

Supplementary Material:

SpaCut: Transcript-aware Morphology Fusion for Robust Cell
Segmentation in Subcellular Spatial Transcriptomics

Contents

1	Table	2
2	Figure	4
3	Baselines	8

1 Table

Table S1: Runtime (in seconds) and Cell Count on **Stereo seq one patch**

	SCS	SpaCut	Cellpose	UCS	Cellpose-SAM
Runtime (s)	8126.18	435	10.8	418	16.8
Cell Count	875	1133	950	1120	1017

Table S2: Runtime (in seconds) and Cell Count on **CosMx Pancreas fov51**

	Proseg	SpaCut	Cellpose	UCS	Cellpose-SAM
Runtime (s)	763	532	24.94	688	29.8
Cell Count	2540	2514	2667	2520	2679

Table S3: Cell Count on three Xenium datasets

Method	Human Lung Cancer	Human Pancreas	Mouse Colon
StarDist	214235	214235	356014
DeepCell	189276	189276	280116
Cell 10X (GT)	162248	140701	219795
MEDIAR	160532	160532	221992
Cellpose3	158028	158028	214610
Nuclei 10X	156612	136520	206898
SpaCut (Ours)	156591	136446	206808
Baysor (prior)	155139	155139	213335
UCS	154574	134729	205090
Cellpose	150070	137412	201714
Cellpose-SAM	148870	136360	136360
SAM	125916	125916	114190
Baysor (no prior)	110508	98823	107452

To evaluate the contributions of the NGAF and GWA modules to SpaCut, we conducted a series of ablation experiments on three Xenium datasets (Table S4). We specifically aimed to determine how different designs of nucleus-guided refinement and gene-wise encoding affect segmentation accuracy.

The experiments included a baseline variant without either Spatial Attention or DSC_CA (None), two NGAF variants, and two GWA variants. For NGAF, we first removed the Spatial Attention block (w/o SA) and then further removed the concatenation between the modulated nucleus mask and the original nucleus mask (w/o concat). Removing SA already leads to consistent degradation: on Xenium Human Lung Cancer, AJI decreases from 0.7465 to 0.7363 and F1 from 0.9395 to 0.9236 (drops of 1.0 and 1.6 points), with similar but slightly smaller drops on the Xenium FFPE Human Pancreas and Mouse Colon datasets. Dropping the concatenation step causes additional losses, especially on Xenium Mouse Colon, where AJI falls to 0.5934 and F1 to 0.8820 (2.2- and 1.6-point drops relative to the full model). These results indicate that combining raw and modulated nuclear priors, together with spatially selective aggregation, is crucial for robust boundary refinement.

For GWA, we first ablated the Depthwise Separable Convolution with Channel Attention (w/o DSC_CA), which introduces noticeable performance degradation across all datasets. For example, on

Table S4: Ablation studies on NGAF and GWA modules across three Xenium datasets.

Method	Xenium Human Lung Cancer		Xenium FFPE Human Pancreas		Xenium Mouse Colon	
	AJI	F1 Score	AJI	F1 Score	AJI	F1 Score
w/o SA (NGAF)	0.7363	0.9236	0.7404	0.8859	0.6013	0.8834
w/o concat (NGAF)	0.7258	0.9284	0.7437	0.8913	0.5934	0.8820
w/o DSC_CA (GWA)	0.7284	0.9148	0.7351	0.8764	0.5938	0.8745
w/o fusion (GWA)	0.6805	0.8871	0.7351	0.8652	0.5910	0.8865
None (no SA / DSC_CA)	0.7228	0.9107	0.7384	0.8804	0.5842	0.8621
Ours (Full)	0.7465	0.9395	0.7451	0.8920	0.6151	0.8980

Xenium Human Lung Cancer, AJI and F1 drop by 1.8 and 2.5 points, respectively. When we further remove the feature fusion block and directly use only the deepest gene-wise feature map as the encoder output (w/o fusion), performance deteriorates most severely among all single-component variants: on Xenium Human Lung Cancer, AJI decreases from 0.7465 to 0.6805 and F1 from 0.9395 to 0.8871 (6.6- and 5.2-point drops). This confirms that the dedicated fusion block, which aggregates intermediate gene-wise representations, is critical for decoding sparse transcript signals.

When both SA and DSC_CA are removed simultaneously (None), the model suffers additional degradation on all datasets. The Xenium Mouse Colon dataset is the most sensitive, with AJI dropping from 0.6151 to 0.5842 and F1 from 0.8980 to 0.8621 (3.1- and 3.6-point drops). Overall, the ablation experiments verify that each component makes a non-trivial contribution, and the full SpaCut model consistently achieves the best AJI and F1 scores on all three Xenium datasets.

2 Figure

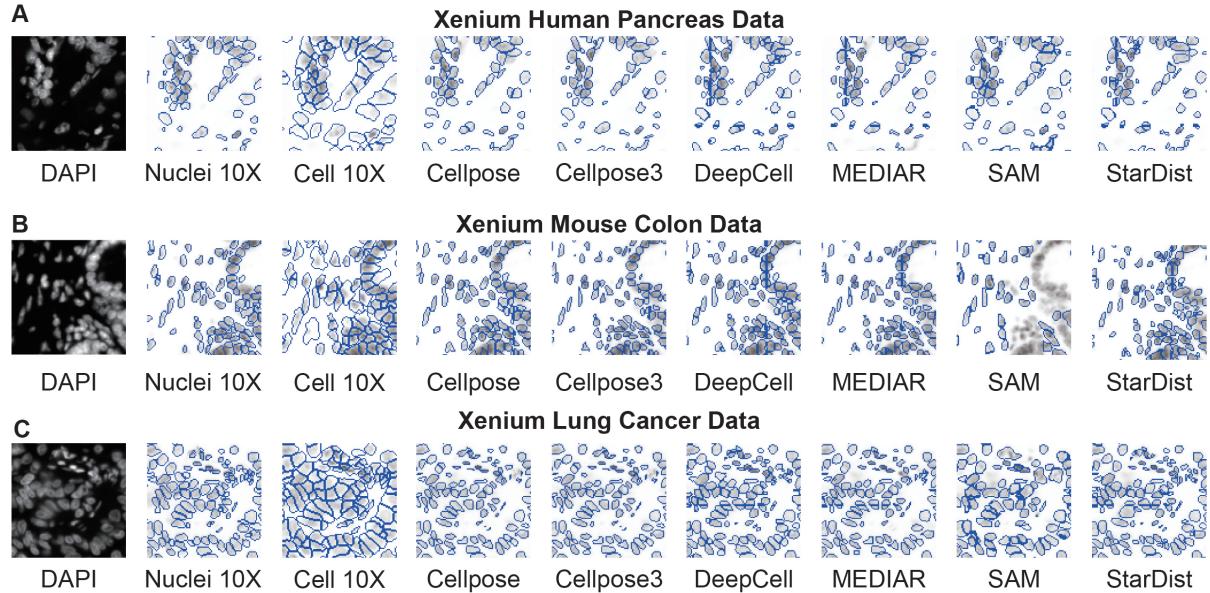


Figure S1: Segmentation Results of image-based models on three Xenium datasets

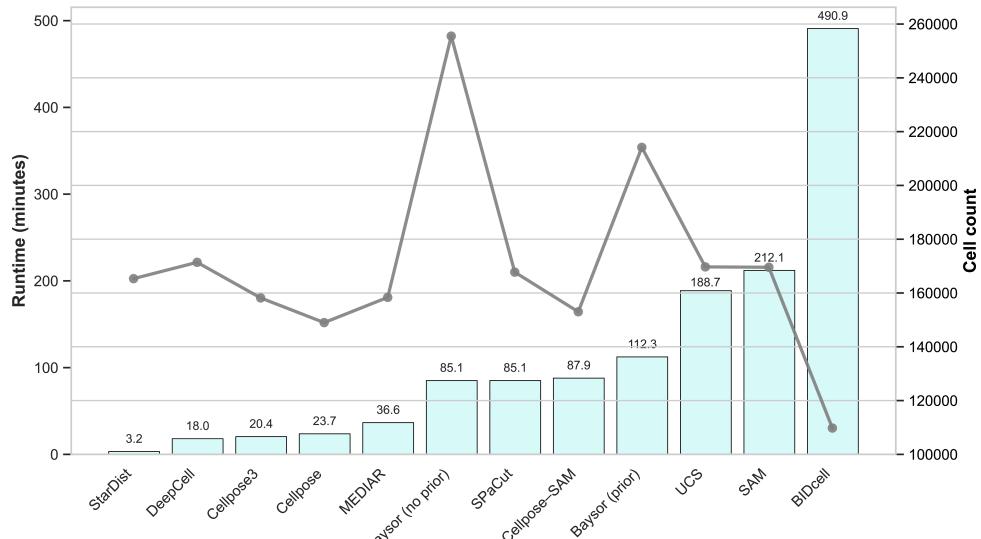


Figure S2: Runtime (bars; left y-axis, minutes) and segmented cell count (grey line; right y-axis) of representative cell-segmentation methods on the Xenium Breast Cancer dataset. Our method, SpaCut, achieves a favorable runtime–count trade-off (competitive cell counts at moderate runtime), highlighting its efficiency relative to alternatives. StarDist is the fastest (3.2 min), BIDCell is the slowest (490.9 min), and Bayso (no prior) attains the highest cell count.

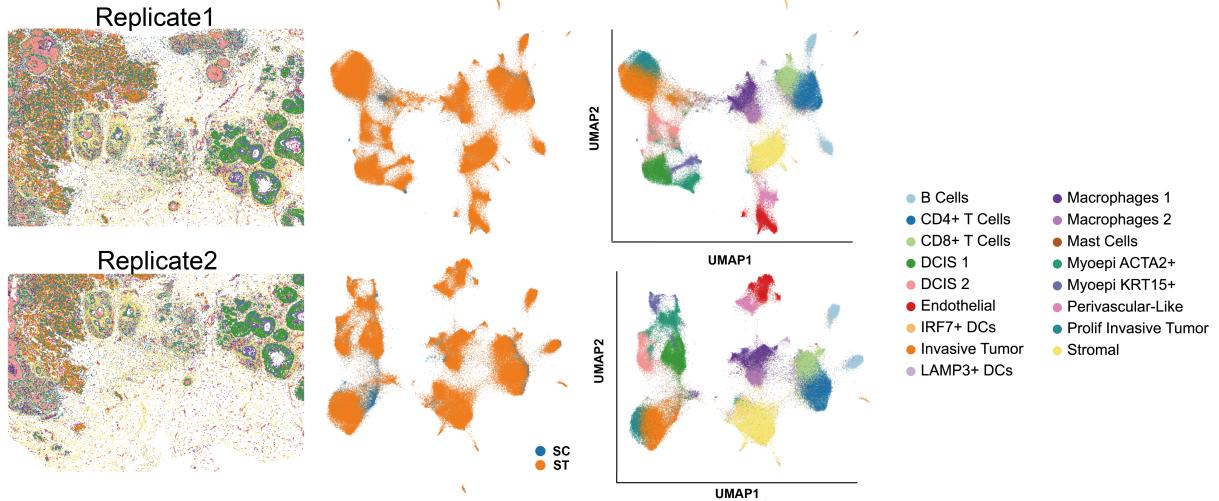


Figure S3: Cross-replicate evaluation on the Xenium Breast Cancer dataset. We train **SpaCut** on Replicate1 and validate on Replicate2 (no additional fine-tuning). Left: tissue-wide spatial cell-type maps rendered from SpaCut-based segmentations and downstream annotation. Middle: joint UMAP embedding of single-cell (SC, blue) and spatial-transcriptomic (ST, orange) profiles, showing that ST embeddings align well to the SC reference. Right: UMAP colored by predicted cell types.

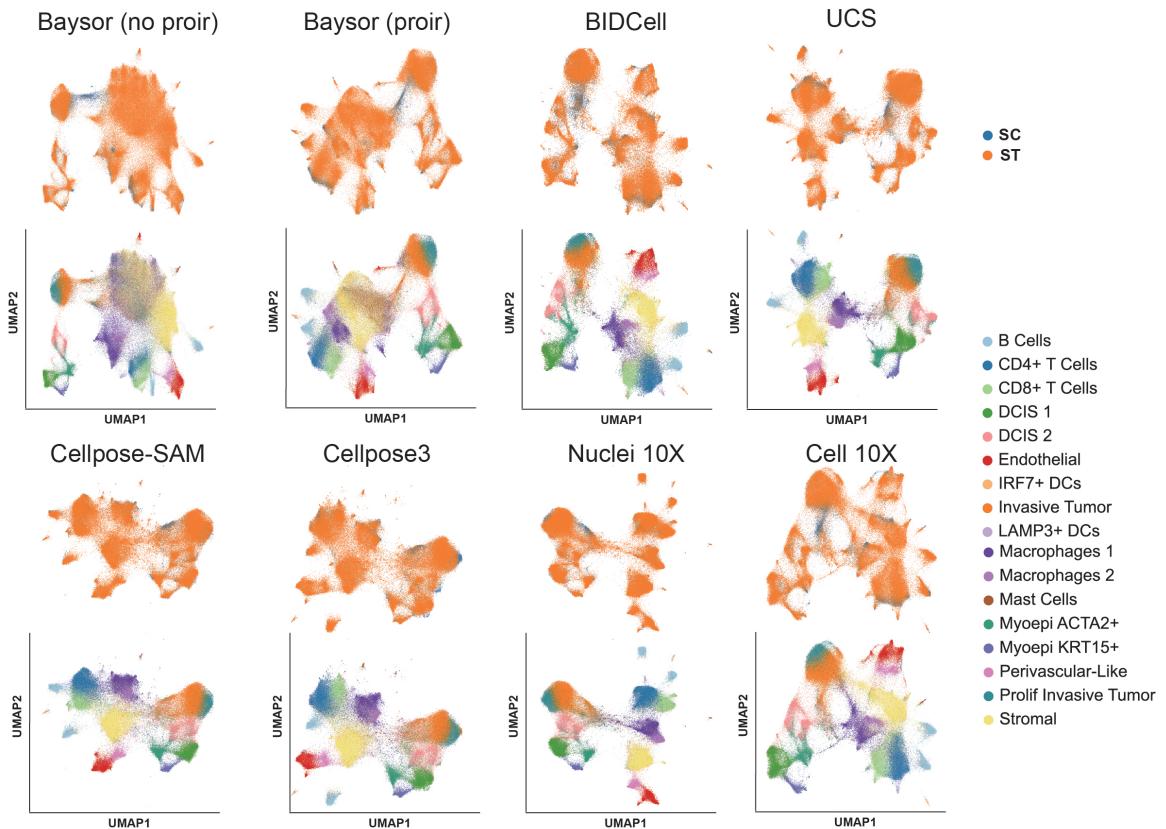


Figure S4: Qualitative comparison of segmentation benchmarks on the Xenium Breast Cancer dataset (Replicate 1). Top: Joint embeddings showing the alignment between single-cell reference (SC, blue) and segmented spatial data (ST, orange). Bottom: Corresponding UMAPs colored by annotated cell types to show clustering resolution.

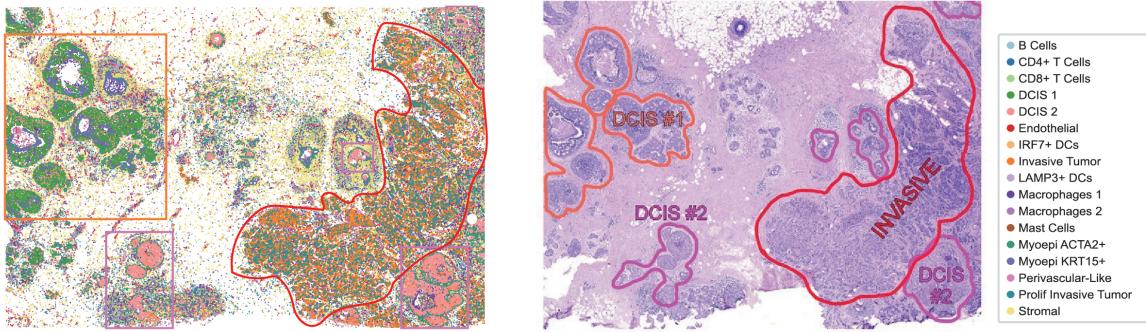


Figure S5: Identification of distinct cancer regions via spatial cell type mapping. The left panel visualizes the algorithmic identification of tumor microenvironments based on cell types. The right panel displays the ground-truth histopathology with annotations for specific cancer subtypes (Invasive, DCIS #1, and DCIS #2). The close correspondence between the cell-type clusters and the manual annotations[1] validates the accurate identification of these cancer regions.

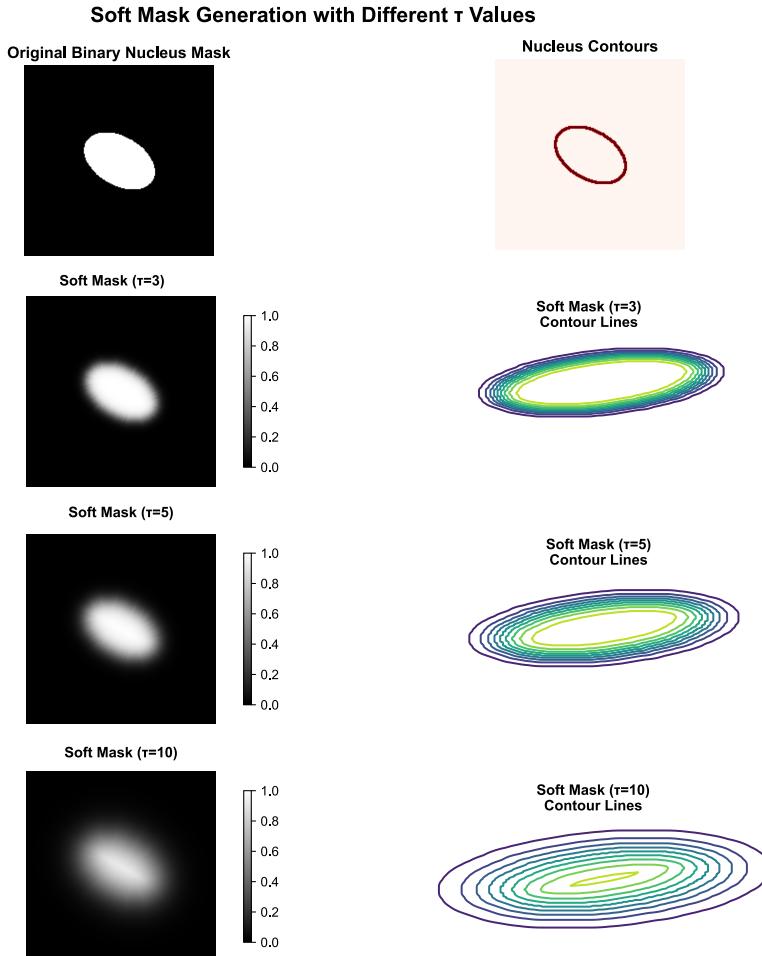
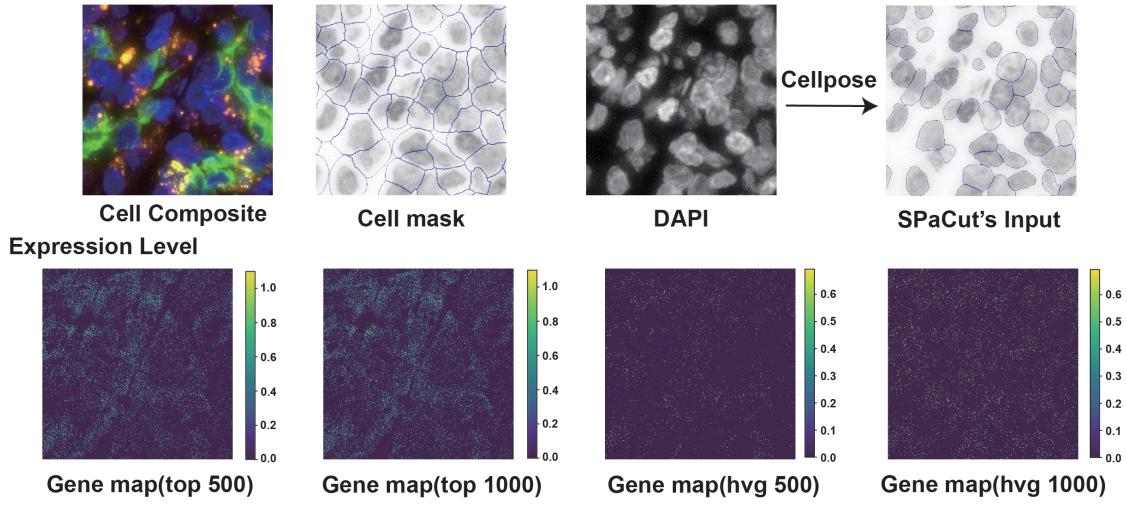
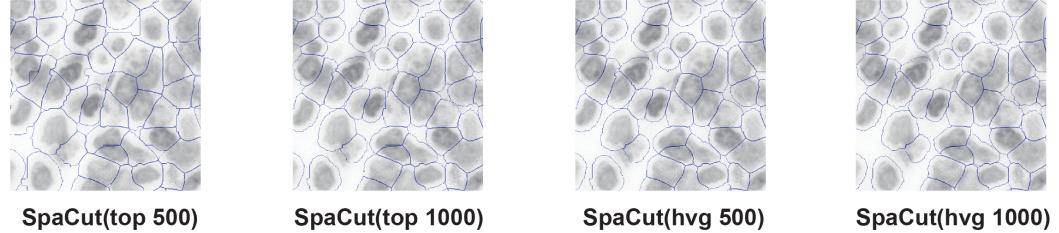


Figure S6: SoftMask generation from a binary nucleus mask with different softness parameters (τ). Left column: the original binary nucleus mask and the resulting SoftMasks for $\tau = 3, 5, 10$ (brighter values indicate higher membership). Right column: corresponding isocontours of the SoftMasks. For a pixel p , the SoftMask value is defined as $\text{SoftMask}(p) = \sigma\left(\frac{d(p)}{\tau}\right)$, where $\sigma(\cdot)$ is the logistic function and $d(p)$ is the signed shortest distance from p to the nucleus contour (positive inside, negative outside; computed via `cv2.pointPolygonTest`). A smaller τ yields a sharper transition that approximates the binary mask, while a larger τ produces smoother, wider halos near boundaries, enabling soft assignment around the nucleus edge.

CosMx Human Pancreas



Segmentation Results



Stereo-seq Mouse Brain

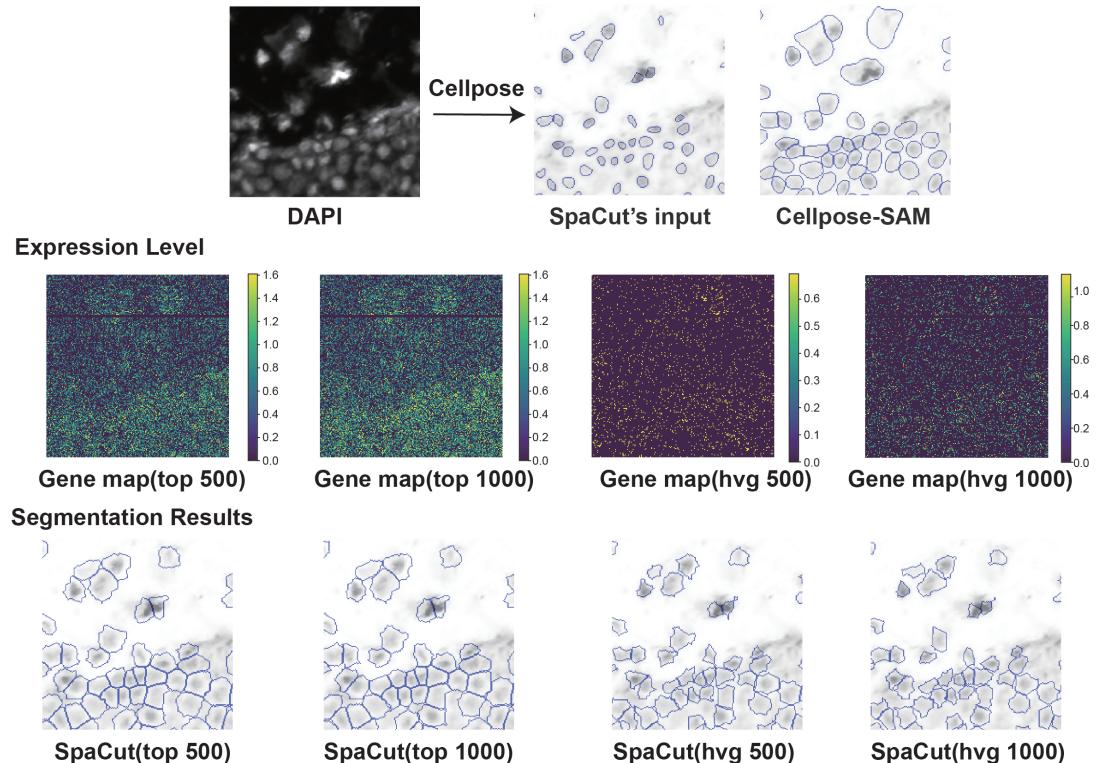


Figure S7: Assessment of gene-set selection for constructing spatial expression maps across two datasets (CosMx Human Pancreas and Stereo-seq Mouse Brain). We compare two criteria—top-expressed genes (top) and highly variable genes (HVG)—at two set sizes (500 and 1000). Visual inspection shows that maps constructed from top-expressed genes exhibit clearer intensity differences between tissue regions than maps constructed from HVG-based gene sets on both datasets. Performance is comparable between top-500 and top-1000; to balance spatial detail and computational cost, we adopt top-500 in subsequent experiments.

3 Baselines

BIDcell [2]: BIDCell is a self-supervised deep learning framework designed for cell segmentation in sub-cellular spatial transcriptomics (SST) data. It innovatively integrates biologically informed loss functions with single-cell RNA-seq references to model the complex relationships between gene expression and cell morphology, thereby accurately delineating cell boundaries even in data with sparse signals.

Proseg [3]: Proseg is a probabilistic cell segmentation method that adapts the Cellular Potts model within a Metropolis–Hastings sampling framework. It infers cell boundaries by optimizing the likelihood of transcript distributions. Notably, it operates without the need for supervised training and incorporates a mechanism to reduce spurious co-expression through transcript repositioning.

SCS [4]: SCS employs a transformer-based neural network to fuse gene expression profiles with high-resolution imaging data. It treats segmentation as a gradient flow tracking problem, where the model learns the relative position of each spot to its corresponding cell center and groups spots into distinct cellular instances based on these spatial relationships.

UCS [5]: UCS utilizes a dual convolutional neural network (CNN) framework specifically designed to integrate nuclei segmentation masks with transcriptomic data. By processing visual and transcripts information in parallel, it enhances the precision of cell segmentation in subcellular spatial transcriptomics.

Baysor [6]: Baysor performs segmentation using a Bayesian mixture model that accounts for both spatial morphology and transcriptional composition. It optimizes 2D or 3D cell boundaries by probabilistically assigning molecules to cells. In this study, the method was implemented via the Sopa framework [7], which provides a standardized interface for spatial omics analysis.

DeepCell [8]: DeepCell employs a deep learning architecture for whole-cell and nuclear segmentation in tissue imaging. It relies on a ResNet50-based backbone combined with a Feature Pyramid Network (FPN), trained on the extensive TissueNet dataset, to capture multi-scale features for robust segmentation.

SAM [9]: The Segment Anything Model (SAM) utilizes a Vision Transformer-based architecture trained on the massive SA-1B dataset (containing over 1 billion masks). It introduces a promptable segmentation paradigm, enabling strong zero-shot generalization to diverse segmentation tasks without task-specific retraining.

Cellpose [10]: Cellpose is a generalist algorithm that employs a U-Net model to predict spatial gradient flow fields. Instead of direct mask prediction, it learns a vector field pointing toward cell centers and defines cellular instances by grouping pixels that converge to common sinks via a gradient tracking process.

Cellpose-SAM [11]: Cellpose-SAM represents a hybrid approach that leverages the foundation model SAM for cellular segmentation. It adapts SAM’s powerful transformer-based image encoder to

predict the vector flow fields characteristic of Cellpose. This combination aims to achieve superior generalization across diverse microscopy images by benefiting from SAM’s large-scale pre-training.

Stardist [12]: StarDist assumes that cell nuclei can be approximated by star-convex polygons. It employs a convolutional neural network to simultaneously predict the distances to the polygon boundary along several radial directions and the associated object probabilities, facilitating robust instance segmentation even in crowded environments.

MEDIAR [13]: MEDIAR is a transformer-based framework designed for cell instance segmentation in multi-modality microscopy images. It incorporates “human-in-the-loop” strategies, combining data-centric and model-centric approaches to predict cell masks and class probabilities, ensuring robustness across varying imaging conditions.

References

- [1] Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and *in situ* analysis. *Nature Communications*, 14(1):8353–8368, 2023.
- [2] Xiaohang Fu, Yingxin Lin, David M Lin, Daniel Mechtersheimer, Chuhan Wang, Farhan Ameen, Shila Ghazanfar, Ellis Patrick, Jinman Kim, and Jean YH Yang. Bidcell: Biologically-informed self-supervised learning for segmentation of subcellular spatial transcriptomics data. *Nature Communications*, 15(1):509–524, 2024.
- [3] Daniel C Jones, Anna E Elz, Azadeh Hadadianpour, Heeju Ryu, David R Glass, and Evan W Newell. Cell simulation as cell segmentation. *Nature Methods*, pages 1–12, 2025.
- [4] Hao Chen, Dongshunyi Li, and Ziv Bar-Joseph. SCS: cell segmentation for high-resolution spatial transcriptomics. *Nature Methods*, 20(8):1237–1243, 2023.
- [5] Yuheng Chen, Xin Xu, Xiaomeng Wan, Jiashun Xiao, and Can Yang. UCS: a unified approach to cell segmentation for subcellular spatial transcriptomics. *Small Methods*, 9(5):2400975, 2025.
- [6] Viktor Petukhov, Rosalind J Xu, Ruslan A Soldatov, Paolo Cadinu, Konstantin Khodosevich, Jeffrey R Moffitt, and Peter V Kharchenko. Cell segmentation in imaging-based spatial transcriptomics. *Nature Biotechnology*, 40(3):345–354, 2022.
- [7] Quentin Blampey, Kevin Mulder, Margaux Gardet, Stergios Christodoulidis, Charles-Antoine Dutertre, Fabrice André, Florent Ginhoux, and Paul-Henry Cournède. Sopa: a technology-invariant pipeline for analyses of image-based spatial omics. *Nature Communications*, 15(1):4981, 2024.
- [8] Dylan Bannon, Erick Moen, Morgan Schwartz, Enrico Borba, Takamasa Kudo, Noah Greenwald, Vibha Vijayakumar, Brian Chang, Edward Pao, Erik Osterman, et al. Deepcell kiosk: scaling deep learning-enabled cellular image analysis with kubernetes. *Nature Methods*, 18(1):43–45, 2021.
- [9] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
- [10] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021.
- [11] Marius Pachitariu, Michael Rariden, and Carsen Stringer. Cellpose-SAM: superhuman generalization for cellular segmentation. *bioRxiv*, pages 2025–04, 2025.

- [12] Martin Weigert and Uwe Schmidt. Nuclei instance segmentation and classification in histopathology images with stardist. In *The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, 2022.
- [13] Gihun Lee, SangMook Kim, Joonkee Kim, and Se-Young Yun. Mediar: Harmony of data-centric and model-centric for multi-modality microscopy. *arXiv preprint arXiv:2212.03465*, 2022.