How Do We Visualize Space in Molecular Biology? A Study of Spatial Transcriptomics Visualization Practices

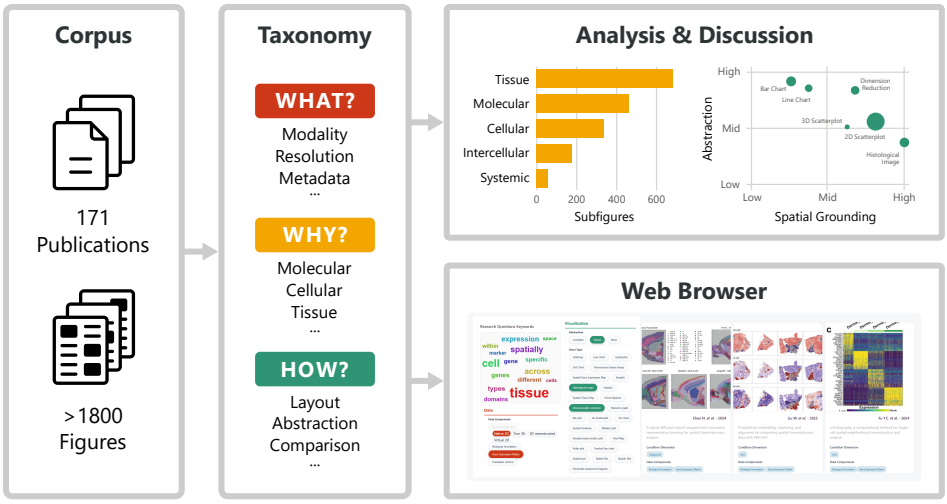

Fig. 1: To characterize how spatial transcriptomics can be visualized in practice, we annotated over 1,800 figures from 171 publications using a structured taxonomy organized around three dimensions: data (what), tasks (why), and visualization design (how). This approach reveals recurring patterns in how biological questions map onto visual representations.

Modern omics technologies capture where and when molecules act within tissues, producing datasets that are structured, interdependent, and multimodal [8]. This growing complexity of biological datasets presents both challenges and opportunities for the design of informative visualizations [10]. In this work we seek to address the question: how can we design visual representations and interactions that enable scientists to reason across space and molecular scales?

Recent advances in single-cell and imaging-based molecular profiling allow researchers to measure thousands of genes per cell while preserving spatial coordinates and sometimes temporal information [5]. These spatiotemporal omics datasets pose challenges in visualization research that extend well beyond biology. Thousands of expression features (i.e., genes) measured for thousands of cells, paired with tissue images or longitudinal time series, necessitate the development of perceptually scalable representations [11, 6]. Visualization approaches need to preserve spatial autocorrelation and temporal continuity to reveal inherent data structures, such as neighboring cells that influence each other, or correlated trajectories formed from consecutive snapshots [11]. Some datasets in this domain are deeply multimodal, integrating transcriptomic, proteomic, and image-based measurements with distinct resolution and uncertainty profiles, calling for coordinated visual representations that support cross-modal reasoning [4, 12]. Many datasets also embody significant measurement uncertainty, ranging from technical noise to biological variability [2, 1], reinforcing the need for visualization strategies that enable robust interpretation.

Despite this rapid growth in data complexity and the proliferation of analysis tools, the visualization community lacks a systematic understanding of how visual representations encode data, support biological reasoning, and align with the questions researchers seek to answer. Existing work has primarily focused on developing new methods or systems, while the relationship between biological inquiry and visualization design remains underexplored. While prior surveys have characterized genomic visualization techniques [9] and single-cell atlas visualization [3], no systematic analysis has addressed the spatial transcriptomics domain through the lens of biological tasks. This gap is consequential: without a principled account of how visualization practice maps onto biological tasks, it is difficult to identify where current approaches fall short, to compare tools in terms of their analytical affordances, or to guide the design of next-generation visual analytics systems. As the field continues to scale—in data volume, modality, and biological complexity—the cost of this blind spot grows accordingly.

To address this gap, we performed a systematic analysis of visualization practices in spatial transcriptomics, grounded in the structured annotation of figure panels from published analysis and visualization tools. We constructed our corpus through targeted searches in Scopus and Google Scholar, complemented by an AI-assisted search using Asta by Ai2. We based our search on keywords such as “spatial transcriptomics visualization”, “spatial omics visualization”, and “spatial data analysis tool”. These searches yielded an initial seed collection that we expanded iteratively by incorporating tools and methods referenced within selected

works or known to be relevant but not retrieved via keyword search. We applied inclusion and exclusion criteria to focus on visualizations that support biological interpretation: we included computational and statistical methods for spatial transcriptomics that produce visual representations for analysis, as well as tools specifically designed for interactive visualization of spatial biological data. This filtering process resulted in a curated set of 171 papers, from which we extracted over 3,000 figure panels. After filtering, more than 1,800 figures were retained and annotated using a coding system of over 50 metadata terms (Figure 1).

We adopt a question-driven perspective that centers biological inquiry as the primary organizing principle. Building on Munzner's *what–why–how* framework [7] and adapting it to the spatial transcriptomics domain, each figure is annotated along three interconnected dimensions: data (*what*), tasks (*why*), and visualization design (*how*). The data dimension captures data modality and resolution, among others; the task dimension encodes biological questions across multiple scales, including molecular, cellular, intercellular, tissue, and integrative analyses; and the visualization dimension characterizes design choices such as layout, abstraction level, comparative strategies, and scalability mechanisms. For each figure we additionally annotate the underlying biological research question it aims to address, capturing an abstracted representation of the analytical intent driving the visualization. We formalize these dimensions in a detailed taxonomy. To ensure consistency, two annotators independently coded a shared subset of figures. Through iterative discussion, we reconciled disagreements and refined the taxonomy before proceeding with full annotation.

Analysis of the annotated corpus reveals several clear and recurring patterns. First, visualization practice is heavily skewed toward tissue structure and cell-type identification tasks, which together account for the large majority of figure panels. This reflects the dominance of exploratory and descriptive analysis in current spatial transcriptomics workflows. Second, when comparative analysis is supported, juxtaposition through small multiples is by far the most common strategy, typically used to contrast gene expression patterns across tissue regions or experimental conditions. Third, a fraction of visualizations discard spatial context entirely, reproducing single-cell transcriptomics conventions despite the inherently spatial nature of the data. To make these findings accessible and reusable, all annotated results are made available through an interactive web-based figure browser. This browser allows researchers to query and filter the full corpus by data type, biological task, visualization technique, or research question.

Currently, our annotated corpus captures what visualizations show and how they are designed in a structured taxonomy; specific research questions addressed by the figures were individually added based on the descriptions in the publications. It remains a challenge and opportunity for future work to also systematically organize and abstract these biological questions themselves. To this end, we are actively exploring complementary methods for deriving higher-level task structures from the corpus. These ideas include (*i*) LLM-based mappings to existing task taxonomies; (*ii*) embedding-based clustering of research questions in a latent space; and (*iii*) manual affinity-based organization, in which related questions are grouped iteratively to surface higher-level patterns in how biologists reason about their data.

Taken together, our collection of resources and the findings we gained from them offer a dual contribution to the field. For the visualization research community, our work moves beyond anecdotal observations about biological visualization to provide an empirically grounded map of the design space: what has been built, what tasks have been addressed, and where lie the gaps? For domain scientists and tool developers, our perspective on existing approaches and their limitations supports more informed selection and use of visualization techniques, and provides a vocabulary for articulating what a given tool does and does not support. More broadly, we hope this work demonstrates the value of systematic analysis of visualization practice as a complement to the development of new methods, an approach that can be extended to other rapidly evolving domains beyond spatial biology.

[1] Á. Arzalluz-Luque et al. "Delineating biological and technical variance in single cell expression data". In: *The International Journal of Biochemistry & Cell Biology* 90 (2017), pp. 161–166. DOI: 10.1016/j.biocel.2017.07.006.

[2] G. Gorin, J. J. Vastola, and L. Pachter. "Studying stochastic systems biology of the cell with single-cell genomics data". In: *Cell Systems* 14.10 (2023), 822–843.e22. DOI: 10.1016/j.cels.2023.08.004.

[3] M. S. Keller et al. "The State of Single-Cell Atlas Data Visualization in the Biological Literature". In: *IEEE Computer Graphics and Applications* 45.5 (2025), pp. 18–34. DOI: 10.1109/mcg.2025.3583979.

[4] M. S. Keller et al. "Vitessce: integrative visualization of multimodal and spatially resolved single-cell data". In: *Nature Methods* 22.1 (2024), pp. 63–67. DOI: 10.1038/s41592-024-02436-x.

[5] D. Klein et al. "Mapping cells through time and space with moscot". In: *Nature* 638.8052 (2025), pp. 1065–1075. DOI: 10.1038/s41586-024-08453-2.

[6] L. Liu et al. "Spatiotemporal omics for biology and medicine". In: *Cell* 187.17 (2024), pp. 4488–4519. DOI: 10.1016/j.cell.2024.07.040.

[7] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. ISBN: 9781498759717.

[8] Y. Nam et al. "Harnessing Artificial Intelligence in Multimodal Omics Data Integration: Paving the Path for the Next Frontier in Precision Medicine". In: *Annual Review of Biomedical Data Science* 7.1 (2024), pp. 225–250. DOI: 10.1146/annurev-biodatasci-102523-103801.

[9] S. Nusrat, T. Harbig, and N. Gehlenborg. "Tasks, Techniques, and Tools for Genomic Data Visualization". In: *Computer Graphics Forum* 38.3 (2019), pp. 781–805. DOI: 10.1111/cgf.13727.

[10] C. Seydel. "Beyond cell atlases: spatial biology reveals mechanisms behind disease". In: *Nature Biotechnology* 43.6 (2025), pp. 841–844. DOI: 10.1038/s41587-025-02699-5.

[11] B. Velten and O. Stegle. "Principles and challenges of modeling temporal and spatial omics data". In: *Nature Methods* 20.10 (2023), pp. 1462–1474. DOI: 10.1038/s41592-023-01992-y.

[12] T. Xia et al. "Empowering integrative and collaborative exploration of single-cell and spatial multimodal data with SGS genome browser". In: *Cell Genomics* 5.5 (2025), p. 100848. DOI: 10.1016/j.xgen.2025.100848.