Scientific discovery in biomedicine from spatially resolved omics

Jovan Tanevski and Julio Saez-Rodriguez

Institute for Computational Biomedicine, Heidelberg University Hospital, Germany

Advances in technologies for multiplexed imaging and spatially resolved sequencing enable a more complex look at the structure and function of tissues [1]. Such data enables a new venue for basic scientific discovery and advancement of its clinical applications [2]. As initial approaches to the analysis of spatially resolved omics mature [3], we improve our understanding of the data and establish better relationships to prior knowledge. We now have the opportunity to construct more complex models to acquire novel insights into structure, function, and the emergence of structure-function relationships. These models can provide, on one hand, a better understanding of biological relationships that explain organization and regulation [4]. On the other hand, they can go beyond the scope of current histo(patho)logy for patient stratification and following of disease progression [5].

The aforementioned technologies measure the abundance of a set of molecular entities (e.g. number of gene transcripts or molecules of protein) per spatial unit. Because of the different resolutions of the technologies, each spatial unit has physical coordinates that correspond to either the location of a specific molecule, the center of a cell, or the center of a larger spot covering a number of cells. Typically, the spatial resolution of a unit is traded-off with the number of captured distinct entities.

Our goal is to capture relationships that are consistent across studied samples while taking into account spatial organization. To this end we proposed a multiview learning approach, capturing flexibly and explicitly different aspects of the data [4]. First, an intrinsic view is comprised of features capturing the biological context of interest per spatial unit. Subsequent views capture different spatial or functional aspects of the data. For example, a spatial-context-specific view can summarize measurements coming from the immediate neighborhood of the spatial unit, or capture the broader spatial context by summarizing measurements in a fixed radius around a spatial unit. View-specific models are then trained using intrinsic features as targets. Finally, a meta-model fuses the predictions from the view-specific models.

In this frame, to explore the spatial organization of the tissue, prior biological knowledge in the form of markers or atlases can be used to assign or deconvolve the cell-type composition of a spatial unit as an intrinsic representation [6]. The spatial views then capture the cell composition at different radiuses around the spatial unit. The models, in turn, capture persistent patterns of cell-type composition in different spatial contexts. Similarly, knowledge [7, 8] about functionally-related genes (genesets) can be used to estimate activities of processes such as signaling pathways, allowing to model pathway crosstalks. To capture intercellular communication more explicitly, each spatial unit can be represented by a subset of genes coding for receptors, whose expressions can then be related to the expression of ligands from different spatial or cell-type contexts [9]. The ligand-receptor relationships captured by the model provide insight into the potential channels of communication.

The primary use of the models is not in a predictive setting, but to generate hypotheses about the relationships in the data. Therefore, the output consists of the contribution of each view to the meta-model and the estimated relevance of the predictor-target relationships within the view-specific models. Ground truth about the existing relationships in the data is not complete

or not available. Therefore, we estimate the predictive performance (e.g. variance explained) of the constructed models as a proxy for the relevance of the information captured by the models. Additionally, a limited set of relationships from prior knowledge can be used as support for the correctness of the output relationships, while the remaining candidates define a set of hypotheses for further experimental validation.

The view- and target-specific models and their interpretability are central to the framework, as they provide insight into the underlying biology of the studied system. This framework lends itself readily to symbolic approaches to explore different, more explicit forms of relationships in the data informed by prior knowledge. The generated hypotheses can be also used to constrain the design of smaller experimental panels for patient stratification or the generation of measurements needed to uncover more direct mechanistic relationships. Generalizing across multiple experimental data opens a venue for a systems approach to the discovery of higher-order relationships underlying emergent molecular programs and functional tissue organization.

References

1. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nature Biotechnology* **40,** 308–318. https://doi.org/10.1038/s41587-021-01182-1 (2022).
2. Elhanani, O., Ben-Uri, R. & Keren, L. Spatial profiling technologies illuminate the tumor microenvironment. *Cancer Cell* **41,** 404–420. https://doi.org/10.1016/j.ccell.2023.01.010 (2023).
3. Zeng, Z., Li, Y., Li, Y. & Luo, Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biology* **23.** https://doi.org/10.1186/s13059-022-02653-7 (2022).
4. Tanevski, J., Ramirez Flores, R. O., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology* **23.** https://doi.org/10.1186/s13059-022-02663-5 (2022).
5. Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* **608,** 766–777. https://doi.org/10.1038/s41586-022-05060-x (2022).
6. Rahimi, A., Vale-Silva, L. A., Savitski, M. F., Tanevski, J. & Saez-Rodriguez, J. DOT: Fast Cell Type Decomposition of Spatial Omics by Optimal Transport. *arXiv.* https://doi.org/10.48550/arXiv.2301.01682 (2023).
7. Türei, D. *et al.* Integrated intra and intercellular signaling knowledge for multicellular omics analysis. *Molecular Systems Biology* **17.** https://doi.org/10.15252/msb.20209923 (2021).
8. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* **2.** https://doi.org/10.1093/bioadv/vbac016 (2022).
9. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nature Communications* **13.** https://doi.org/10.1038/s41467-022-30755-0 (2022).