

Supplementary information for RNA2seg: a generalist model for cell segmentation in image-based spatial transcriptomics

1 Supplementary Methods

1.1 Manual annotation protocol

To generate a high-quality validation set for RNA2seg, we manually annotated cells across five MERFISH datasets from different organs: breast, colon, melanoma, and lung from VHFI, as well as the MERFISH Mouse Ileum dataset [1]. In total, 724 cells were annotated from regions with both high- and low-quality staining (see Supplementary Table 2 for the number of annotated cells per dataset). Manual annotations were performed using the Napari software, leveraging a combination of available membrane stainings, DAPI, and the spatial distribution of RNAs. RNA signals were visualized using the co-expression colorization method introduced in our manuscript, enabling annotators to delineate cell boundaries with high confidence even in regions with weak or ambiguous staining. RNA co-expression helped us to refine boundaries and resolve ambiguities where structural signals were incomplete. The goal of this manual annotation was not to exhaustively label all cells but rather to generate a high-quality validation set. Therefore, only cells with clearly identifiable contours based on the combined signals were annotated.

Multiple annotators carried out annotations across a wide range of organ types and species (see Supplementary Table 2). To ensure the annotated cells were representative of the broader population and to minimize selection bias, we randomly selected crops of size 1200×1200 pixels from the tissue samples. Within each crop, cells were chosen from both densely and sparsely populated regions and across various morphological contexts. Additionally, to avoid bias related to staining performance, which may vary depending on cell type, we included cells from regions with low staining quality and high staining quality, respectively, provided the RNA-based colorization enabled unambiguous identification of cell boundaries.

1.2 Reproducibility analysis

To further evaluate the impact of different segmentation methods on downstream single-cell analysis, we performed a reproducibility analysis. The purpose of this analysis is to assess whether a segmentation method produces consistent results on similar tissue regions.

Method.

We followed the procedure introduced in [2]. In brief, we sampled 200 square crops (each 120μ square) from the publicly available Merscope lung dataset. To group tissue crops with similar cell-type composition, we assumed that crops sharing comparable RNA expression profiles originate from similar tissue type. Therefore, the crops were clustered by their RNA profiles using k -means, in a manner analogous to clustering bulk RNA-seq samples. The number of clusters was set to $K_{\text{crop}} = 6$ based on the silhouette score. For each segmentation method, we computed a cell-by-gene expression matrix for every crop. Similarly than for the crops, each segmented cell was then clustered based on its RNA profile into $K_{\text{cell-type}} = 6$ groups, yielding a representation of the cell-type

composition of the crop. Then, following [2], each crop was represented by a cell-type vector V_c : a vector of length $K_{\text{cell-type}}$ whose entries correspond to the number of cells of each cell type in the crop.

Reproducibility measure.

Finally, like in [2], for crops assigned to the same RNA profile cluster, we measured reproducibility by computing the Pearson correlation of the corresponding V_c vectors across segmentation methods. The reproducibility of the 4th best method for IoU metrics and GeneSegNet [2] are displayed on Supplementary Figure S8. While RNA2seg achieves higher cell-type reproducibility than GeneSegNet, Cellpose, and ProSeg, its performance is comparable to the vendor’s segmentation (VHFI). However, the vendor method may produce more segmentation errors in regions with low-quality staining, as shown in the main manuscript (Figures 3 and 5), potentially introducing a consistent bias. These systematic errors are not accounted for in our reproducibility analysis. Altogether, this analysis further suggests that RNA2seg provides more reproducible results while maintaining superior segmentation quality. The code to reproduce this study is available at https://github.com/tdefa/downstream_reproducibility

2 Supplementary Tables

Dataset name	Imaging approach	number of Cells
VHFI Breast	MERFISH	713,121
VHFI Lung	MERFISH	353,762
VHFI Col on	MERFISH	677,451
VHFI Melanoma	MERFISH	468,138
VHFI Ovarian	MERFISH	358,485
VHFI Prostate	MERFISH	721,668
VHFI Uterine	MERFISH	843,285
CosMx-lung (Lung5_rep1) [3]	CosMx	87955

Table 1 Summary of the datasets used for the training. The number of cells of each VHFI dataset comes from the Vizgen data released. The number of cells for the CosMx dataset is inferred with Cellpose on membrane staining

Dataset	Number of annotated cells
VHFI Breast	185
VHFI Colon	95
VHFI Lung	181
VHFI Melanoma	193
Mouse ileum	70

Table 2 Numbers of manually annotated cells per dataset.

3 Supplementary Figures

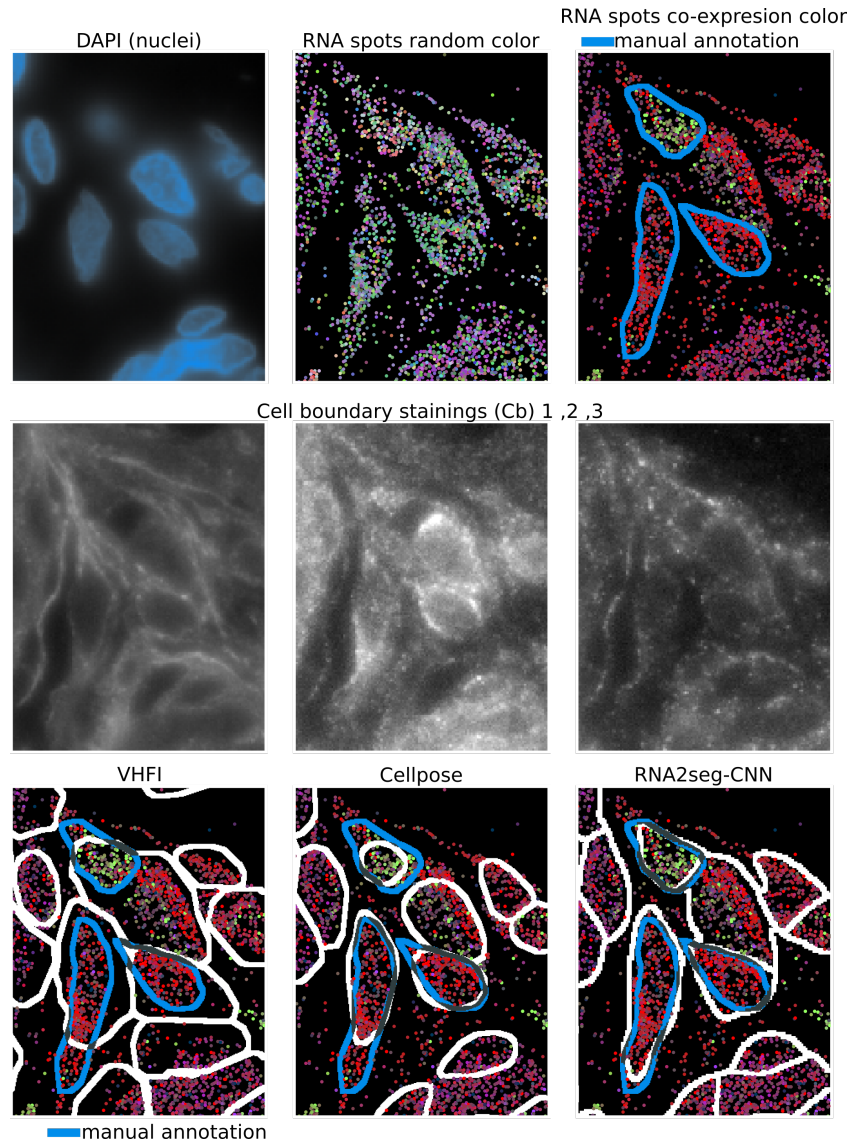


Fig. S1 A) Example of staining and RNA point cloud. From left to right in the first row: DAPI staining, RNA spots with random color, RNA spots with co-expression color and manual annotation in blue. The second row display the three different available stainings in the VHF1 dataset. The third row display examples of segmentation from VHF1, Cellpose, and RNA2seg-CNN

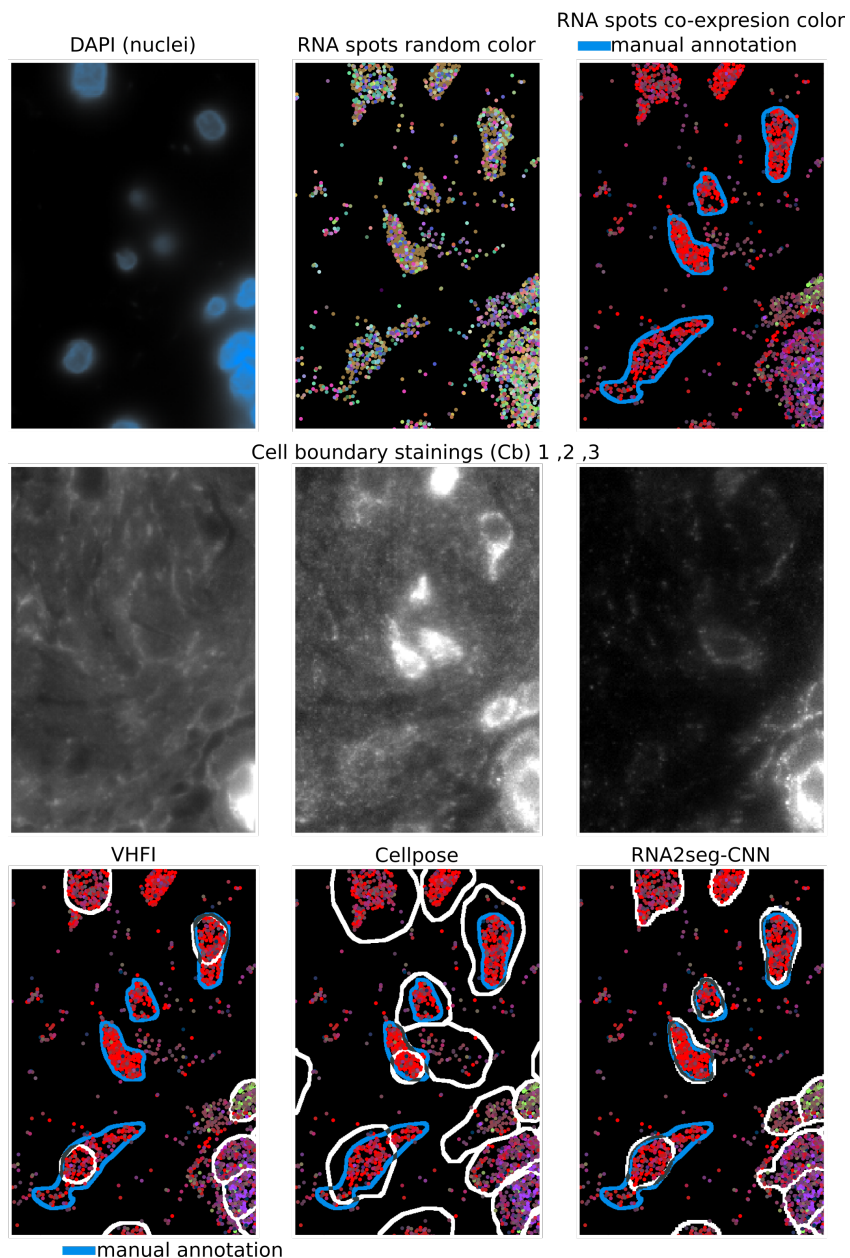


Fig. S2 A) Example of staining and RNA point cloud. From left to right in the first row: DAPI staining, RNA spots with random color, RNA spots with co-expression color and manual annotation in blue. The second row display the three different available stainings in the VHF1 dataset. The third row display examples of segmentation from VHF1, Cellpose, and RNA2seg-CNN

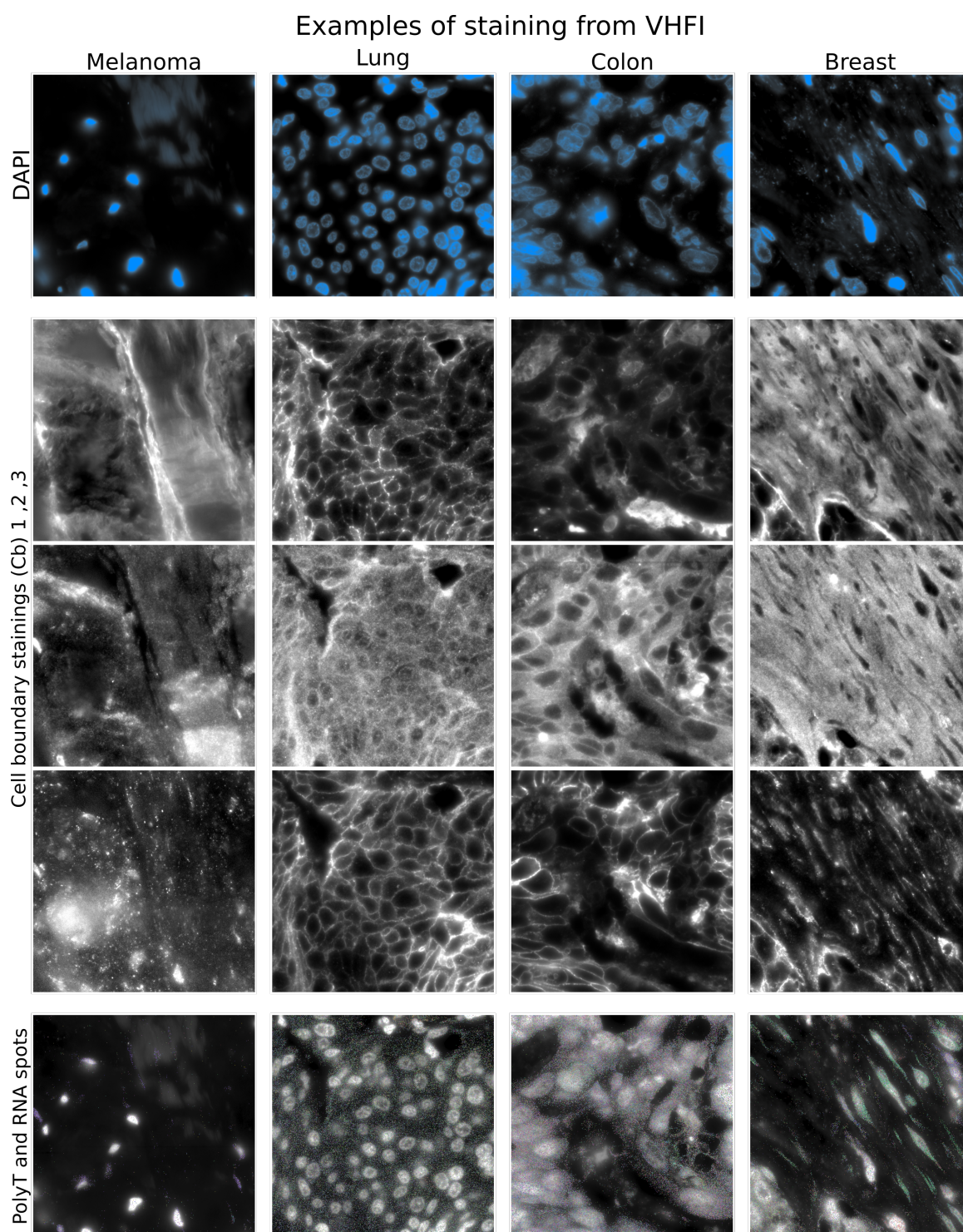


Fig. S3 Examples of staining and detected RNA spots from the VHFI datasets.

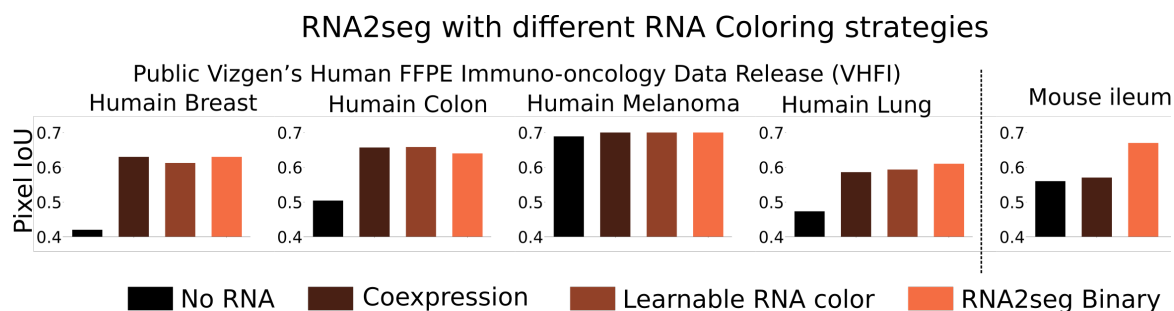


Fig. S4 Benchmark of RNA2seg performances using different RNA representations strategies. All models are trained on the VHFI datasets having the same 500 genes spatially resolved. No RNA: RNA2seg trained solely using stainings. Co-expression: RNA2seg trained on RNA images using our co-expression-based coloring method. Learnable RNA Color: RNA2seg where the color of each RNA species is treated as a learnable parameter during training. However, this approach is not transferable to new datasets and is therefore inapplicable to the mouse ileum in a zero-shot setting. Binary: The RNA2seg version retained in this study, trained using only positional information.

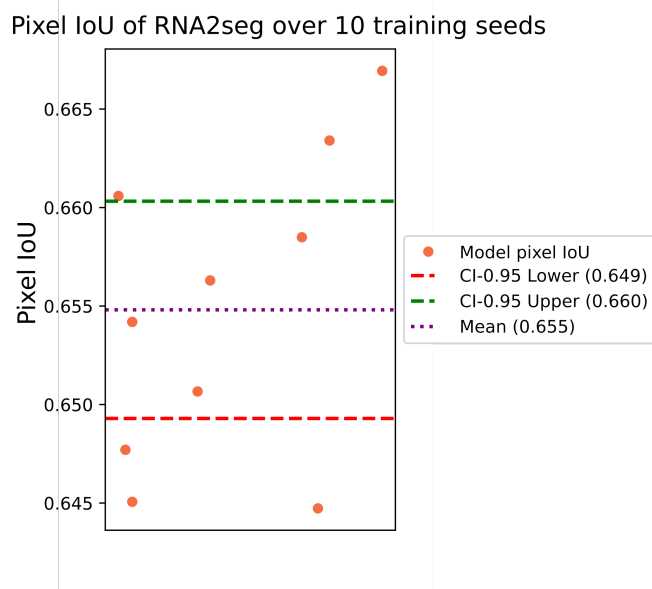


Fig. S5 Pixel IoU of RNA2seg-CNN averaged over 10 independent training initializations on all annotated data from the VHFI datasets. The confidence interval (CI) is computed with a t-distribution

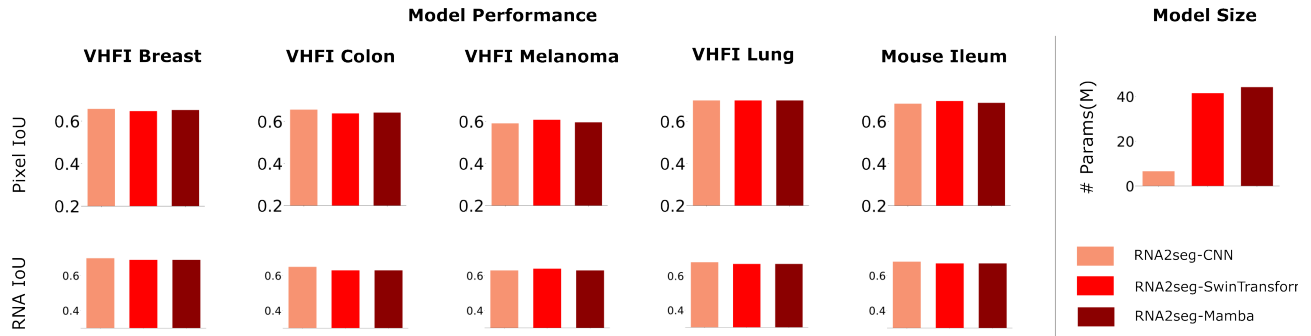


Fig. S6 Benchmark of different RNA2seg architectures across multiple datasets. Each model (CNN-Unet, Swin-Unet, and VM(Mamba)-Unet) uses the Unet backbone, while the underlying architecture varies. Pixel-level and RNA-level IoU are shown for each model, along with the number of parameters per model. Performance was nearly identical across all five organs, indicating that the approaches are essentially equivalent. RNA2seg-CNN was retained as the primary model due to its comparable accuracy, lighter weight, and greater computational efficiency.

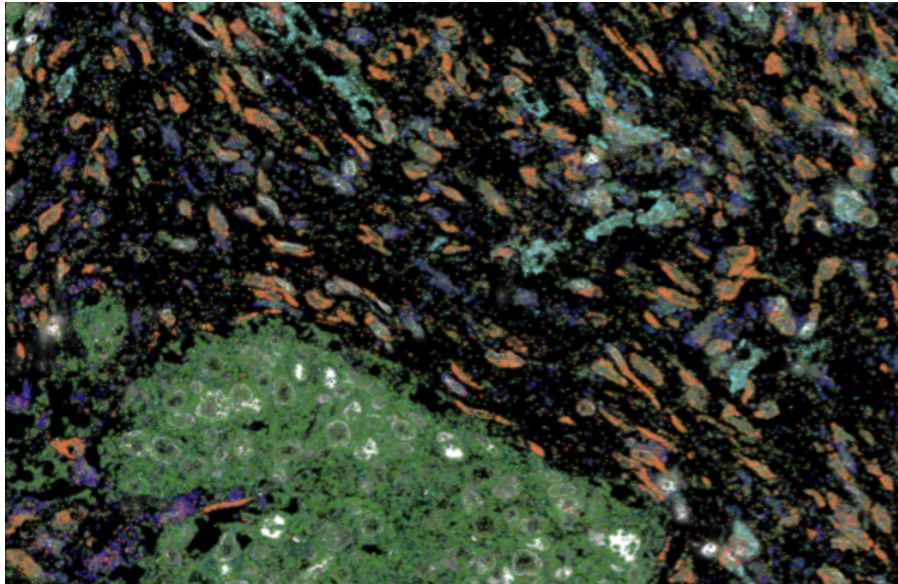


Fig. S7 RNA co-expression visualization of breast tissue (MERSCOPE). RNA species are assigned RGB colors based on their co-expression vectors, with co-expressed genes appearing in similar colors. Heterogeneous regions show distinct local expression patterns, enabling visual separation of neighboring cells (e.g., purple, orange, blue), whereas homogeneous regions appear uniformly colored, reflecting similar expression profiles (green area).

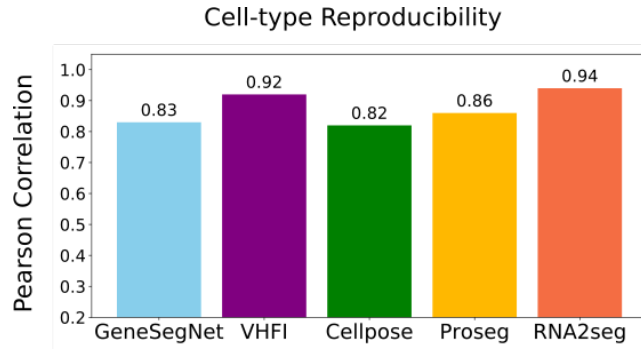


Fig. S8 Reproducibility analysis across segmentation methods. Reproducibility was quantified for each segmentation method by computing the Pearson correlation of cell-type vectors V_c across crops assigned to the same tissue type. Bars represent the mean correlation values of the cell-type vectors V_c of the same sample for each method: GeneSegNet (light blue) VHFI (purple), Cellpose (green), Proseg (yellow), and RNA2seg (orange). Higher values indicate greater consistency of the segmentation method in reproducing similar single-cell representations across similar tissue regions.

References

- [1] Petukhov, V., Xu, R. J., Soldatov, R. A. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nature Biotechnology* **40**, 345–354 (2022). URL <https://doi.org/10.1038/s41587-021-01044-w>.
- [2] Wang, Y., Wang, W., Liu, D. *et al.* Genesegnet: a deep learning framework for cell segmentation by integrating gene expression and imaging. *Genome Biology* **24**, 235 (2023). URL <https://doi.org/10.1186/s13059-023-03054-0>.
- [3] He, S., Bhatt, R., Brown, C. *et al.* High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature Biotechnology* **40**, 1794–1806 (2022). URL <https://doi.org/10.1038/s41587-022-01483-z>.