cells. Thus, three computational procedures are keys to the cell-type annotation of imaging-based ST data: segmentation, clustering, and marker gene assignment. First, numerous methods are available to delineate cell boundaries in fluorescent images. For example, in a recent study, CellProfiler was used to segment cells from an in situ sequencing dataset of a mouse testis section [8]. However, given the various sizes and shapes of testicular cells, the accurate delineation of testicular cell boundaries remains challenging. Emerging machine learning-based approaches such as Cellpose [17] may overcome this challenge. Second, to cluster the segmented cells, the clustering approaches designed for scRNA-seq data, such as those implemented in Seurat [15] and SCANPY [18], are usually used. These approaches leverage the Leiden or Louvain clustering on a set of feature genes selected by performing dimensionality reduction with methods such as the principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), or uniform manifold approximation and projection (UMAP) on gene expression data. However, these approaches do not consider the

spatial information of gene expression. Recent spatial clustering work combines spatial distance information or tissue morphological information with gene expression profiles. For example, BayesSpace builds on a Bayesian model to identify spatial clusters with similar expressions [19]. SpaGCN [20] and SEDR [21] use an autoencoder to integrate the spatial distance and gene expression for a low-dimensional embedding of the gene expression data for clustering. Finally, after clustering, cells in the same cluster are considered to be the same cell type. Prior biological knowledge can be leveraged to manually assign cell type identities based on the marker genes extracted from differentially expressed genes (DEGs) in each cell cluster. However, there are caveats associated with this approach. For example, some marker genes may be expressed in more than one cluster. A combination of marker genes may be required to distinguish two cell types from the same origin. Moreover, prior studies may provide conflicting information on the choice of marker genes for a given cell type. Therefore, relevant domain knowledge is crucial when choosing marker genes for the cell-type annotation of a specific tissue type.

For array-based ST technologies such as Slide-seqV2 and 10X Genomics Visium, the cell-type annotation faces a different set of challenges. In these approaches, tissue mRNAs are captured by patterned spots, and each spot is considered the minimum processing unit (i.e., a pseudo-cell) even though one spot may capture transcripts from two or more "real" cells in a tissue slice. This spatial misalignment between the array spots and cells in the tissue requires the deconvolution of the mRNA signals captured by each spot. Current ST data deconvolution methods estimate the composition of mRNA signals within each spot mostly by transferring cell-type signatures defined by a reference scRNA-seq dataset of a matching tissue type to the ST data. For instance, robust cell-type decomposition (RCTD) estimates the cell-type proportion of signals in each spot by treating the ST data of each spot as a linear combination of the transcriptome of different cell types measured by scRNA-seq [22]. RCTD was recently used to assign cell-type information in the mouse and human testis Slide-seqV2 data [12]. The results matched the known spatial localization of major testicular cell types. Besides RCTD, other emerging deconvolution methods may also be applied to the testis ST data. For example, SPOTlight employs a seeded non-negative matrix factorization regression and reference scRNA-seq data to infer the location of cell types and states within a tissue [23]. Tangram, a deep-learning framework, learns a spatial alignment of scRNA-seq data from a paired ST data [24]. In Cell2location, cell-type signatures are first extracted from a scRNA-seq reference dataset and then used to decompose the mRNA counts in the ST data using a hierarchical Bayesian model [25]. More recently, CytoSPACE allows for the mapping of individual cells from a reference scRNA-seq dataset to spatial locations in an ST dataset by formulating the single-cell/spot assignment as a convex optimization problem and solving this problem using the Jonker–Volgenant shortest augmenting path algorithm [26]. Redeconve goes beyond cell-type-level deconvolution to map fine-grained cell states by treating each cell of the reference scRNA-seq data as a distinct cell state reference [27]. It is worth noting that annotations of cell types in ST data can only be as good as those of the reference scRNA-seq datasets. Therefore, it is challenging to identify new cell types or cell states that are not already present in the scRNA-seq datasets.

4 Tissue-Wide Gene Expression Exploration

One of the key advances of ST technologies compared to scRNA-seq technologies is the ability to capture the spatial distribution of gene expression within intact tissue slices or even within individual cells. These genes, termed spatially variable genes (SVGs), are similar to the DEGs proposed in scRNA-seq studies. The main difference is that SVGs are detected to study the spatial expression patterns since genes with nonrandom spatial expression patterns may exert specialized functions in tissues or cells, while DEGs are calculated to define the specificity of a certain cluster

compared to other clusters. A plethora of computational approaches have been developed to detect SVGs using ST data. For example, SPARK identifies SVGs based on nonparametric Gaussian process (GP) regression [28]. It has been recently applied to testis Slide-seqV2 data to identify SVGs in individual seminiferous tubules [8]. Recent machine learning-based methods may also be applied to the testis Slide-seqV2 data. For instance, sepal identifies SVGs by assuming that genes with spatial patterns will demonstrate a lower degree of randomness (diffusions) and a higher degree of structure. Therefore, compared to genes with a uniform pattern across different spatial locations, SVGs require more iterations for the gradient algorithm to converge. And a long convergence time of the system is indicative of a structured spatial pattern [29]. GLISS, on the other hand, is a graph-based method that constructs a mutual nearest neighborhood graph from spatial coordinates and relies on a graph-based feature selection procedure to select SVGs [30].

It's worth noting that the abovementioned methods are mostly geared toward the identification of global spatial patterns in gene expression without taking into account the cell-type information. It has been challenging to identify cell-type-specific SVGs for array-based spatial transcriptomics technologies, such as 10x Genomics Visium and Slide-seqV2, as they can capture multiple cell types on individual spots. The presence of cell-type mixtures complicates the estimation of cell-type-specific SVGs because different cell types have different gene expression profiles, independent of spatial locations. Even for single cell-level, imaging-based spatial transcriptomics technologies, such as MERFISH [31, 32] and seqFISH [33], diffusion of reagents or imperfect cellular segmentation may still cause mixing across cell types. Not accounting for cell-type proportions leads to biased estimates of differential gene expression at different spatial locations due to cell-type proportion changes or contamination from other cell types. To this end, cell-type-specific inference of differential expression (C-SIDE) was developed. C-SIDE models gene expression as an additive mixture across cell types of log-linear cell-type-specific expression functions. It takes as input one or more biologically relevant covariates, such as spatial position or cell-type colocalization, and identifies genes, for each cell type, that significantly change their expression as a function of these covariates [34]. Similarly, Niche-DE was recently developed to identify local gene expression differences of a cell type in the presence of another cell type from ST data [35]. Different from C-SIDE, however, niche-DE adjusts for cell-type-specific library size and conducts hierarchical FDR control [35].

5 Inference of Cell-Cell Communications (CCCs)

CCC is a key aspect of cell and tissue biology that mediates cellular characteristics. Cell location affects CCCs and associated gene expression partly because molecules mediating CCCs form concentration gradients as they diffuse from their producing cells and trigger different signaling programs in receiver cells. For this reason, it is important to account for the spatial context of cells to understand how tissues function. For example, in the testis, the self-renewal of SSCs requires glial cell line-derived neurotrophic factor (GDNF) secreted from neighboring somatic cells such as the Sertoli cells and peritubular myoid cells [36]. Additional factors such as fibroblast growth factor 2 (FGF2) and retinoic acid (RA) secreted from surrounding somatic cells are also required for SSC self-renewal or differentiation [37, 38]. ST technologies offer a unique opportunity to dissect CCCs since the physical distance between cells is recorded. A dominant mechanism of CCCs is ligand–receptor (LR) interactions, in which ligands from one cell type bind to their cognate receptors in other cell types. Thus, most computational strategies are focused on LR interactions as a function of cell-cell distance. For example, NICHES has been recently applied to the mouse and human testis Slide-seqV2 data to characterize the LR interactions between SSCs and their surrounding somatic cells [12]. NICHES takes ST data as input and

computes the strength of each LR interaction by multiplying ligand expression on the sending cell with receptor expression on the adjacent receiving cell [39]. Similarly, on the cell-type level, the strength of communications between cell types is created by multiplying mean ligand expression from one cell type with the receptor expression on another cell type [39].

Other available methods may also be adapted to decipher the CCCs in the testis. For instance, Giotto ranks LR pairs based on a so-called CCC score to identify interactions between adjacent cells [40]. The CCC score is calculated by first computing a combined co-expression score for each LR pair for all cells of two interacting cell types. This co-expression score is then compared with a background distribution of co-expression scores based on spatial permutations. The resulting adjusted p-value and log2 fold change are used to generate the CCC score [40]. SpaTalk spatially constrains the LR interaction analysis by using the Euclidean distance between cells to build a cell graph network based on their K-nearest neighbors. It then computes CCCs of connected cells in the graph network [41]. Similarly, CellPhoneDB v3 restricts cell-cell interactions to cell clusters in the same microenvironment defined based on spatial information [42]. Recently, optimal transport-based methods have been developed to dissect CCCs. For example, SpaOTsc first constructs a spatial metric for a scRNA-seq dataset by integrating it with paired ST data using optimal transport. It then reconstructs CCC networks and identifies intercellular regulatory relationships between genes [43]. Another method, COMMOT, introduces an algorithm called collective optimal transport which sets spatial distance constraints on CCCs and enables the transport of multispecies distributions (ligands) to multispecies distributions (receptors) to account for multispecies interactions [44].

Together, these various methods highlight the diverse ways in which spatial information can be explicitly accounted for in CCC analyses. However, it remains unclear which metric represents the best way to evaluate the CCCs inferred from different methods. Therefore, multiple methods may be tested on the same ST data to look for consensus CCCs. It is also important to note that the strength of LR interactions inferred from ST data may be interpreted with caution as the transcriptome is not the equivalent of the proteome. Posttranslational modifications such as glycosylation have further effects on protein interactions, especially in LR binding [45], which is not captured in the ST data.

6 Conclusions and Outlook

This review highlights four topics related to ST data analysis for the understanding of mammalian spermatogenesis. Computational approaches designed to improve ST data quality are expected to eliminate technical noise, bias in sequencing depth, and the batch effect. Methods for cell-type identification and annotation can then be applied to map the spatial distribution of testicular cell types. Tissue-wide data exploration approaches enable the identification of spatially variable genes across seminiferous tubules and the dissection of cell-cell communication networks between various testicular cell types. Furthermore, new developments in computational approaches will continue to place spatial omics in a position to reveal novel insights into mammalian spermatogenesis and beyond.

First, algorithms for ST data analysis are starting to model cell fate trajectory. The so-called pseudotime analysis, which has been used extensively on scRNA-seq data to understand phenomena such as cell differentiation and cancer progression [46], can now incorporate spatial structure information of a tissue. For example, in stLearn, the (re)construction of spatiotemporal trajectories is done by taking a linear combination of nonspatial diffusion pseudotime and spatial distance [11]. Such a measure of spatial pseudotime represents a combined distance in physical and gene expression space. It is anticipated that more computational approaches will be developed to accurately model development and disease progression with ST data.

Second, while most current ST studies use individual tissue slices, an increased number of

efforts have been made to reconstruct tissue structure in three dimensions (3D) by performing ST analysis on consecutive tissue slices. These efforts require algorithms that can spatially align each pair of adjacent slices and then construct a stacked 3D architecture of the tissue. For example, the Scube algorithm from SPACEL builds a mutual nearest neighbor (MNN) graph between the cells/spots of two adjacent slices based on the spatial coordinate information of the cells/spots and constructs an alignment objective function between them. This function serves as the foundation for a unique global optimization strategy for 3D alignment [47]. Following 3D reconstruction, future methods may also enable cell fate trajectory inference and CCC characterization in the 3D space.

Finally, multiomics integration, previously implemented for single-cell omics data [48, 49], may be extended to ST data. Recent research has begun to explore this. For example, a recent study performed simultaneous single-cell spatial profiling of mRNA and chromatin accessibility, along with CNV inference in human melanoma samples, facilitating the understanding of the role of epigenetic regulation in the tumor microenvironment [50]. Future developments in generating multiomics, multidimensional, and multitime-series data may not only provide opportunities for novel discoveries but also create a further need for developing computational approaches for data integration.

References

1. Chen, H., et al.: Human spermatogenesis and its regulation. In: Winters, S.J., Huhtaniemi, I.T. (eds.) Male Hypogonadism: Basic, Clinical and Therapeutic Principles, pp. 49–72. Springer International Publishing, Cham (2017)
2. Hale, D.W.: Mammalian spermatogenesis. In: Verma, R.S. (ed.) Advances in Genome Biology, pp. 249–304. Amsterdam, The Netherlands, JAI (1996)
3. Saunders, P.T.: Germ cell-somatic cell interactions during spermatogenesis. Reprod. Suppl. **61**, 91–101 (2003)
4. Fritz, I.B.: Somatic cell-germ cell relationships in mammalian testes during development and spermatogenesis. Ciba Found. Symp. **182**, 271–274 (1994).; discussion 274–81
5. Zhang, X., et al.: Dissecting mammalian reproduction with spatial transcriptomics. Hum. Reprod. Update. **29**(6), 794–810 (2023)
6. Tian, L., Chen, F., Macosko, E.Z.: The expanding vistas of spatial transcriptomics. Nat. Biotechnol. **41**(6), 773–782 (2023)
7. Moses, L., Pachter, L.: Museum of spatial transcriptomics. Nat. Methods. **19**(5), 534–546 (2022)
8. Chen, H., et al.: Dissecting mammalian spermatogenesis using spatial transcriptomics. Cell Rep. **37**(5), 109915 (2021)
9. Stickels, R.R., et al.: Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat. Biotechnol. **39**(3), 313–319 (2021)
10. Hafemeister, C., Satija, R.: Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. **20**(1), 296 (2019)
11. Pham, D., et al.: Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. Nat. Commun. **14**(1), 7739 (2023)
12. Rajachandran, S., et al.: Dissecting the spermatogonial stem cell niche using spatial transcriptomics. Cell Rep. **42**(7), 112737 (2023)
13. Linderman, G.C., et al.: Zero-preserving imputation of single-cell RNA-seq data. Nat. Commun. **13**(1), 192 (2022)
14. Welch, J.D., et al.: Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell. **177**(7), 1873–1887 e17 (2019)
15. Stuart, T., et al.: Comprehensive integration of single-cell data. Cell. **177**(7), 1888–1902 e21 (2019)
16. Shengquan, C., et al.: stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. Bioinformatics. **37**(Suppl_1), i299–i307 (2021)
17. Pachitariu, M., Stringer, C.: Cellpose 2.0: how to train your own model. Nat. Methods. **19**(12), 1634–1641 (2022)
18. Wolf, F.A., Angerer, P., Theis, F.J.: SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. **19**(1), 15 (2018)
19. Zhao, E., et al.: Spatial transcriptomics at subspot resolution with BayesSpace. Nat. Biotechnol. **39**(11), 1375–1384 (2021)
20. Hu, J., et al.: SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat. Methods. **18**(11), 1342–1351 (2021)
21. Xu, H., et al.: Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med. **16**(1), 12 (2024)
22. Cable, D.M., et al.: Robust decomposition of cell type mixtures in spatial transcriptomics. Nat. Biotechnol. **40**(4), 517–526 (2022)
23. Elosua-Bayes, M., et al.: SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res. **49**(9), e50 (2021)

24. Biancalani, T., et al.: Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat. Methods. **18**(11), 1352–1362 (2021)
25. Kleshchevnikov, V., et al.: Cell2location maps fine-grained cell types in spatial transcriptomics. Nat. Biotechnol. **40**(5), 661–671 (2022)
26. Vahid, M.R., et al.: High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. Nat. Biotechnol. **41**(11), 1543–1548 (2023)
27. Zhou, Z., et al.: Spatial transcriptomics deconvolution at single-cell resolution using Redeconve. Nat. Commun. **14**(1), 7930 (2023)
28. Sun, S., Zhu, J., Zhou, X.: Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat. Methods. **17**(2), 193–200 (2020)
29. Andersson, A., Lundeberg, J.: Sepal: identifying transcript profiles with spatial patterns by diffusion-based modeling. Bioinformatics. **37**(17), 2644–2650 (2021)
30. Zhu, J., Sabatti, C.: Integrative spatial single-cell analysis with graph-based feature learning. bioRxiv, 2020: p. 2020.08.12.248971
31. Xia, C., et al.: Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl. Acad. Sci. USA. **116**(39), 19490–19499 (2019)
32. Wang, G., Moffitt, J.R., Zhuang, X.: Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. Sci. Rep. **8**(1), 4847 (2018)
33. Eng, C.L., et al.: Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature. **568**(7751), 235–239 (2019)
34. Cable, D.M., et al.: Cell type-specific inference of differential expression in spatial transcriptomics. Nat. Methods. **19**(9), 1076–1087 (2022)
35. Mason, K., et al.: Niche-DE: niche-differential gene expression analysis in spatial transcriptomics data identifies context-dependent cell-cell interactions. Genome Biol. **25**(1), 14 (2024)
36. Meng, X., et al.: Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. Science. **287**(5457), 1489–1493 (2000)
37. Takashima, S., et al.: Functional differences between GDNF-dependent and FGF2-dependent mouse spermatogonial stem cell self-renewal. Stem Cell Rep. **4**(3), 489–502 (2015)
38. Busada, J.T., Geyer, C.B.: The role of retinoic acid (RA) in spermatogonial differentiation. Biol. Reprod. **94**(1), 10 (2016)
39. Raredon, M.S.B., et al.: Comprehensive visualization of cell-cell interactions in single-cell and spatial transcriptomics with NICHES. Bioinformatics. **39**(1), btac775 (2023)
40. Dries, R., et al.: Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol. **22**(1), 78 (2021)
41. Shao, X., et al.: Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. Nat. Commun. **13**(1), 4429 (2022)
42. Garcia-Alonso, L., et al.: Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nat. Genet. **53**(12), 1698–1711 (2021)
43. Cang, Z., Nie, Q.: Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat. Commun. **11**(1), 2084 (2020)
44. Cang, Z., et al.: Screening cell-cell communication in spatial transcriptomics via collective optimal transport. Nat. Methods. **20**(2), 218–228 (2023)
45. Castelli, M., et al.: How aberrant N-glycosylation can alter protein functionality and ligand binding: an atomistic view. Structure. **31**(8), 987–1004.e8 (2023)
46. Tritschler, S., et al.: Concepts and limitations for learning developmental trajectories from single cell genomics. Development. **146**(12), dev170506 (2019)
47. Xu, H., et al.: SPACEL: deep learning-based characterization of spatial transcriptome architectures. Nat. Commun. **14**(1), 7603 (2023)
48. Adossa, N., et al.: Computational strategies for single-cell multi-omics integration. Comput. Struct. Biotechnol. J. **19**, 2588–2596 (2021)
49. Rautenstrauch, P., et al.: Intricacies of single-cell multi-omics data integration. Trends Genet. **38**(2), 128–139 (2022)
50. Russell, A.J.C., et al.: Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. Nature. **625**(7993), 101–109 (2024)