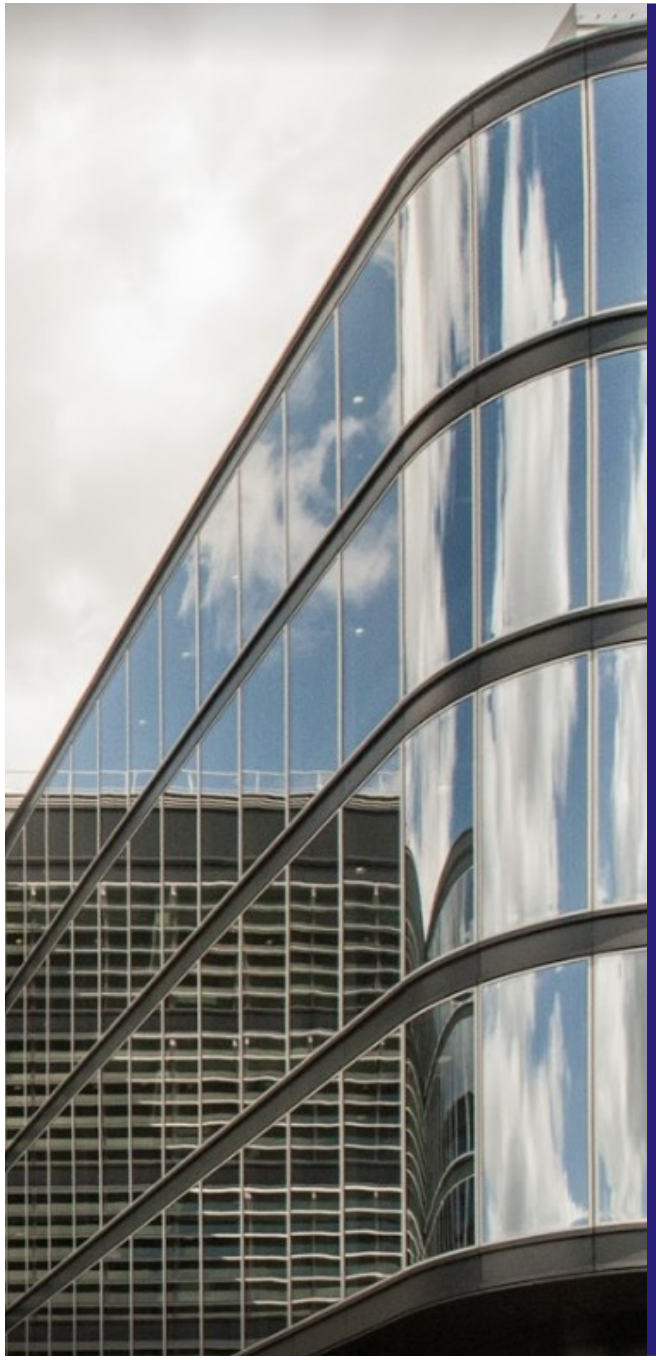

Spatially-resolved transcriptomics meets Deep Learning:

denoising omics data matrix using Optimal Transport and Graph Attention Networks

Lucas RINCÓN DE LA ROSA

02/10/2024

01.Scientific background

Single cell vs Bulk.

- **Proportion of zeros** in **bulk** RNA-seq data is usually **10–40%** *(Jiang et al. Genome Biol 2022)*.
- That proportion can be as high as **90%** in single cell RNA sequencing (**scRNA-seq**) data.
- The high data sparsity provoked the use of **zero-inflated** models **and** the development of **imputation methods** for reducing zeros.

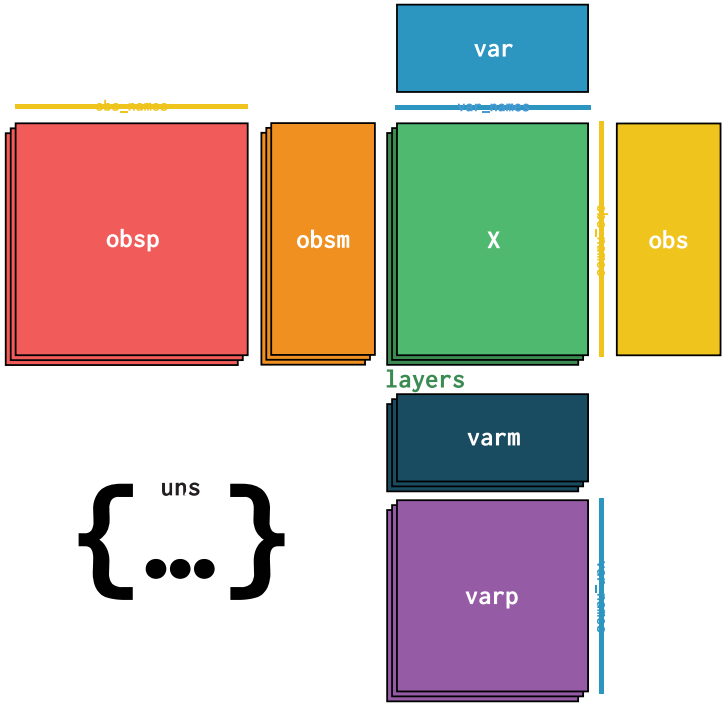

***Figure 2**: RNA-seq AnnData example. .X: containing the count matrix; .obs: containing annotations for the cells/spots and .var: contains annotations for the genes. (Virshup el al. 2021)*

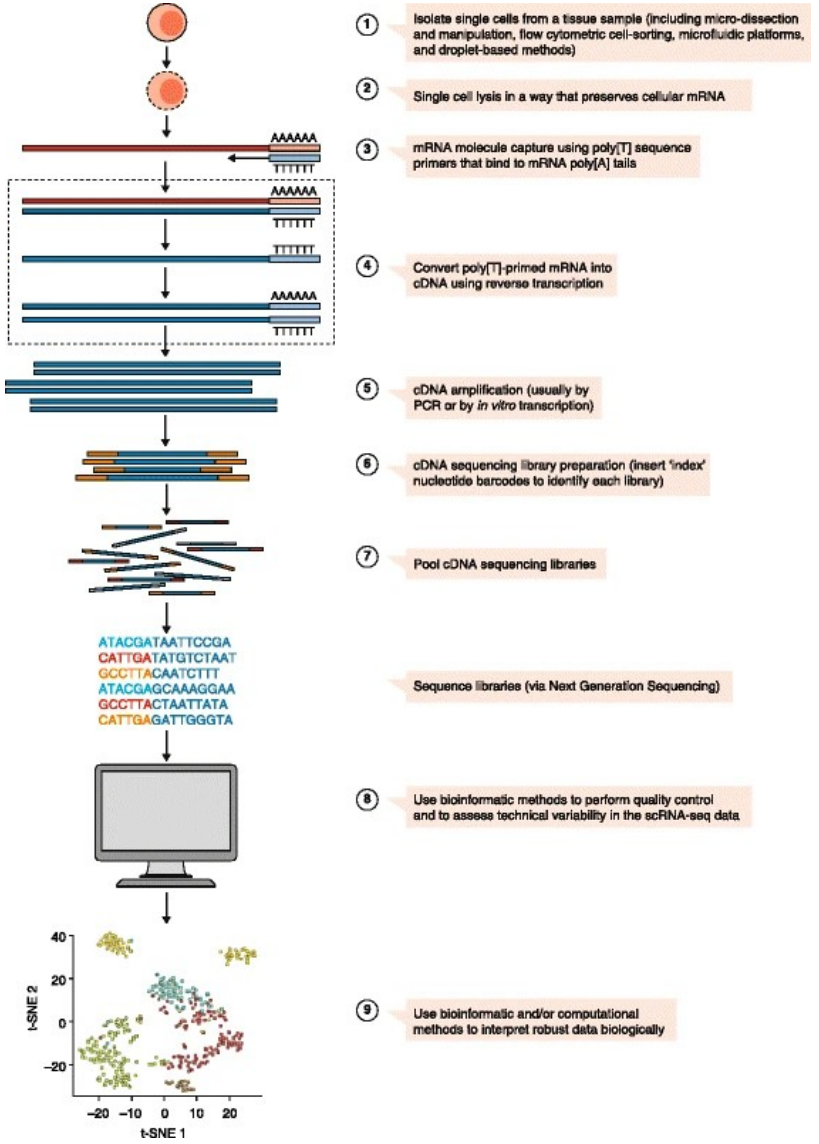

***Figure 1**: General workflow of single-cell RNA-sequencing (scRNA-seq) experiments (Haque et al. Genome Med 2017)*

02.Data availability

DLPFC data by 10x Visium

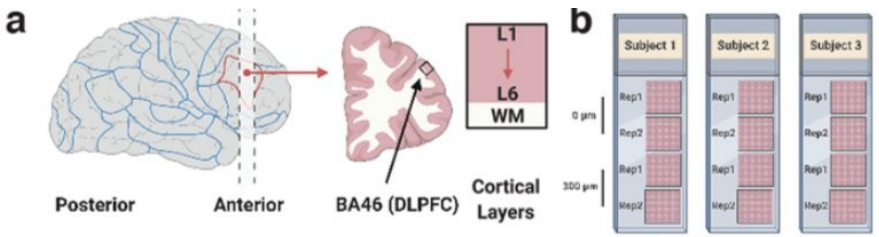

***Figure 3**: (a) Tissue blocks for experiments. (b) Schematic representation of experiments. (Maynard et al. Nat Neurosci 2021)*

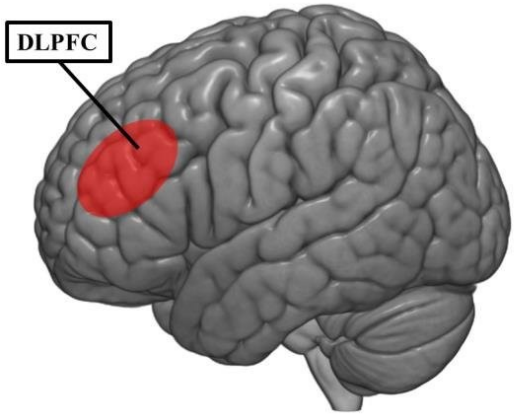

***Figure 4**: Location of the Dorsolateral Prefrontal Cortex (Duprat 2017)*

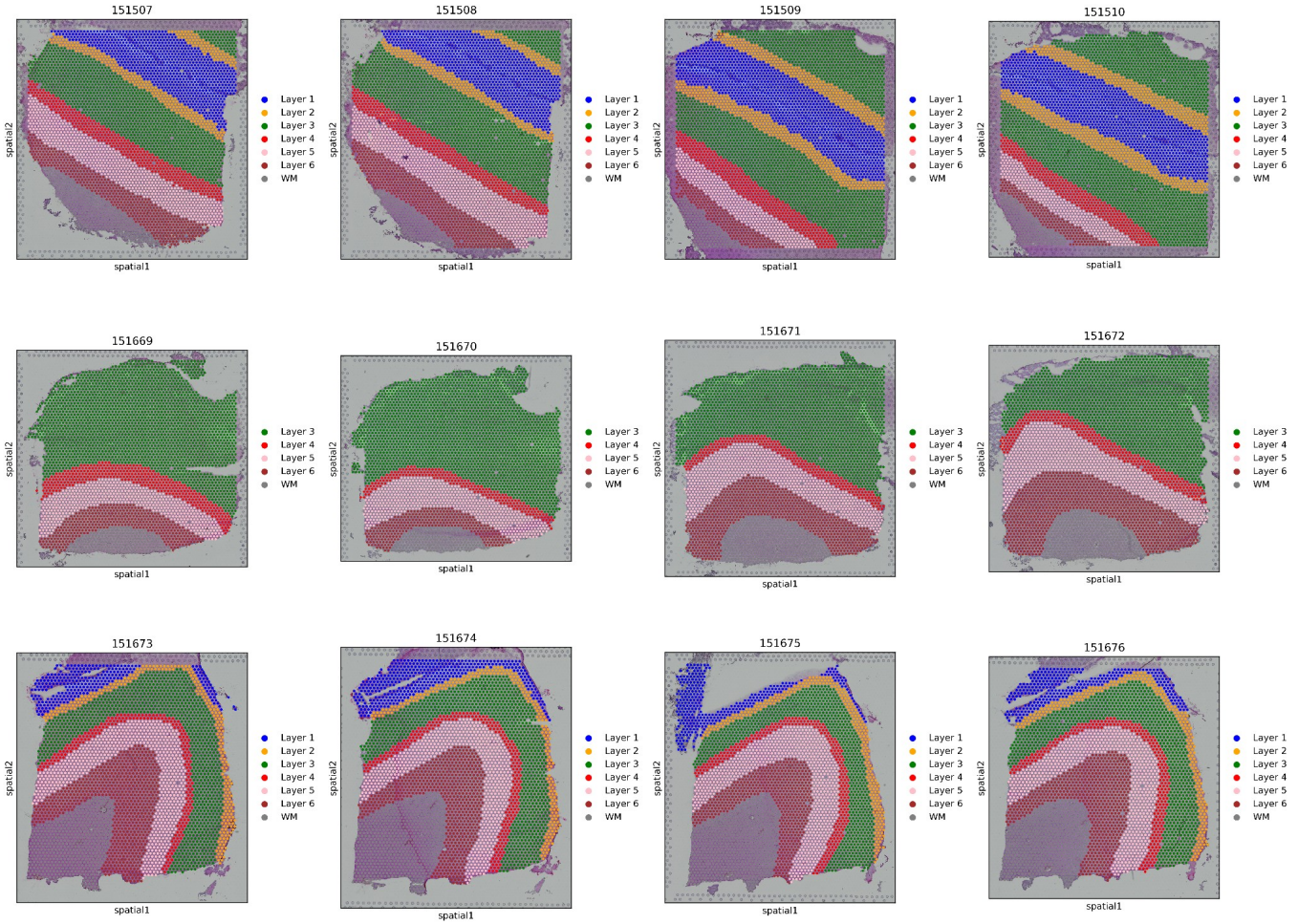

***Figure 5**: 12 samples from the DLPFC dataset by (Maynard et al. Nat Neurosci 2021) showing the 6 regions plus white matter.*

Visualization of the Maynard 2021 dataset in SODB: https://gene.ai.tencent.com/SpatialOmics/dataset?datasetID=6

Paris Brain Institute ICM

04.Proposed method

Graph with Optimal Transport, Transformers and Time diffusion (GO3T)

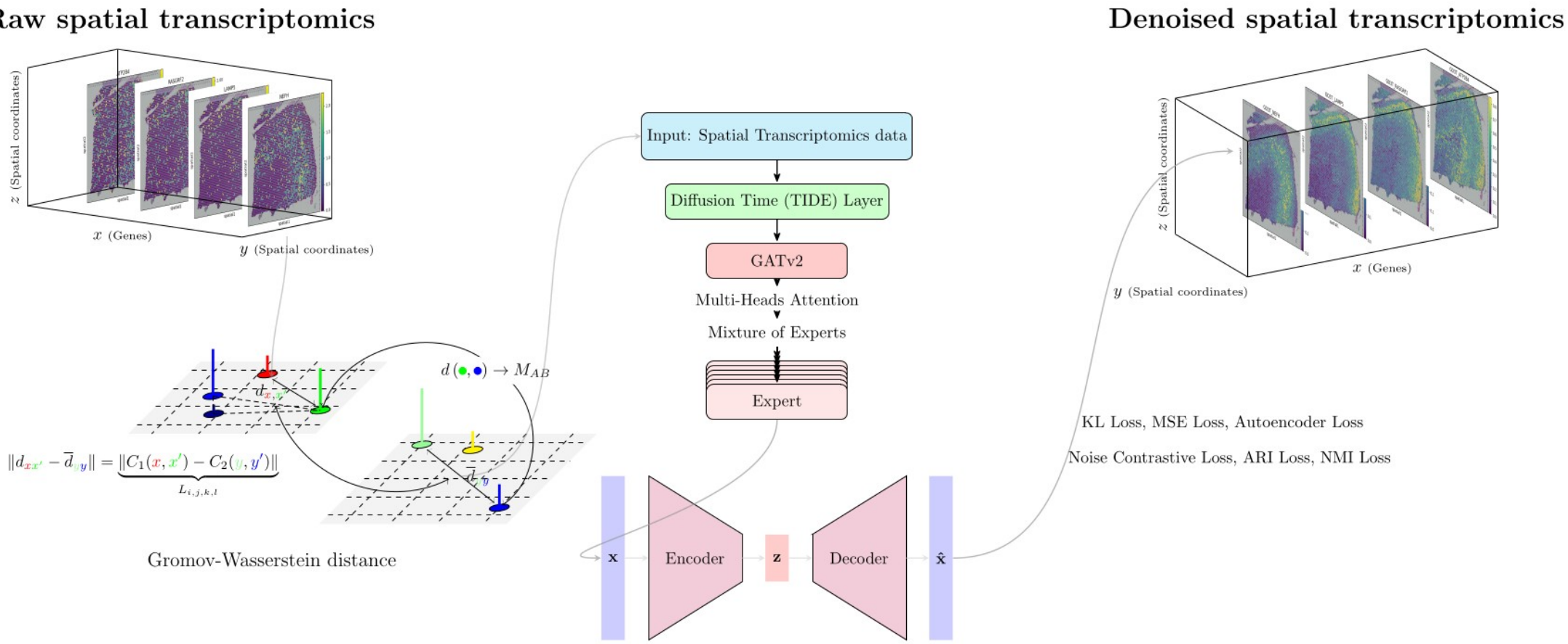

***Figure 6**: overview of GO3T architecture.*

05.Results

Boxplots for metrics comparison

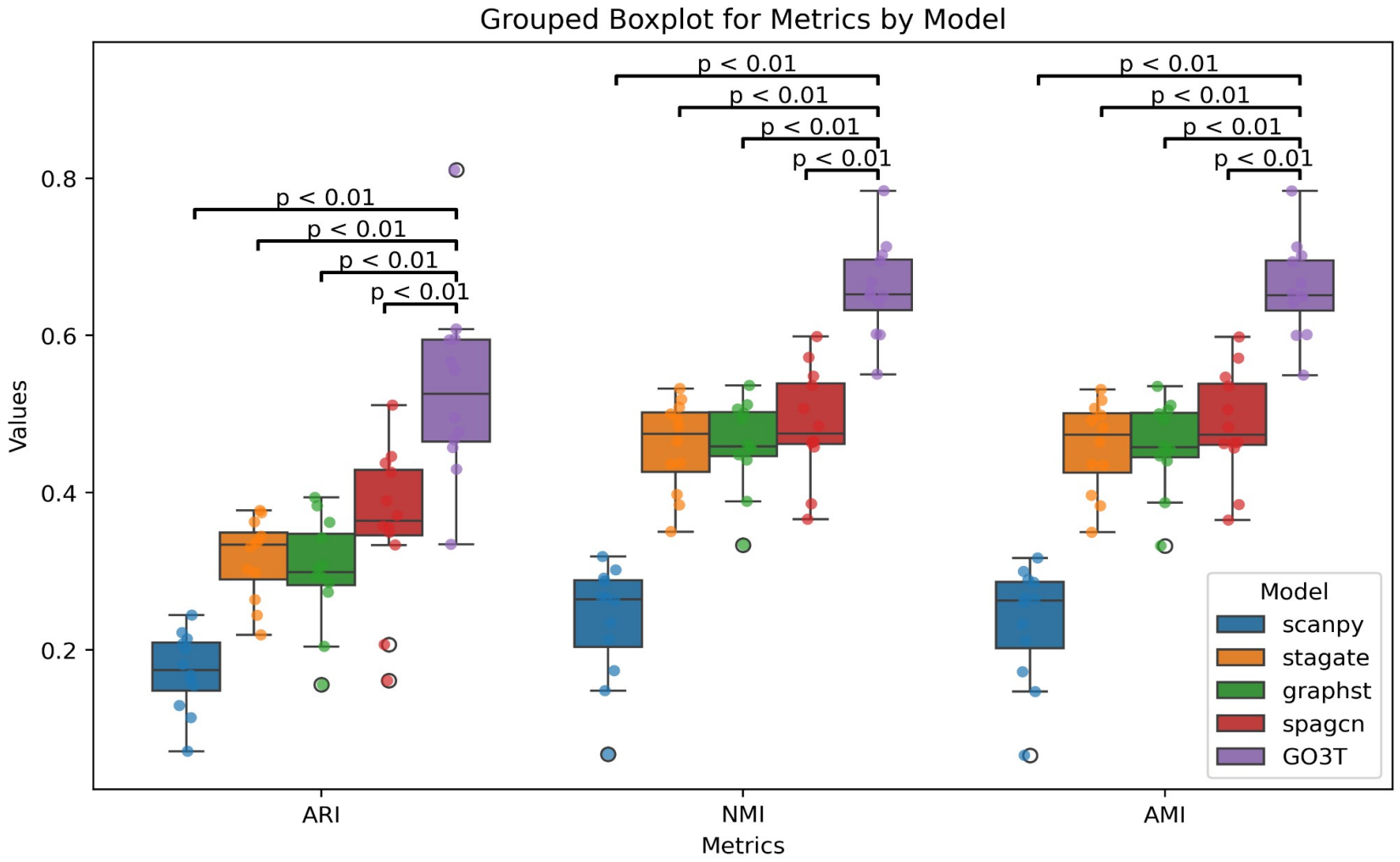

***Figure 7**: boxplots grouped by the three metrics used (ARI, NMI and AMI) for the 12 DLPFC datasets along with the t-test value comparison between four common methods and ours (GO3T).*

Other comparisons for metrics and gene regions

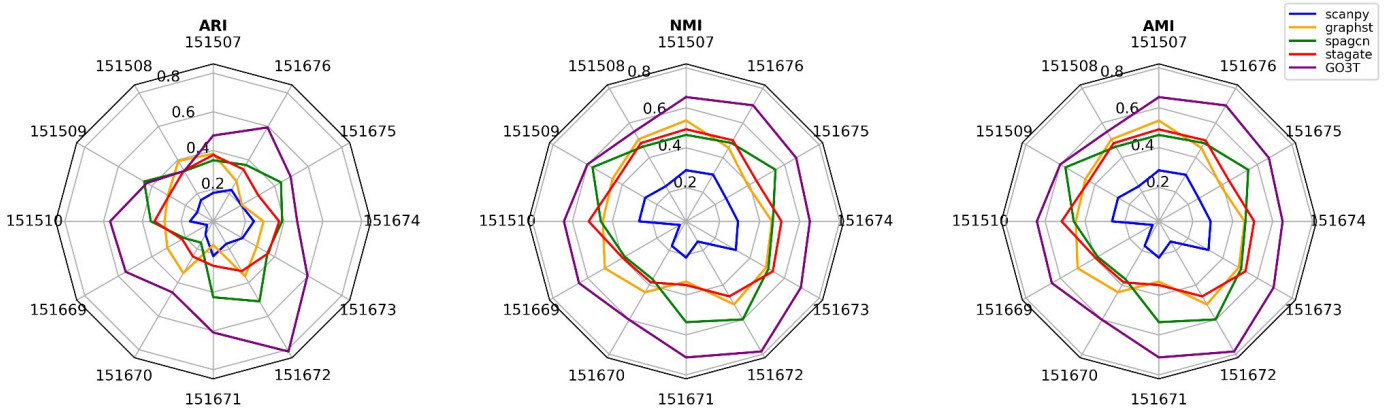

***Figure 8**: radarplots grouped by the three metrics used (ARI, NMI and AMI) for the 12 DLPFC datasets using four common methods and ours (GO3T).*

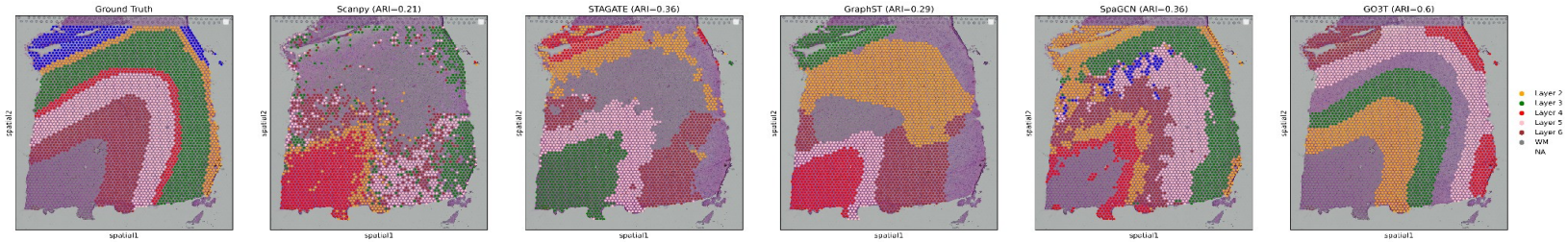

***Figure 9**: spatial plots for the 151673 dataset showing the ground truth regions along with the predicted ones for 4 common methods and ours (GO3T).*

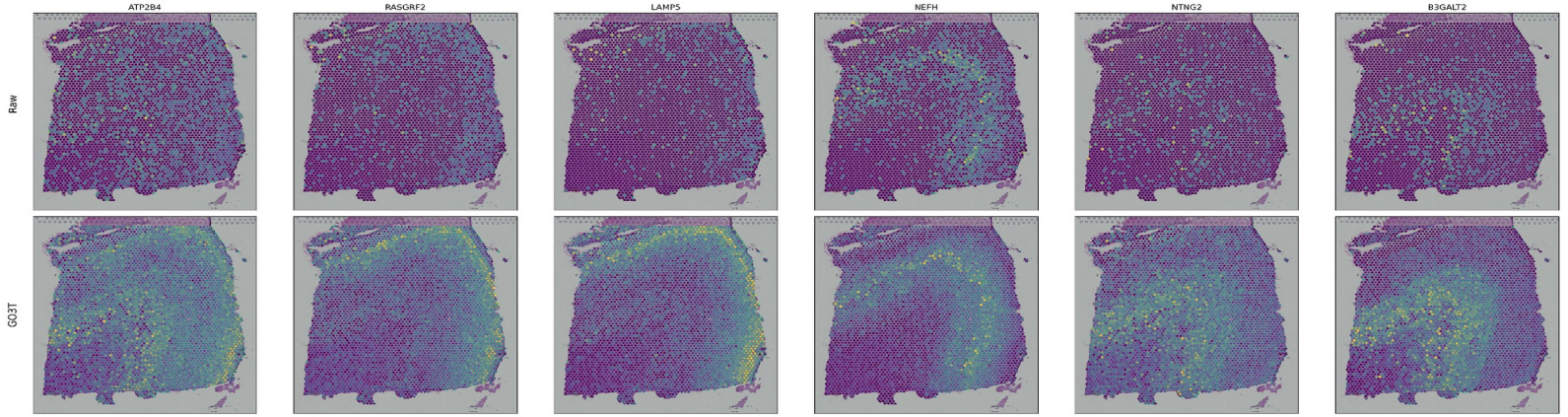

***Figure 10**: spatial plots comparing 6 gene expression using scanpy (above) and our method GO3T (below) for denoising in the 151673 dataset.*

Paris Brain Institute ICM

Acknowledgments

Thank you for your time

- Any questions ? Join me at Poster #18
- Find poster and slides at *lucas-rdlr.github.io*

Agustí ALENTORN - PhD supervisor

References

Haque, A., Engel, J., Teichmann, S.A. *et al.* A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 9, 75 (2017). https://doi.org/10.1186/s13073-017-0467-4

Jiang, R., Sun, T., Song, D. et al. Statistics or biology: the zero-inflation controversy about scRNA-seq data. Genome Biol 23, 31 (2022). https://doi.org/10.1186/s13059-022-02601-5

Li, X., Wang, CY. From bulk, single-cell to spatial RNA sequencing. Int J Oral Sci 13, 36 (2021). https://doi.org/10.1038/s41368-021-00146-0

Virshup, I., Rybakov, S., Theis, F.J., Angerer, F. and Wolf, F.A. bioRxiv 2021 Dec 19. doi: 10.1101/2021.12.16.473007.

Maynard, K.R., Collado-Torres, L., Weber, L.M. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci 24, 425–436 (2021). https://doi.org/10.1038/s41593-020-00787-0

Duprat, Romain. (2017). Accelerated intermittent theta-burst stimulation in major depressive disorder. Focus on the reward system.

Dong, K., & Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nature communications, 13(1), 1739. https://doi.org/10.1038/s41467-022-29439-6

Long, Y., Ang, K.S., Li, M. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun 14, 1155 (2023). https://doi.org/10.1038/s41467-023-36796-3

Hu, J., Li, X., Coleman, K. 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, 1342–1351 (2021). https://doi.org/10.1038/s41592-021-01255-8

Haviv, D., Kunes, R. Z., Dougherty, T., Burdziak, C., Nawy, T., Gilbert, A., & Pe'er, D. (2024). Wasserstein Wormhole: Scalable Optimal Transport Distance with Transformers. ArXiv, arXiv:2404.09411v4.

Krahn, Maximilian & Behmanesh, Maysam & Ovsjanikov, Maks. (2022). TIDE: Time Derivative Diffusion for Deep Learning on Graphs. 10.48550/arXiv.2212.02483.

Wu, X., Huang, S., Wang, W., and Wei, F. Multi-Head Mixture-of-Experts (2024). https://doi.org/10.48550/arXiv.2404.15045

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. and Bengio, Y. Graph Attention Networks (2018). https://doi.org/10.48550/arXiv.1710.10903

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W. and Sun, Y. Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification (2021). https://doi.org/10.48550/arXiv.2009.03509