



Principles of Spatial Transcriptomics Analysis: A Practical Walk-Through in Kidney Tissue

Teia Noel¹, Qingbo S. Wang^{2,3,4,5*}, Anna Greka^{1,6*} and Jamie L. Marshall^{1*}

¹ Kidney Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, United States, ² Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, United States, ³ Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, United States, ⁴ Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, United States, ⁵ Department of Statistical Genetics, Graduate School of Medicine, Osaka University, Osaka, Japan, ⁶ Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

OPEN ACCESS

Edited by:

Bruce Molitoris,
Indiana University, United States

Reviewed by:

Christoph Kuppe,
RWTH Aachen University, Germany

*Correspondence:

Qingbo S. Wang
qingbow@broadinstitute.org

Anna Greka
agreka@broadinstitute.org

Jamie L. Marshall
jmarshall@broadinstitute.org

Specialty section:

This article was submitted to
Renal and Epithelial Physiology,
a section of the journal
Frontiers in Physiology

Received: 04 November 2021

Accepted: 26 November 2021

Published: 06 January 2022

Citation:

Noel T, Wang QS, Greka A and
Marshall JL (2022) Principles
of Spatial Transcriptomics Analysis:
A Practical Walk-Through in Kidney
Tissue. *Front. Physiol.* 12:809346.
doi: 10.3389/fphys.2021.809346

Spatial transcriptomic technologies capture genome-wide readouts across biological tissue space. Moreover, recent advances in this technology, including Slide-seqV2, have achieved spatial transcriptomic data collection at a near-single cell resolution. To-date, a repertoire of computational tools has been developed to discern cell type classes given the transcriptomic profiles of tissue coordinates. Upon applying these tools, we can explore the spatial patterns of distinct cell types and characterize how genes are spatially expressed within different cell type contexts. The kidney is one organ whose function relies upon spatially defined structures consisting of distinct cellular makeup. Thus, the application of Slide-seqV2 to kidney tissue has enabled us to elucidate spatially characteristic cellular and genetic profiles at a scale that remains largely unexplored. Here, we review spatial transcriptomic technologies, as well as computational approaches for cell type mapping and spatial cell type and transcriptomic characterizations. We take kidney tissue as an example to demonstrate how the technologies are applied, while considering the nuances of this architecturally complex tissue.

Keywords: slide-seqV2, spatial transcriptomics, kidney spatial transcriptomics, slide-seq, kidney transcriptomics

APPLICATIONS OF SPATIAL TRANSCRIPTOMICS

Unbiased spatial transcriptomics adds *in situ* spatial context to single cell RNA data providing a powerful tool to characterize the spatial location for whole transcriptome sequencing (10x Genomics, 2007; Chen et al., 2015; Stegle et al., 2015; Ståhl et al., 2016; Rodriques et al., 2019; Liao et al., 2021; Longo et al., 2021; Nature Methods, 2021; Stickels et al., 2021). This is accomplished with two technologies, Spatial Transcriptomics (ST or Visium) and Slide-seqV2 (10x Genomics, 2007; Ståhl et al., 2016; Rodriques et al., 2019; Stickels et al., 2021). Other technologies, such as MERFISH (targeted), GeoMxTM (targeted), DBiT-seq, and Stereo-seq (BGI), allow for higher resolution, in some cases even subcellular, detection of RNA (Chen et al., 2015, 2021; Liu et al., 2020; Zollinger et al., 2020; Su et al., 2021). However, they tend to be more complex to implement. Both ST and Slide-seqV2 use uniquely barcoded beads with Oligo(dT) to capture polyadenylated RNA. ST has a larger feature size (avg 1–10 cells per spot, 50 μm beads with 100 μm spacing between beads (10x Genomics, 2007) than Slide-seqV2 [avg 1–3 cells per spot, 10 μm beads (Rodriques et al., 2019; Stickels et al., 2021)] and thus has relatively limited resolution compared to the near single cell resolution in Slide-seqV2.

On the other hand, ST allows for profiling of a large area of tissue (typically the entire cross section) while a single Slide-seqV2 array only covers a 3 mm diameter of tissue so multiple arrays on serial cryosections are needed to cover the entire tissue cross section. The other advantage of ST is the ability to co-stain the same tissue slice from which the spatial transcriptome is captured with hematoxylin and eosin staining (H&E) for histology or targeted antibodies. Slide-seqV2 and ST capture similar numbers of UMIs across the same spot area (Stickels et al., 2021).

Human and mouse kidneys across health and disease have been profiled with both ST and Slide-seqV2 (Raghubar et al., 2020; Lake et al., 2021; Melo Ferreira et al., 2021). Raghubar et al. (2020) used ST to examine spatial transcriptomic differences between sex and species of human and mouse kidney tissue revealing differences in gene expression correlating with male versus female kidneys and human versus mouse kidneys. Raghubar et al. (2020) used ST to characterize spatial transcriptomics and in particular immune cell clusters in human kidney tissue and mouse models of kidney tissue subjected to ischemia/reperfusion injury and cecal ligation puncture (Melo Ferreira et al., 2021). Raghubar et al. (2020) used both ST and Slide-seqV2 to spatially define altered injury states and a fibrotic niche in human kidney biopsies from healthy, acute kidney injury (AKI), and chronic kidney disease (CKD) participants (Lake et al., 2021).

Our group used Slide-seqV2 to develop a spatial transcriptomic atlas of human and mouse kidney tissue in health and disease (Marshall et al., 2021). We profiled two mouse models of disease, early diabetic kidney disease (DKD, BTBR *ob/ob*) and uromodulin autosomal dominant tubulointerstitial kidney disease (ADTKD, UMOD-KI). The study also contains nine human participants with both cortex and medulla biopsies, one with early DKD and one with injury due to sustained tumor compression. Slide-seqV2 revealed the spatial location of *LYVE1*+ macrophages in human medulla with injury due to sustained tumor compression. In mice with early DKD, we revealed changes in the cellular organization of spatially restricted glomeruli. In UMOD-KI mice, we identified the spatial locations of diseased fibroblasts, macrophages, and *Trem2* + macrophages as well as an upregulation of the unfolded protein response (UPR) pathway in thick ascending limb (TAL) tubules. These results altogether demonstrated the utility of spatial transcriptomics technologies combined with downstream computational analysis to uncover previously unknown human and mouse kidney physiology. Throughout this review, we will discuss such computational approaches for cell type mapping as well as spatial cell type and transcriptomic characterizations focusing on kidney tissue.

PREPROCESSING OF SPATIAL TRANSCRIPTOMICS DATA

The Slide-seqV2 reads are first aligned and mapped to the human or mouse reference transcriptome using tools such as STAR aligner (Dobin et al., 2013). After standard quality control protocols, including filtering based on the number of genes and

UMIs per bead, the Slide-seqV2 data is turned into an expression matrix where rows correspond to beads and columns to genes. Additionally, per bead, native spatial coordinates are preserved across 2-D tissue space.

MAPPING THE CELL TYPE FROM scRNA-SEQ DATA

The first step in the analysis of Slide-seqV2 data is to assign cell type identities to each bead. Accurate cell type classification is aided by external scRNA-seq datasets from published materials, where cell types are already annotated (Halpern et al., 2017; Lake et al., 2021; Subramanian et al., 2021). Different computational techniques have been developed to perform such analysis. First, we will focus on a method that was originally developed (Rodrigues et al., 2019), NMFreg. Since typical scRNA-seq data is high dimensional (e.g., >10 thousands cells \times >20 thousands genes, where each cell is annotated with specific known cell type), the scRNA-seq expression matrix is projected to a low-dimensional basis of factors by selection of highly variable genes followed by NMF (Lee and Seung, 1999). Choosing the number of dimensions of the low-dimensional space is not trivial; In the Slide-seq paper (Rodrigues et al., 2019), the authors evaluated different numbers of dimensions k to semi-manually assign a value to the parameter (they showed that the biological interpretation is roughly consistent across different k s). Every factor is then assigned a unique cell type. To do so, for every cell in the scRNA-seq data, the method computes a loading for all factors, and assigns the cell the factor of maximum weight. Since each cell was previously annotated with a cell type identity, the cell type distribution for the cells assigned to each factor can be calculated. Every factor is then assigned the cell type of maximum count.

Having acquired a set of “metagene” factors for each cell type category, the method next utilizes this set of features to map Slide-seqV2 beads to cell types. Each bead’s gene expression profile is approximated as a non-negative sum of the factors using non-negative least square regression (NNLS). Since each factor can be mapped to a unique cell type, the factor loading for each bead can be converted to cell type loading. The cell type with highest load is selected to define the cell type identity for each bead, and the beads that do not have a single dimension with clearly highest load (e.g., beads having uniform load over all the dimensions, or having >1 dimensions with $\gg 0$ load) are filtered out. More recent developments of spatial transcriptomics cell type mapping techniques will be discussed in the next chapter.

CELL TYPE IDENTIFICATION WITH SEURAT TRANSFER LEARNING

A second approach for assigning cell type identities to beads is Seurat’s transfer learning method (Stuart et al., 2019). Similar to NMFreg, an scRNA-seq data set is utilized to elucidate cell type labels in a spatial transcriptomics query data set. First, the data of both the reference and the query are projected to

a shared lower-dimensional space. Anchors are then defined to map pairs of similar data points between the reference and the query, based on the amount of shared overlap between their respective neighborhoods. Next, weights are allocated to each anchor in order to emphasize mappings between similar biological states. The weights are based on the following criteria: (1) The distance between the query bead and k nearest reference neighbors and (2) The number of mutual nearest neighbors between the query bead and reference cell. Finally, cell type annotations from scRNA-seq are transferred along these anchors to the spatial transcriptomics reference using a weighted voter classifier (Kuncheva and Rodríguez, 2014). As a result of either the NMFreg or Seurat transfer learning methodologies, not only is each bead informed by a gene expression profile and spatial location, but also its overarching cell type identity. In Marshall et al. (2021), Seurat transfer learning presented higher consistency with physiologically known structures than NMFreg for all cell types in human tissue and in collecting ducts (CD), vascular smooth muscle cells (vSMC), and distal tubules in mouse tissue. An overview of mapped cell types in mouse and human tissue is shown in **Figure 1**.

DECOMPOSING SPATIAL TRANSCRIPTOMIC SPOTS INTO MULTIPLE CELL TYPES

One caveat of many spatial transcriptomics methods, including Slide-seqV2, is that data points are not necessarily representative of single cells. Even in near single-cell resolution spatial transcriptomic technologies such as Slide-seqV2, while data points are about the size of a single cell, their fixed locations may overlap with and capture mRNA from multiple cells. This is particularly apparent when spatial transcriptomics technologies are applied to kidney tissue, wherein the cell type landscape is dense and complex (Chabardès-Garonne et al., 2003; Raghubar et al., 2020). As a result, one collective shortcoming of methods like NMFreg and Seurat transfer learning is that they assume a one-to-one relationship between cell types and data points. More recent approaches have tried to dissect the multifaceted cell type profile of spatial transcriptomics spots, leading to a higher-resolution understanding of cell type distributions across tissue space. One method is RCTD, or robust cell type decomposition (Cable et al., 2021). RCTD leverages a generalized linear model for the total counts per gene on each data point. Specifically, counts are assumed to be Poisson distributed with the rate parameter being the product of the total molecules corresponding to a query data point and a mixture metric representative of the contribution of each gene to all cell types. In particular, one component of the mixture metric is the weighted sum of the average gene expression profiles per cell type (derived from a scRNA-seq reference dataset). One notable asset of RCTD is that it explicitly incorporates into its model platform effects that may mask true biological signals. Ultimately the weights that best fit the observed gene counts in a query data point indicate the proportions of each cell type captured by this data point.

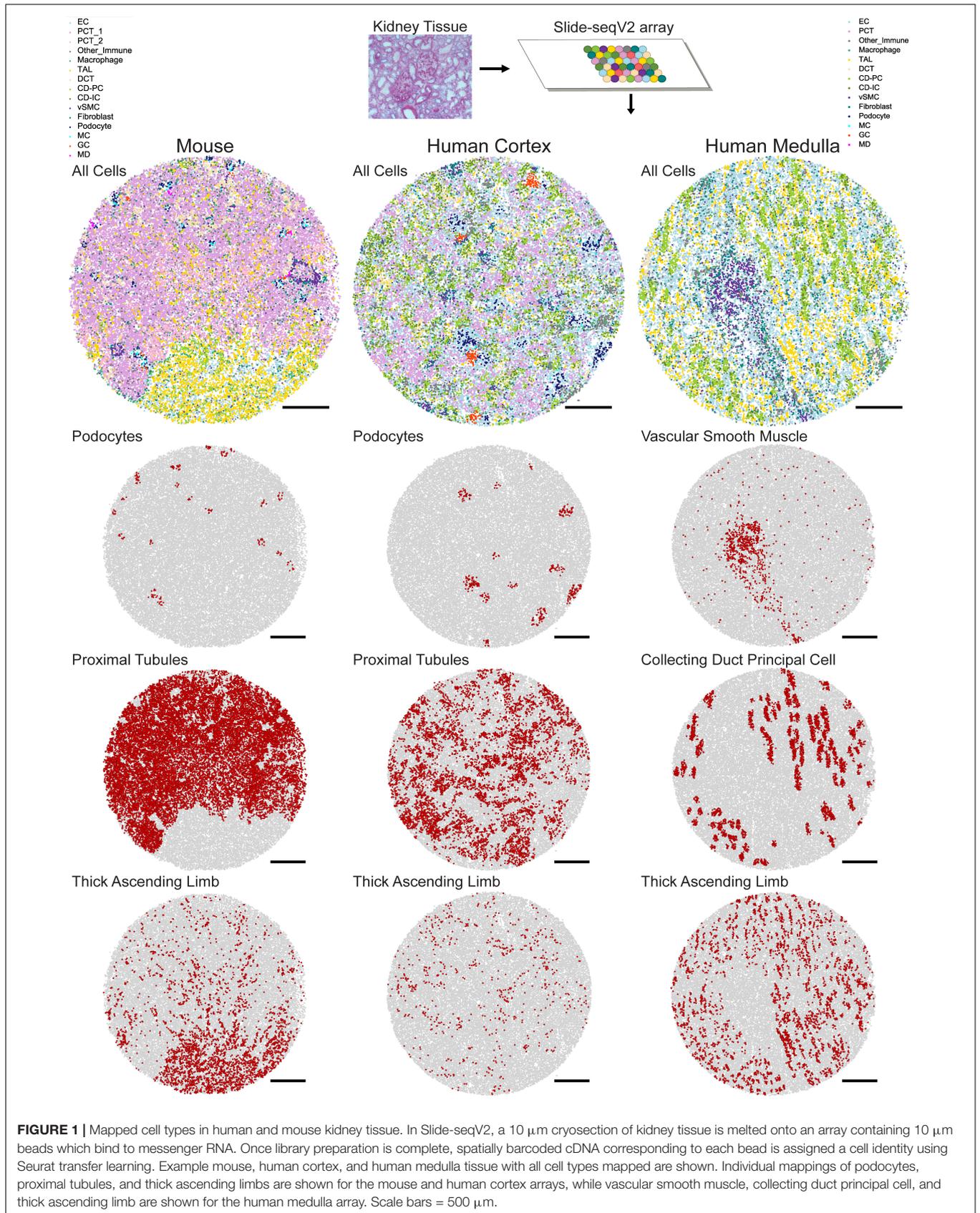
A second example of a method that is able to decompose spatial transcriptomic spots into multiple cell types is SPOTLight (Elosua-Bayes et al., 2021). SPOTLight follows the same trajectory as NMFreg, using both NMF and NNLS at its essence, but incorporates a more biologically driven initialization step and additional NNLS step. First, SPOTLight learns a k -dimensional set of factors (k = number of cell type categories) from a scRNA-seq reference *via* NMF. While the initial factorization may be chosen to be random, here, the authors chose to initialize the matrices so that they encode gene markers for each cell type in the first matrix, and cell type identities of cells in the second matrix. NMF is then run to find the factorization of genes by factors and factors by cells that best fits the scRNA-seq reference. An initial round of NNLS derives the loadings of the basis of factors defined in the previous step for every spatial transcriptomics spot. A second round of NNLS formalizes the relationship between spots and cell types as follows: First, it defines a matrix of factors by cell types, where every column represents the median factor profile per cell type (derived from the first NMF step). Second, the spatial transcriptomic spots are decomposed into a weighted combination of these cell type profiles.

In comparison to RCTD, SPOTLight does not account for platform effects. On the other hand, RCTD has two modes: doublet, which assumes a maximum of two cell types per data point and full, which allows for >2 cell types. The authors forewarn that performance is largely dictated by the mode that the user opts for; in particular, full mode may be subject to overfitting the data. In the context of the kidney, which is complex and tightly structured, doublet mode is not a safe assumption to make. On the other hand, SPOTLight does not make this distinction and searches for the maximum number of cell types, and its proven accuracy is based on this framework.

Additional notable methods have been developed to increase our confidence of both singular and multiple cell type calls per spot. Confidence in cell type classifications can be hindered by the sparse UMI-capture rate of spatial transcriptomic spots in comparison to scRNA-seq. FIST addresses this issue by imputing transcriptomes based on spatial relationships of data points (Li Z. et al., 2021). SPICEMIX utilizes both intrinsic transcriptomic information and the spatial relationships of data points to infer cell type identities (Chidester et al., 2021). Cell2Location utilizes a Bayesian model to predict cell type identities (Kleshchevnikov et al., 2020). Lastly, CellDART tackles the problem of identifying the cell type composition of each spot by leveraging neural networks (Bae et al., 2021).

SPATIAL CURATION OF CELL TYPE IDENTITIES

The cell type identities assigned from the observed gene expression markers through computational techniques is a noisy estimate. For example, the set of mRNAs detected in a 2-D plane might not be representative of the global distribution. As a result, it is advantageous to have prior knowledge of the



spatial distribution of cell types across cross-sectional tissue space as a method of further filtering out noisy cell type calls. For example, we know that podocytes occur solely within circular structures called glomeruli (Garg, 2018), and distal convoluted tubules (DCT) and principal cells of the collecting duct (CD-PC) form tubular structures (McCormick and Ellison, 2015). In our work, we have implemented automated methods to identify and filter out isolated podocyte or distal tubule calls that do not occur within these denser structures. Specifically, cell type calls that have less than k nearest neighbors with the cell type identity of interest are filtered out (Marshall et al., 2021).

Furthermore, we can use prior knowledge about the known vicinity of multiple cell type classes to hone in on biologically sensible cell type calls. For example, granular cells, macula densa, and glomeruli are known to be adjacent to each other in a structure called the juxtaglomerular apparatus (JGA) (Briggs and Schnermann, 1996). The polygon encapsulating each of these structures can be found and a filtration system can be set up where only those within x units of each other are maintained. Similarly, intercalated cells of the collecting duct (CD-IC) occur within larger tubular structures made up of CD-PC (Rao et al., 2019). Again, we can isolate instances of CD-IC that occur within x units of the edge of any CD-PC structure.

Lastly, certain cell types are thought to occur only within certain regions of tissue. In the kidney, two such regions are the cortex and the medulla. Our prior knowledge (Kriz and Kaissling, 2008) suggests that cell types belonging to glomerular structures (i.e., podocytes, mesangial cells, and endothelial cells) and proximal tubule cells occur only within the cortex. On the other hand, DCT, CD-IC, CD-PC, and the thick ascending limb (TAL) are enriched and form dense structures within the medulla. We can use our knowledge of these distributions of cells across the cortex and medulla to delineate the two regions, and further filter cell types that are called in the unlikely zones.

Not only are cell type calls trusted based on the agreement between their spatial distribution and known morphology of kidney tissue, but we can also verify their identities by looking at the relative expression of known cell type marker genes. Furthermore, cell type proportions can be computed and compared between Slide-seqV2 arrays and Hybridization Chain Reaction (HCR) images (Choi et al., 2018; Marshall et al., 2021).

DOWNSTREAM ANALYSES

Once cell type identities are well-defined across spatial transcriptomic data, the spatial distribution of data points can be explored at two levels: (1) cell types and (2) gene expression profiles within certain cell types. In terms of the former, we can simply ask what the cell type composition looks like in a tissue section at large, or in the medulla or cortex regions. Furthermore, we may be interested in the neighborhoods of certain cell types. That is, what cell types often neighbor a cell type of interest. In our work, we target cell type neighborhoods of *Trem2*⁺ and *LYVE1*⁺ macrophages using k nearest neighbors (Marshall et al., 2021). Lastly, we can look at the morphology of cell type structures. Utilizing computational geometry (Gillies

et al., 2007¹; Bellock et al., 2021), we can compute the convex hull or alpha shape of the coordinates of cell types, and compute the area of these polygons. We can then ask if, on average, these metrics shift across groupings of arrays (e.g., healthy and diseased samples).

The full potential of spatial transcriptomics arises when we look into the spatial distribution of cell types, paired with their gene expression profiles. Searching for spatially non-random distributions of gene expression can be done independent of cell type information, but it is hard to distinguish gene hits that are simply markers for cell types, which themselves are characterized by distinct spatial organization. For this reason, it is useful to look for spatially non-random genes within specific cell types. Several methods currently exist that accomplish this task of discovering spatially non-random genes. In the first Slide-seq paper (Rodrigues et al., 2019), the authors presented a permutation-based approach, which compares the distribution of randomly selected beads (while accounting for the total number of transcripts for each bead) with the distribution of beads expressing the specific gene of interest. They also defined spatially overlapping/anticorrelating gene pairs, regional gene enrichment and other interpretable measures using similar permutation based methods, to tackle important biological questions such as quantification and visualization of local transcriptional responses to injury. It is notable, however, that the permutation processes they used were time-consuming and included manual downstream filtering.

Another study introduces SPARK (Sun et al., 2020), which utilizes a generative, generalized linear model, argued to be more statistically robust and computationally efficient than preceding methods. It assumes that gene counts can be modeled by an over dispersed Poisson distribution, where the rate parameter is dictated by a non-spatial component (weighted sum of covariates, including batch information, cell-cycle information etc.), and a spatial component, dictated by either a gaussian distribution (representative of a localized gene expression pattern) or periodic distribution. Each component has a different variance, and the method addresses the null hypothesis that the variance of the spatial component is 0. The SPARK developers argue that a generative statistical model is far more efficient than permutation-based approaches.

Another unique advantage of coupling gene expression data with spatial information is inferring cell interactions. Previously, cell interactions were drawn from scRNA-seq, identifying ligand-receptor co-expression across cell type pairs (Cabello-Aguilar et al., 2020; Lu et al., 2021). However, there was no way of knowing if identified signaling was truly occurring between cell types in close vicinity to each other. With the onset of new spatial transcriptomics technologies, new methods have been developed to increase our confidence in identifying signaling between cell types by accounting for spatial information. Giotto looks for coexpression of ligand-receptor pairs, specifically within cell types in close vicinity to each other (Dries et al., 2021b), while MESSI is a method that uses multi-task learning to

¹Gillies, S., et al. (2007). *Shapely: Manipulation and Analysis of Geometric Objects*. Available online at: <https://github.com/shapely/shapely/blob/main/CITATION.txt>

predict response genes to intracellular and intercellular signals, considering expression from neighboring cells (Li D. et al., 2021).

CONCLUSION

In this manuscript, we have reviewed the biological and computational basis of spatial transcriptomics analysis, with an example of cell type mapping and downstream applications in kidney tissue. Spatial transcriptomics technologies are evolving at a rapid pace and extending beyond transcriptomics into metabolomics and proteomics (Lundberg and Borner, 2019; Ganesh et al., 2021; Yuan et al., 2021). The integration of unbiased spatial omics technologies will provide a powerful set of tools to characterize disease processes in intact tissue (Dries et al., 2021a). We hope that these technologies will not only develop data rich atlases of healthy and diseased tissues, but also provide a platform for advances in the fundamental understanding of disease mechanisms and highlight new therapeutic targets.

REFERENCES

- 10x Genomics (2007). *Spatial Gene Expression – 10x Genomics*. Available online at: <https://www.10xgenomics.com/products/spatial-gene-expression> (accessed October 31, 2021).
- Bae, S., Na, K. J., Koh, J., Lee, D. S., Choi, H., Kim, Y. T., et al. (2021). CellDART: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. *bioRxiv* [Preprint]. doi: 10.1101/2021.04.26.441459
- Bellock, K., Godber, N., and Kahn, P. (2021). *bellockk/alphashape: v1.3.1 Release*. Zenodo. doi: 10.5281/ZENODO.4697576
- Briggs, J. P., and Schnermann, J. B. (1996). Whys and wherefores of juxtaglomerular apparatus function. *Kidney Int.* 49, 1724–1726. doi: 10.1038/ki.1996.255
- Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., Colinge, J., et al. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* 48:e55. doi: 10.1093/nar/gkaa183
- Cable, D. M., Murray, E., Zou, L. S., Goeva, A., Macosko, E. Z., Chen, F., et al. (2021). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* 1–10. doi: 10.1038/s41587-021-00830-w [Epub ahead of Print].
- Chabardès-Garonne, D., Mejéan, A., Aude, J. C., Cheval, L., Di Stefano, A., Gaillard, M. C., et al. (2003). A panoramic view of gene expression in the human kidney. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13710–13715.
- Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., et al. (2021). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball patterned arrays. *bioRxiv* [Preprint]. doi: 10.1101/2021.01.17.427004
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090.
- Chidester, B., Zhou, T., and Ma, J. (2021). SPICEMIX: integrative single-cell spatial modeling for inferring cell identity. *bioRxiv* [Preprint]. doi: 10.1101/2020.11.29.383067
- Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., et al. (2018). Third-generation hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145:dev165753. doi: 10.1242/dev.165753
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dries, R., Chen, J., del Rossi, N., Khan, M. M., Sistig, A., and Yuan, G.-C. (2021a). Advances in spatial transcriptomic data analysis. *Genome Res.* 31, 1706–1718.
- Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., et al. (2021b). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* 22:78. doi: 10.1186/s13059-021-02286-2
- Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* 49:e50. doi: 10.1093/nar/gkaa183
- Ganesh, S., Hu, T., Woods, E., Allam, M., Cai, S., Henderson, W., et al. (2021). Spatially resolved 3D metabolomic profiling in tissues. *Sci. Adv.* 7:eabd0957.
- Garg, P. A. (2018). Review of podocyte biology. *Am. J. Nephrol.* 47(Suppl. 1), 3–13. doi: 10.1159/000481633
- Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542, 352–356.
- Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H. W., Li, T., et al. (2020). Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *bioRxiv* [Preprint]. doi: 10.1101/2020.11.15.378125
- Kriz, W., and Kaissling, B. (2008). Structural organization of the mammalian kidney. *Seldin Giebischs Kidney* 8, 479–563. doi: 10.1016/b978-012088488-9.50023-1
- Kuncheva, L. I., and Rodríguez, J. J. (2014). A weighted voting framework for classifiers ensembles. *Knowled. Inform. Syst.* 38, 259–275. doi: 10.1007/s10115-012-0586-6
- Lake, B. B., Rajasree, M., Seth, W., Qiwen, H., Melo, S. R., Kian, K., et al. (2021). An atlas of healthy and injured cell states and niches in the human kidney. *bioRxiv* [Preprint]. doi: 10.1101/2021.07.28.454201
- Lee, D. D., and Seung, H. S. x (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565
- Li, D., Ding, J., and Bar-Joseph, Z. (2021). Identifying signaling genes in spatial single-cell expression data. *Bioinformatics* 37, 968–975.
- Li, Z., Song, T., Yong, J., and Kuang, R. (2021). Imputation of spatially-resolved transcriptomes by graph-regularized tensor completion. *PLoS Comput. Biol.* 17:e1008218. doi: 10.1371/journal.pcbi.1008218
- Liao, J., Lu, X., Shao, X., Zhu, L., and Fan, X. (2021). Uncovering an Organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.* 39, 43–58. doi: 10.1016/j.tibtech.2020.05.006
- Liu, Y., Yang, M., Deng, Y., Su, G., Enniful, A., Guo, C. C., et al. (2020). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 183, 1665–1681.e18. doi: 10.1016/j.cell.2020.10.026
- Longo, S. K., Guo, M. G., Ji, A. L., and Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue

AUTHOR CONTRIBUTIONS

TN, QW, and JM designed, wrote, and edited the manuscript. JM prepared the figure. AG edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

BroadIgnite (JM) CZI Seed Networks grant number CZI2019-02447 (AG).

ACKNOWLEDGMENTS

The authors thank the entire Chen and Greka labs at the Broad Institute of MIT and Harvard for helpful discussions throughout this project.

- dynamics. *Nat. Rev. Genet.* 22, 627–644. doi: 10.1038/s41576-021-00370-8
- Lu, M., Kuo, P. H., Hsu, C. Y., Chiu, Y. H., Liu, Y. W., Lu, M. L., et al. (2021). LR hunting: a random forest based cell-cell interaction discovery method for single-cell gene expression data. *Front. Genet.* 12:708835. doi: 10.3389/fgene.2021.708835
- Lundberg, E., and Borner, G. H. H. (2019). Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* 20, 285–302. doi: 10.1038/s41580-018-0094-y
- Marshall, J. L., Noel, T., Wang, Q. S., Bazua-Valenti, S., Chen, H., Murray, E., et al. (2021). High resolution slide-seqV2 spatial transcriptomics enables discovery of disease-specific cell neighborhoods and pathways. *bioRxiv* [Preprint]. doi: 10.1101/2021.10.10.463829
- McCormick, J. A., and Ellison, D. H. (2015). Distal convoluted tubule. *Compr. Physiol.* 5, 45–98. doi: 10.1002/cphy.c140002
- Melo Ferreira, R., Sabo, A. R., Winfree, S., Collins, K. S., Janosevic, D., Gulbranson, C. J., et al. (2021). Integration of spatial and single-cell transcriptomics localizes epithelial cell-immune cross-talk in kidney injury. *JCI Insight* 6:e147703. doi: 10.1172/jci.insight.147703
- Nature Methods (2021). Method of the Year 2020: spatially resolved transcriptomics. *Nat. Methods* 18:1. doi: 10.1038/s41592-020-01042-x
- Raghubar, A. M., Pham, D. T., Tan, X., Grice, L. F., Crawford, J., Lam, P. Y., et al. (2020). Spatially resolved transcriptome profiles of mammalian kidneys illustrate the molecular complexity of functional nephron segments, cell-to-cell interactions and genetic variants. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.29.317917
- Rao, R., Bhalla, V., and Pastor-Soler, N. M. (2019). Intercalated cells of the kidney collecting duct in kidney physiology. *Semin. Nephrol.* 39, 353–367. doi: 10.1016/j.semnephrol.2019.04.005
- Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467. doi: 10.1126/science.aaw1219
- Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82.
- Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145.
- Stickels, R. R., Murray, E., Kumar, P., Li, J., Marshall, J. L., Di Bella, D. J., et al. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* 39, 313–319. doi: 10.1038/s41587-020-0739-1
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. III, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21.
- Su, G., Qin, X., Enninfu, A., Bai, Z., Deng, Y., Liu, Y., et al. (2021). Spatial multi-omics sequencing for fixed tissue via DBiT-seq. *STAR Protoc.* 2:100532. doi: 10.1016/j.xpro.2021.100532
- Subramanian, A., Vernon, K., Zhou, Y., Marshall, J. L., Alimova, M., Zhang, F., et al. (2021). Obesity-instructed TREM2high macrophages identified by comparative analysis of diabetic mouse and human kidney at single cell resolution. *bioRxiv* [Preprint]. doi: 10.1101/2021.05.30.446342
- Sun, S., Zhu, J., and Zhou, X. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* 17, 193–200. doi: 10.1038/s41592-019-0701-7
- Yuan, Z., Zhou, Q., Cai, L., Pan, L., Sun, W., Qumu, S., et al. (2021). SEAM is a spatial single nuclear metabolomics method for dissecting tissue microenvironment. *Nat. Methods* 18, 1223–1232.
- Zollinger, D. R., Lingle, S. E., Sorg, K., Beechem, J. M., and Merritt, C. R. (2020). GeoMx™ RNA assay: high multiplex, digital, spatial analysis of RNA in FFPE tissue. *Methods Mol. Biol.* 2148, 331–345. doi: 10.1007/978-1-0716-0623-0_21

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Noel, Wang, Greka and Marshall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.