COMPUTATIONAL BIOLOGY

Integrative analysis methods for spatial transcriptomics

Computational methods use different integrative strategies to tackle the challenges of spatially resolved transcriptomics data analysis.

Shaina Lu, Daniel Fürth and Jesse Gillis

Multicellular organisms are defined by the cells that compose them as well as the relationships between those cells, partially captured by cells' spatial organization. Although single-cell transcriptome sequencing (scRNA-seq) has had a transformative impact in characterizing cells as independent elements, many aspects of the cells' relationships are lost with this technique, including spatial distribution. Newly developed tools have focused on assaying the spatial organization of cells in tissues, but there are often trade-offs between spatial resolution and the number of unique RNA transcripts assayed. In this issue of *Nature Methods*, Scalia et al.[1] and Hu et al.[2] introduce computational tools to integrate spatially resolved transcriptomic data with scRNA-seq and/or histology data to bridge these trade-offs and provide a better understanding of the spatial organization of tissues.

Although focusing on different parts of the analysis process, both SpaGCN[2] and Tangram[1], the methods of Scalia et al. and Hu et al., respectively, are computational methods for data integration to improve the interpretation of spatial expression (Fig. 1). SpaGCN focuses on incorporating existing histology to identify spatial domains and subsequently identify genes differentially expressed between the spatial clusters. Though Tangram also incorporates aspects of these steps, its principal focus is on providing cross-modality data integration with scRNA-seq data. After this integration, a number of analysis tasks can be accomplished using Tangram, such as imputing additional genes in spatial data that are not transcriptome wide or deconvolving spatial data that are not of cellular resolution into cell-type proportions. The different forms of analysis accomplished by Tangram and SpaGCN are largely complementary.

SpaGCN and Tangram are part of a broader trend toward the development of computational methods for spatial transcriptomics[3,4]. This development is driven by the increased availability of spatially resolved data and techniques for generating it[5]. SpaGCN is analytically unusual within this cohort for its combined approach to resolving spatial domains and computing differential expression (rather than just one or the other). Like SpaGCN, Tangram uses histology data, but its focus is on aligning any type of single-cell (or single-nucleus) RNA-seq to spatial data packaged with a breadth of methodological tools after integration. Tangram's use as a single-cell and spatial integration tool will be helpful in meeting the popular demand for a straightforward tool to visualize in situ clusters obtained from scRNA-seq[6,7]. Whereas some earlier tools are specific to one type or class of spatial experiment, both SpaGCN and Tangram can be applied across experimental assays and are meant to be universal tools for the spatial field.

As experimental technologies continue to improve[5], the gap between high spatial resolution and percentage of the transcriptome assayed continues to shrink[8]. However, until new techniques that promise to cover the whole transcriptome with subcellular resolution are readily available and accessible, computational data integration is necessary to bridge this gap. Although recent methods are customized for spatial data, the fundamental models are often more general. In essence, information is shared between cells within a dataset in a structured way to minimize noise, and then cells are aligned across datasets. If spatial metadata are available for one of those sets of cells, or the way information is shared between cells is defined by known location, then these data integration methods become spatial data integration methods.

A prominent discussion point in Scalia et al.[1] is the promise of data integration approaches for bringing us closer to a truly multimodality understanding of biology through the creation of large, integrated datasets such as the Human Cell Atlas[9]. Because cellular location is among the most fundamental types of metadata, integration of spatial data is important for large-scale data integration into a common framework. This will allow evaluation of the underlying data and methods, currently a major challenge within the field. As methods improve and reference data emerge, uncovering novel drivers of variability that contribute to disease or other phenotypic

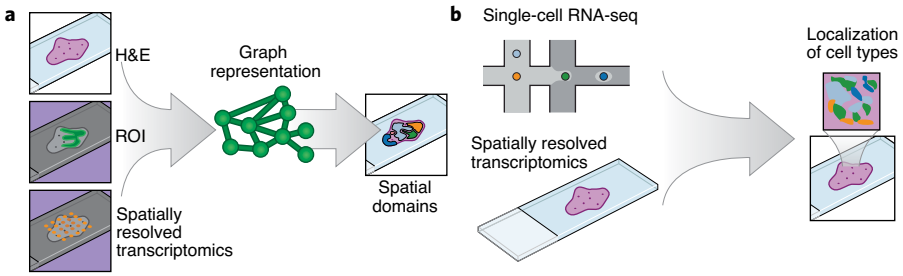

Fig. 1 | Schematic comparison of SpaGCN and Tangram analysis methods for spatially resolved transcriptomics. a, SpaGCN integrates histological information, user-defined region of interest (ROI) and spatial transcriptomics into a graph convolutional network (GCN) and performs unsupervised clustering on the graph representation to arrive at a set of spatial domains. H&E, hematoxylin and eosin histochemical staining. **b**, Tangram aligns single-cell data with spatially resolved data to arrive at imputed and deconvolved spatial domains with single-cell-like qualities.

differences should also become possible. Although some phenotypic differences reflect cell-autonomous variability, a substantial fraction is likely emergent from the relationships between cells. Uncovering the logic of how these cell-cell relationships contribute to tissue function is an important avenue opened up by these integrative methods and the data underlying them.

An important area for technical improvement in analysis methods rests on the fact that current assessments are quite qualitative in nature. Although this does not place a direct limit on the efficacy of methods, it does place a limit on our understanding of how best to apply them or improve upon them. Spatial clustering methods or identification of spatial distributions of cell types, for example, are often visualized with microscopy images and are said to be good representations when these computationally defined features match the cytoarchitecture and morphology of the tissue. There are some popular statistical measures, such as those for determining spatial autocorrelation, but these do not capture the performance of all classes of spatial analysis tasks. In addition to the advances in spatial analysis represented by Tangram and SpaGCN, other spatial tools, not detailed here, are also useful. As with any new field, to better understand the pros and cons of the many spatial analysis tools, an independent, rigorous and quantitative benchmarking across spatially resolved transcriptomics analysis tools is needed.

Moving forward, tools such as SpaGCN[2] and Tangram[1] will be invaluable in establishing spatial regions directly derived from gene expression data, rather than defined from traditionally agreed anatomical boundaries. Although gene expression need not be the be-all and end-all, it provides a unified and quantitative framework to link activity at the cellular and tissue levels. Boundaries defined from spatial expression will link processes such as cell-cell communication, cell migration and morphogenesis in organ formation. Analysis tools for spatially resolved transcriptomics usually take a data-first approach to understanding biology, sometimes described as 'unbiased', but integration with existing biological knowledge to understand causal mechanisms will ultimately require testable hypotheses in combination with high-quality data.

Particularly important for future study are questions relating to evolution and development, as well as their interplay, as modular expansion of spatial domains to create new functions is a repeated theme of both. Evolution and development offer a vast space from which to collect data, with a new class of integration to consider, for which systematic tools such as SpaGCN and Tangram will be essential. Although these tools can capture biological phenomena such as morphological patterns in the brain, clusterings have difficulty in distinguishing between byproducts of evolution and phenotypic traits that are the direct products of selection. Spatial expression across development should provide valuable insight into molecular mechanisms, whereas spatial expression across species helps to capture selection and conservation.

The rapid parallel development of molecular tools available both in spatial genomics[5] and in lineage tracing and clonal identification[10] will, together with computational methods like SpaGCN and Tangram, enable a new era of experimental design and discovery. Spatially resolved transcriptomics has the potential to be the revolution of this decade, much as single-cell techniques were for the previous one; these analysis tools will help to realize that potential. ❐

Shaina Lu, Daniel Fürth and Jesse Gillis ✉

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.

✉*e-mail: JGillis@cshl.edu*

Published online: 28 October 2021
https://doi.org/10.1038/s41592-021-01272-7

References

1. Scalia, G. et al. *Nat. Methods* https://doi.org/10.1038/s41592-021-01264-7 (2021).
2. Hu, J. et al. *Nat. Methods* https://doi.org/10.1038/s41592-021-01255-8 (2021).
3. Svensson, V., Teichmann, S. A. & Stegle, O. *Nat. Methods* **15**, 343–346 (2018).
4. Edsgärd, D., Johnsson, P. & Sandberg, R. *Nat. Methods* **15**, 339–342 (2018).
5. Asp, M., Bergenstråhle, J. & Lundeberg, J. *Mol. Cell. Dev. Biol.* **42**, e1900221 (2020).
6. La Manno, G. et al. *Nature* **596**, 92–96 (2021).
7. Di Bella, D. J. et al. *Nature* **595**, 554–559 (2021).
8. Srivatsan, S. R. et al. *Science* **373**, 111–117 (2021).
9. Regev, A. et al. *eLife* **6**, e27041 (2017).
10. Wagner, D. E. & Klein, A. M. *Nat. Rev. Genet.* **21**, 410–427 (2020).

Acknowledgements

S.L. is supported by the Edward and Martha Gerry Fellowship funded by The William Stamps Farish Fund and the Gladys and Roland Harriman Foundation. D.F. is supported by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. J. G. is supported by NIH grants R01MH113005 and R01LM012736.

Competing interests

The authors declare no competing interests.

Check for updates

ORGANOIDS

Towards spheroid-omics

The MISpheroID knowledgebase records and organizes experimental parameters from thousands of cancer spheroid experiments, revealing heterogeneity and a lack of transparency in key spheroid research reporting practices.

Timothy L. Downing

For more than 40 years, researchers have explored the development of cell culture models that recapitulate biological processes as they occur within three-dimensional (3D) physiological contexts. However, within the past 10 years, there has been a sharp increase in the rate of spheroid studies published, owing to the valuable insights that these models provide into cancer pathophysiology (including cell migration and matrix invasion), as well as pharmacological response through drug testing[1]. 3D spheroid cultures are established through the aggregation of suspended (non-adherent) cells derived