

ENABLING SINGLE-CELL RESOLUTION IN SPATIAL TRANSCRIPTOMICS FOR COMPREHENSIVE CELLULAR PROFILING WITH VARIATIONAL AUTOENCODERS AND OPTIMAL TRANSPORT

Long Yuan*, Xuyang Li, Janis Taube, Alexander Szalay. *Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Background Spatial transcriptomics (ST) is a promising technique for understanding intercellular dynamics within their spatial context. However, existing ST technologies lack the ability to profile at the single-cell level.^{1–2} Here we propose a method that combines optimal transport (OT) with variational autoencoder (VAE)-embedded latent spaces, allowing us to translate information from single-nuclei images obtained from the standard H&E imaging in the ST pipeline to RNA expression profiles.^{3–9} Thereafter, we can achieve ‘self-deconvolution’ and extrapolation from ST data.

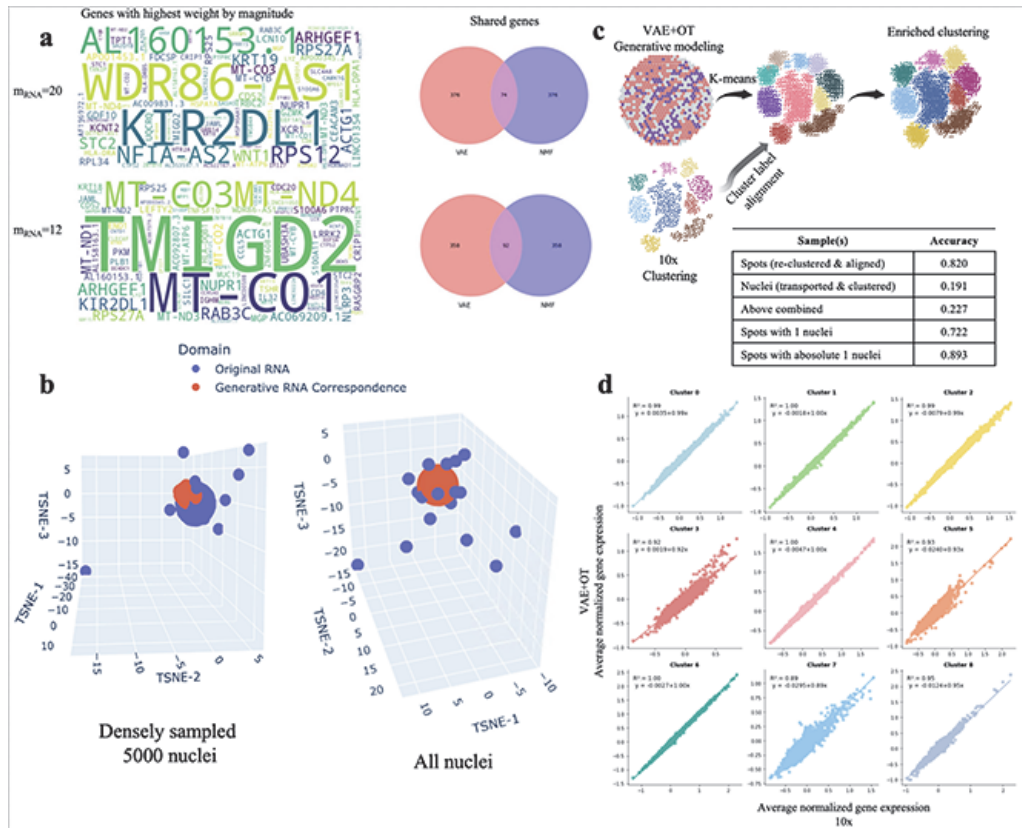
Methods We analyzed 219,096 single nuclei from a breast cancer sample using 10x Visium and StarDist for segmentation. To determine the optimal latent dimensions, we employed various intrinsic dimensionality (ID) detection methods on single-nuclei images and pre-processed transcriptomic data.^{10–13} We developed a Sequencing-VAE with an auxiliary classification task to extract spot identity features and an Imaging-VAE with a nuclei painting proxy task to distill meaningful nuclei morphological features. Through Monge mapping, we translated single-nuclei images into coupling points in transcriptomic latent spaces, which could be decoded by the Sequencing-VAE to generate RNA profiling correspondence.

Results We highlighted the importance of selecting optimal latent dimensions to extract meaningful information from the ambient spaces. Choosing minimal intrinsic dimensions resulted in higher concordance of gene importance compared to a non-negative matrix factorization (NMF)-based method (92/450 versus 74/450) (figure 1a). It also facilitated the sensible distribution of original spot-based sequencing data with RNA profiles from densely-sampled nuclei (figure 1b). The generated single-nuclei transcriptomic profiles exhibited a strong correlation with the original spot-level sequencing data (average correlation coefficient: 0.96) (figure 1d) while capturing cell-level heterogeneity (Jaccard index for spots with 1 nuclei versus more than 1 nuclei: 0.893 versus 0.191) (figure 1c).

Conclusions Our research highlights the valuable information embedded within nuclei morphologies, which can be extracted and translated into gene expression through deep learning and proper mapping functions. This is evidenced by the strong correlation observed between the translated RNA samples and the original RNA samples. Our approach enables higher spatial resolution profiling of the tissue and captures heterogeneity within spots containing multiple nuclei. It also showcases its potential for deconvoluting spot-level RNA sequencing data into single-cell resolution using more informative cell imaging techniques such as multiplexed immunofluorescence (mIF). Considering the emerging use of mIF in precision medicine and its relative cost-effectiveness, our approach opens up possibilities for extrapolating localized gene expression profiles to larger tissue regions profiled with mIF.

REFERENCES

1. SK Longo, MG Guo, AJ Ji, PA Khavari. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics, *Nat. Rev. Genet.* 2021;**22**:627–644.
2. V Svensson, A Gayoso, N Yosef, L Pachter. Interpretable factor models of single-cell RNA-seq via variational autoencoders, *Bioinfo.* 2020;**36**:3418–3421.
3. KD Yang, *et al.* Predicting cell lineages using autoencoders and optimal transport, *PLOS Comp. Biol.* 2020.
4. U Schmidt, *et al.* Cell detection with star-convex polygons, In Proceedings of the 21st ICMICCAI, Granada (2018).
5. M Weigert, *et al.* Star-convex polyhedra for 3D object detection and segmentation in microscopy, *The IEEE Winter Conference on Applications of Computer Vision* 2020.
6. J Bac, *et al.* Scikit-Dimension: A Python Package for intrinsic dimension estimation, *Entropy* 2021;**23**:1368.
7. L Albergante, J Bac, A Zinovyev. Estimating the effective dimension of large biological datasets using Fisher separability analysis, In Proceedings of the 2019 IJCNN, Budapest 2019:1–8.
8. K Johnsson, C Soneson, M Fontes. Low bias local intrinsic dimension estimation from expected simplex skewness, *IEEE Trans. Pattern Anal. Mach. Intell.* 2015;**37**:196–202.
9. E Facco, M D’Errico, A Rodriguez, A Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information, *Sci. Rep.* 2017;**7**:12140.
10. P Grassberger, I Procaccia. Measuring the strangeness of strange attractors, *Phys. D Nonlinear Phenom.* 1983;**9**:189–208.
11. E Levina, PJ Bickel. Maximum Likelihood estimation of intrinsic dimension, In Proceedings of the 17th NeurIPS, Vancouver 2014:777–784.
12. V Little, M Maggioni, L Rosasco. Multiscale geometric methods for data sets I: Multiscale SVD, noise and curvature, *Applied and Computational Harmonic Analysis* 2017;**43**:504–567.
13. A Deshpande, *et al.* Uncovering the spatial landscape of molecular interactions within the tumor microenvironment through latent spaces, *Cell Syst.* 2023;**19**:285–30



Abstract 1316 Figure 1

<http://dx.doi.org/10.1136/jitc-2023-SITC2023.1316>