

Supplementary file to Histology Images in Spatial Transcriptomics: A Critical Evaluation of Their Effectiveness in Spatial Domain Detection

Abdullah Al Noman^{1,2}, Asif Ajrof¹, Md. Abul Hassan Samee^{3,*}, and M Saifur Rahman^{1,*}

¹Bangladesh University of Engineering and Technology, Department of Computer Science and Engineering, Dhaka, 1000, Bangladesh

²Virginia Tech, Department of Computer Science, Blacksburg, Virginia, 24060, United States of America

³Baylor College of Medicine, Department of Integrative Physiology, Houston, Texas, 77030, United States of America

*Corresponding author(s): Md. Abul Hassan Samee (Md.AbulHassan.Samee@bcm.edu) and M Saifur Rahman (mrahman@cse.buet.ac.bd)

ABSTRACT

This supplementary document provides a comprehensive methodological extension and exhaustive performance benchmarking to support the evaluation of histology images in spatial transcriptomics (ST). While the main manuscript focuses on representative results from Sample 151507, this file expands the analysis to the full 15-sample cohort, including 12 human dorsolateral prefrontal cortex (DLPFC) samples and 3 distinct cancer datasets (HBC, Melanoma, and BCDC).

We first formalize a diagnostic clustering framework (Algorithm S1) to quantify the intrinsic structural signal in H&E images, demonstrating that even when high-resolution visual signal is present, current integration models like SpaGCN¹ and conST² often fail to achieve significant performance gains. To ensure data quality, we detail the image enhancement process using SwinIR³, validated by Peak Signal-to-Noise Ratio (PSNR) metrics across all samples. We further provide a massive robustness analysis over 20 random seeds, evaluating performance through Adjusted Rand Index (ARI), Silhouette scores, and spatial autocorrelation metrics (Moran's I and Geary's C). Finally, we introduce an extended cross-method concordance framework. To evaluate pairwise Adjusted Mutual Information (AMI), and Boundary IoU, we reveal a consistent dichotomy: while computational methods successfully converge on biological cores, they exhibit high entropy and disagreement at transition zones. These findings offer a rigorous characterization of the current limitations in multi-modal data fusion for spatial transcriptomics.

Seed generation

To evaluate the stochastic robustness of the spatial domain detection methods, we initialized our experimental runs using 20 randomly generated seeds. These seeds were automatically generated prior to the experiments using the NumPy library, specifically via the function `np.random.randint(1000, size=20)`.

Median ARI for run with 20 random seeds

In addition to the 12 human dorsolateral prefrontal cortex (DLPFC) samples, we included three cancer datasets—human breast cancer, melanoma, and breast cancer ductal carcinoma to evaluate method robustness across distinct tissue types. Each method was executed using the 20 different random seeds, and performance was summarized using the median Adjusted Rand Index (ARI). Table S1 reports the median ARI values for configurations with and without histology. Values shown in **bold** indicate cases where incorporating histology information resulted in an equal or higher median ARI compared to using gene expression data alone.

Significance of using H&E image along with Gene expression in spatial domain detection in various methods vs ARI found by clustering the H&E image

The first four columns of Table-S2 display the statistical significance of spatial domain detection performance improvement, if any, with histology image (see Algorithm-1 in the manuscript). Additionally, we have clustered the high-resolution histology images. To achieve this, we extracted the pixels under the ST spots and applied k-means clustering to obtain labels for each

Table S1. Median ARIs for human DLPFC samples, human breast cancer, melanoma cancer, and breast cancer ductal carcinoma sample for different methods, for 20 random seeds.

Sample	SpaGCN median ARI without histology	SpaGCN median ARI with histology	conST me- dian ARI without histology	conST median ARI with histology	ScribbleDom median ARI without histology	ScribbleDom median ARI with histol- ogy
151507	0.45	0.43	0.37	0.33	0.43	0.45
151508	0.43	0.38	0.28	0.25	0.44	0.42
151509	0.43	0.41	0.33	0.27	0.44	0.57
151510	0.44	0.44	0.30	0.29	0.51	0.53
151669	0.22	0.25	0.37	0.33	0.33	0.46
151670	0.23	0.22	0.33	0.30	0.29	0.68
151671	0.54	0.52	0.51	0.46	0.61	0.55
151672	0.57	0.55	0.48	0.45	0.62	0.65
151673	0.51	0.43	0.47	0.44	0.53	0.50
151674	0.41	0.36	0.46	0.50	0.25	0.52
151675	0.43	0.39	0.40	0.49	0.38	0.48
151676	0.39	0.35	0.45	0.42	0.36	0.37
hbc	0.17	0.18	0.05	0.16	0.14	0.37
Melanoma	0.31	0.30	0.27	0.19	0.48	0.48
bcde	0.43	0.42	0.54	0.35	0.83	0.81

pixel. Subsequently, for each spot, we determined the mode of the labels obtained from k-means clustering on the pixels. The ARIs obtained for the samples are shown in the fifth column of the table. This indicates that, although some samples (e.g., 151673, 151674, 151675, 151676 of the human DLPFC dataset along with human breast cancer and melanoma cancer samples) exhibit higher ARI from clustering on the histology image, they do not show sufficient improvement in ARI with the usage of histology images for conST and SpaGCN.

To calculate the ARI from the raw histology image, we use Algorithm -S1. This algorithm first identifies the spot coordinates of the tissue sample and extracts the pixels of interest under these spots (Lines 1-5), considering the radius of the spots (Line 3). It uses the ‘meshgrid’ and ‘range’ functions from the Python library ‘numpy’ to extract the pixel positions under each spot. Then, the ‘filter’ function (Line 6) extracts the pixel intensities of the histology images for the pixels of interest, and k-means⁴ clustering is performed on these pixel intensities (Line 7). The algorithm then extracts the cluster labels for each spot by finding the mode of the labels obtained for each pixel under that spot. Finally, the algorithm returns the ARI from this cluster labeling by comparing it with the manual annotation of the ST data.

Algorithm S1 Clustering histology image, and finding ARI

Input: Histology_image, spot_radius, spot_coordinate, num_clusters, Manual_annotation

Output: ARI

```

1: pixels_of_interest  $\leftarrow \emptyset$ 
2: for each x_coordinate, y_coordinate  $\in$  spot_coordinate do
3:   spots_in_this_pixel  $\leftarrow$  meshgrid(range((x - radius), (x + radius) + 1), range((y - radius), (y + radius) + 1))
4:   pixels_of_interest  $\leftarrow$  pixels_of_interest.append(spots_in_this_pixel)
5: end for
6: histology_intensities_of_interest  $\leftarrow$  filter(Histology_image, pixels_of_interest)
7: Cluster_labels_per_pixel  $\leftarrow$  k_means(histology_intensities_of_interest, num_clusters)
8: Cluster_labels_per_spot  $\leftarrow$  Mode(Cluster_labels_per_pixel, spot_coordinate)
9: return ARI(Cluster_labels_per_spot, Manual_annotation)

```

PSNR value between raw and preprocessed image with swinIR

Table S3 shows peak signal to noise ratio improvement for each of the 15 samples under study after using swinIR³.

Table S2. Table showing the significance of using histology image while using different methods and ARI obtained from clustering on histology image.

	SpaGCN	conST	ScribbleDom	ARI found by clustering H&E image
151507	Not Significant	Not Significant	Significant	0.09
151508	Not Significant	Not Significant	Not Significant	0.15
151509	Not Significant	Not Significant	Significant	0.12
151510	Not Significant	Not Significant	Not Significant	0.08
151669	Not Significant	Not Significant	Significant	-0.07
151670	Not Significant	Not Significant	Significant	0.07
151671	Not Significant	Not Significant	Not Significant	0.01
151672	Not Significant	Not Significant	Not Significant	-0.02
151673	Not Significant	Not Significant	Not Significant	0.18
151674	Not Significant	Not Significant	Significant	0.18
151675	Not Significant	Not Significant	Significant	0.21
151676	Not Significant	Not Significant	Significant	0.18
HBC	Not Significant	Significant	Significant	0.14
Melanoma	Not Significant	Not Significant	Not Significant	0.22
BCDC	Not Significant	Not Significant	Not Significant	0.02

Table S3. PSNR Values After SwinIR Enhancement

Sample	151507	151508	151509	151510	151669	151670	151671	151672	151673	151674	151675	151676	HBC	BCDC	Melanoma
PSNR	27.13	27.54	25.57	25.88	32.68	31.10	29.52	29.62	30.90	32.29	31.93	32.63	31.32	26.34	31.69

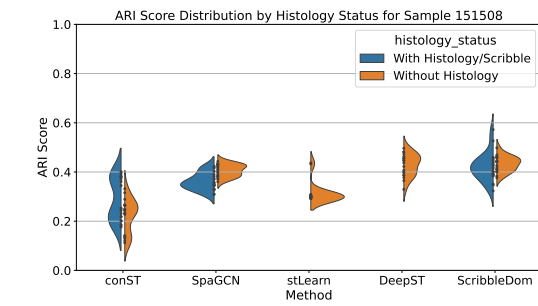
Extended performance measure across tissue samples

In the main manuscript, we presented a detailed performance evaluation using a representative sample (Sample 151507) from the human dorsolateral prefrontal cortex (DLPFC) dataset. To demonstrate the robustness and consistency of our findings across varying tissue architectures, this supplementary section provides the complete benchmark results for the remaining 11 samples of the human DLPFC dataset (Figures S1–S11).

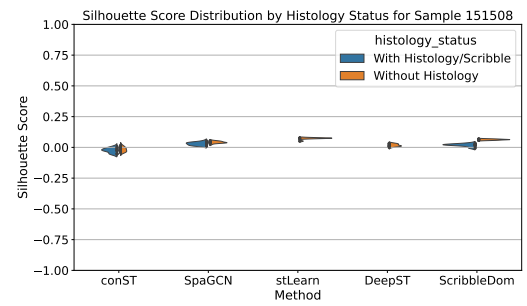
Furthermore, we extend our evaluation to include diverse pathological contexts by providing performance metrics for three additional cancer datasets: human breast cancer, melanoma cancer, and human breast cancer ductal carcinoma (Figures S12–S13). For each sample, we report the distribution of the Adjusted Rand Index (ARI), Silhouette score, Moran’s I, and Geary’s C across 20 random seeds. These results collectively reinforce our observation that the integration of histology images does not consistently improve spatial domain detection, and in some instances, may introduce noise that leads to a decrease in clustering accuracy compared to gene-expression-only models.

Extended analysis of cross method concordance and boundary precision

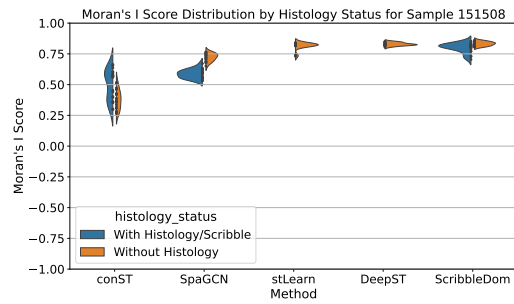
While the main manuscript detailed the cross-method concordance and boundary stability for Sample 151507, here we provide the full comparative analysis for the remaining 11 samples (Figures S15–S25). of the human DLPFC dataset and 3 cancer samples (Figures S26–S28). This comprehensive evaluation allows for a granular look at how different algorithmic architectures achieve consensus or diverge at tissue boundaries.



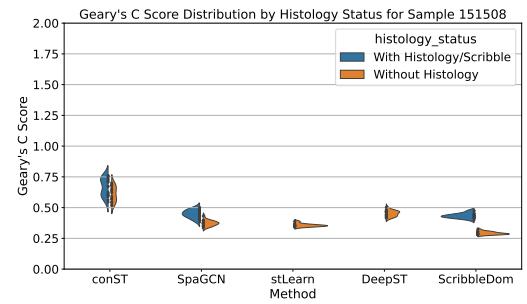
a



b

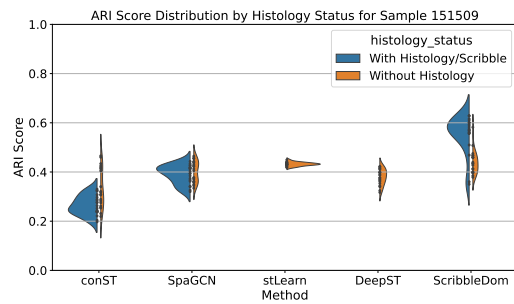


c

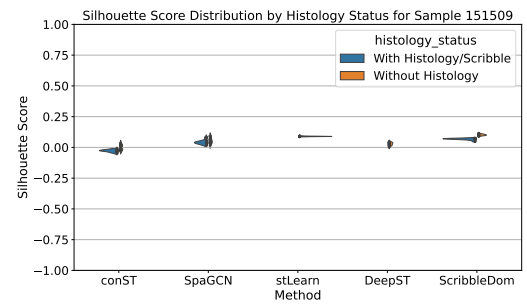


d

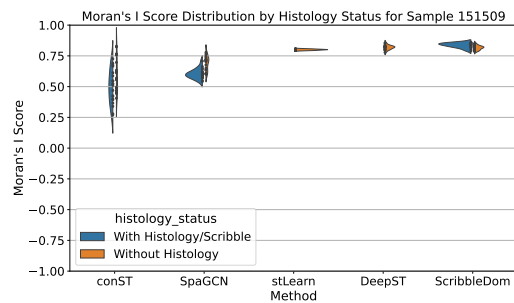
Figure S1. Metrics for sample 151508. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



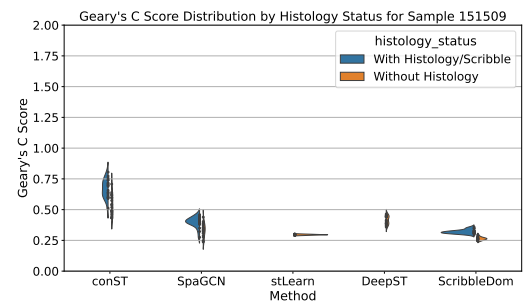
a



b

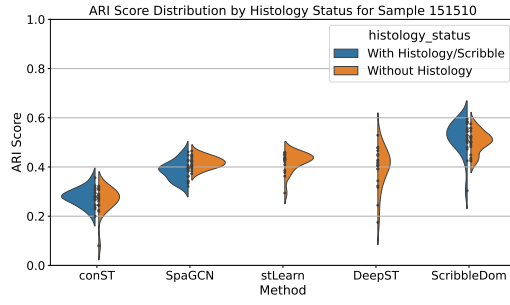


c

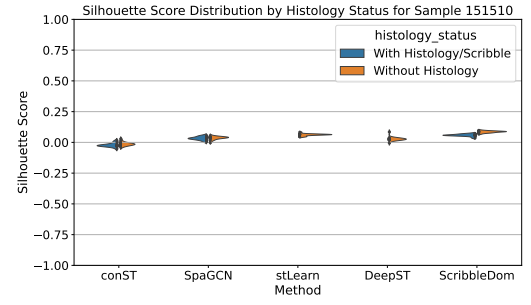


d

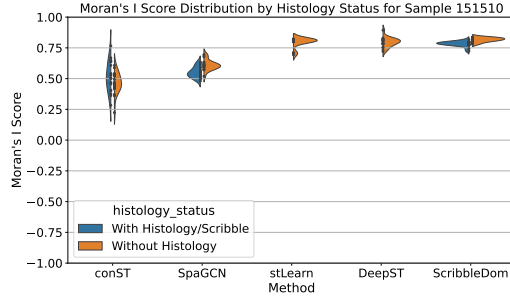
Figure S2. Metrics for sample 151509. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



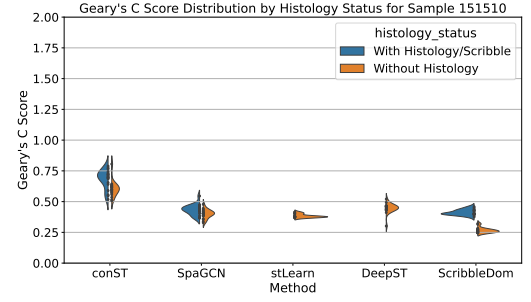
a



b

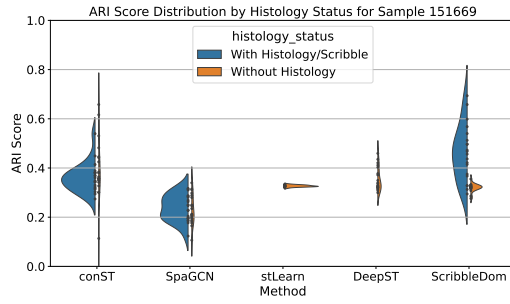


c

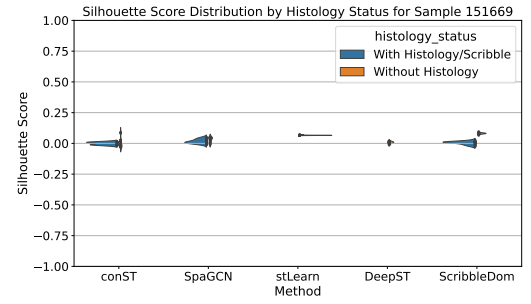


d

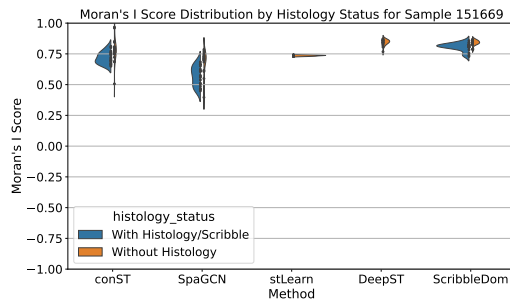
Figure S3. Metrics for sample 151510. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



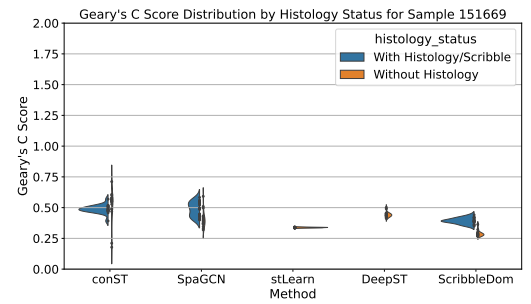
a



b

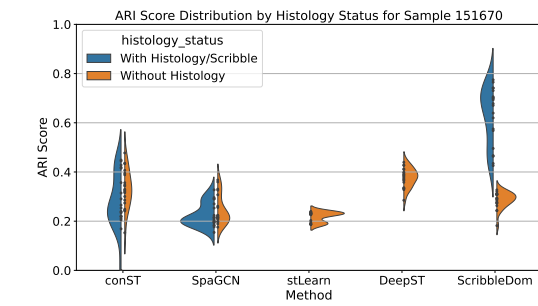


c

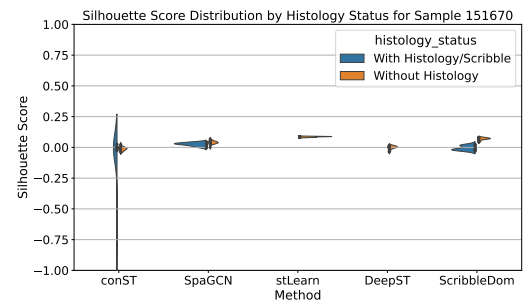


d

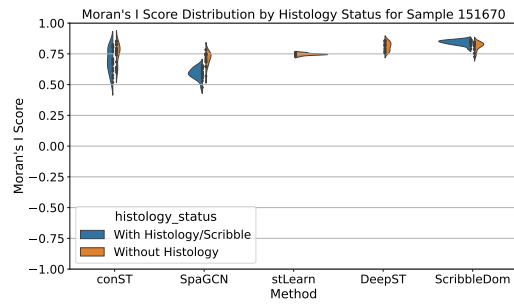
Figure S4. Metrics for sample 151669. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



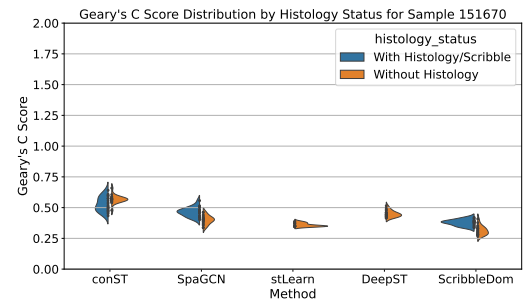
a



b

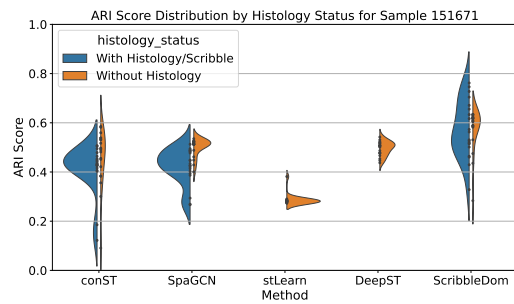


c

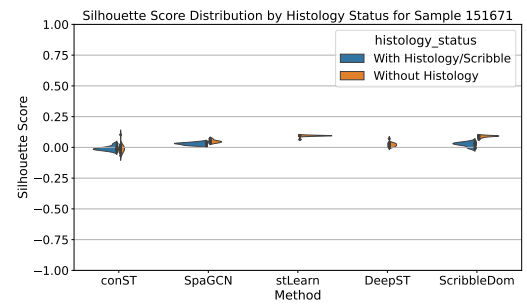


d

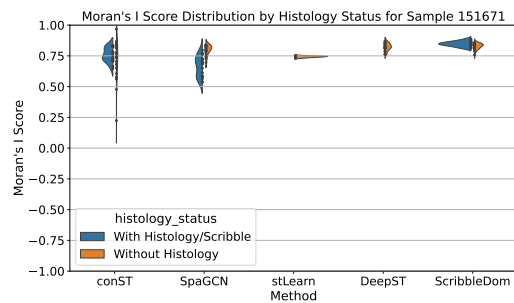
Figure S5. Metrics for sample 151670. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



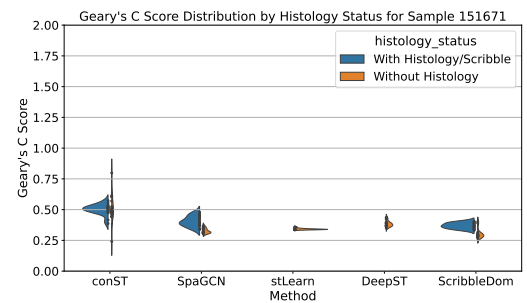
a



b

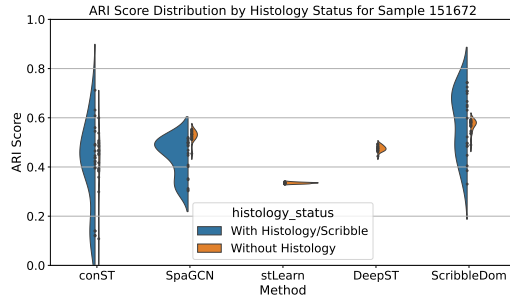


c

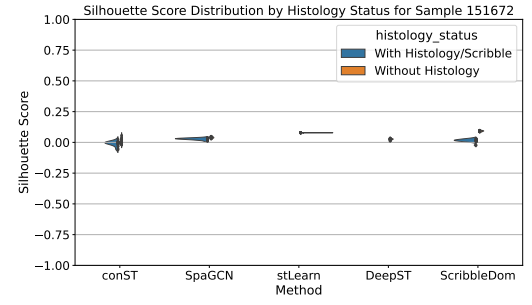


d

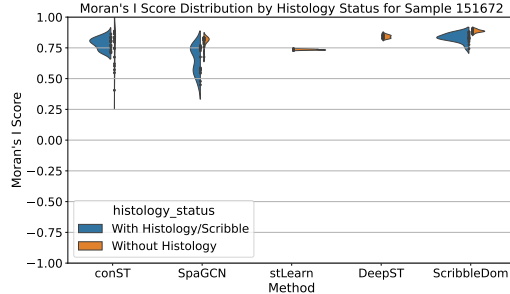
Figure S6. Metrics for sample 151671. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



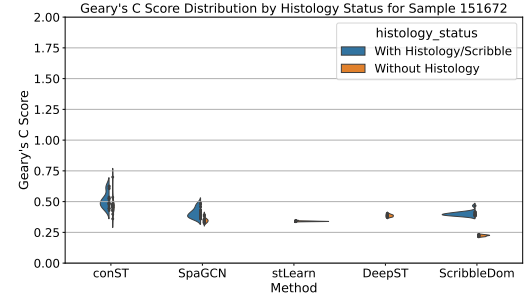
a



b

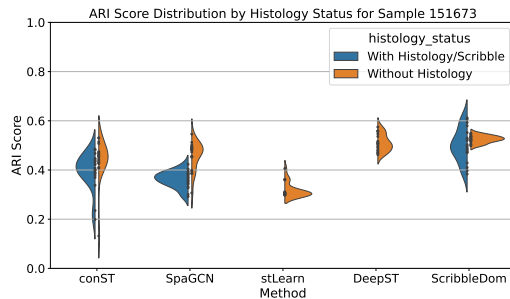


c

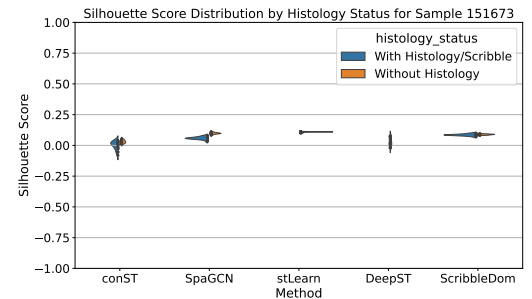


d

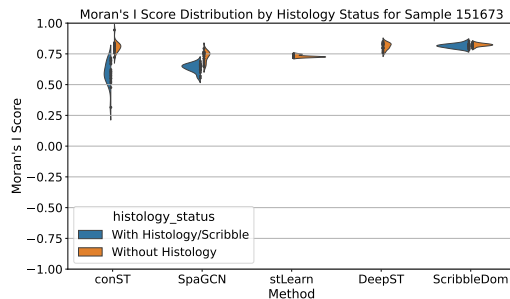
Figure S7. Metrics for sample 151672. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



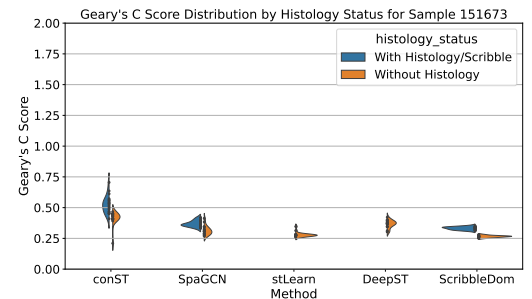
a



b

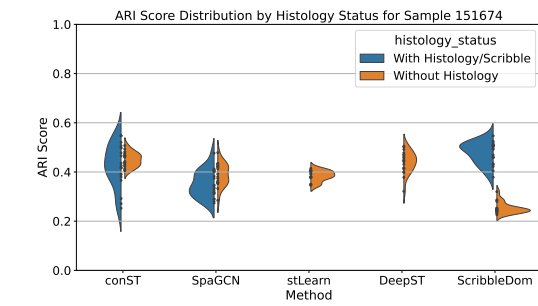


c

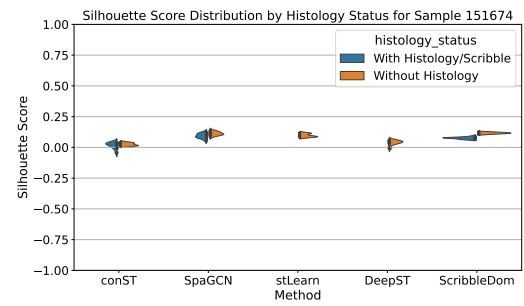


d

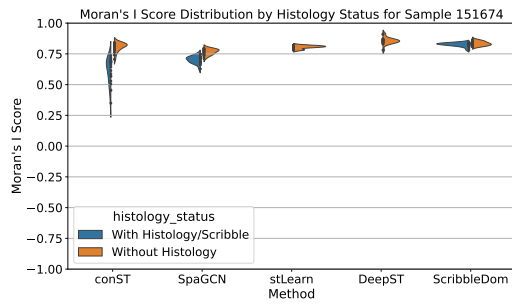
Figure S8. Metrics for sample 151673. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



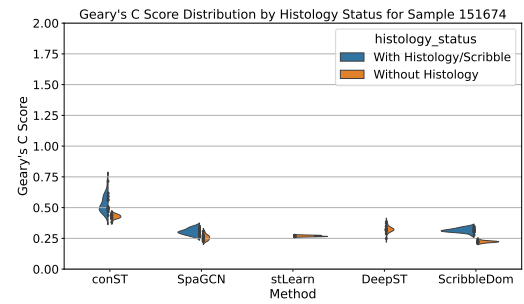
a



b

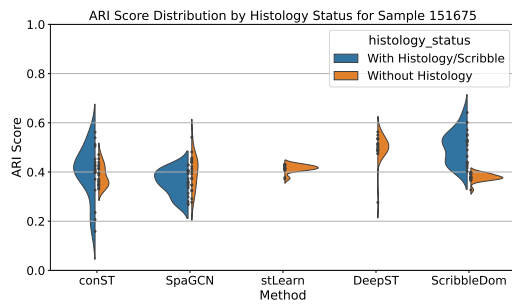


c

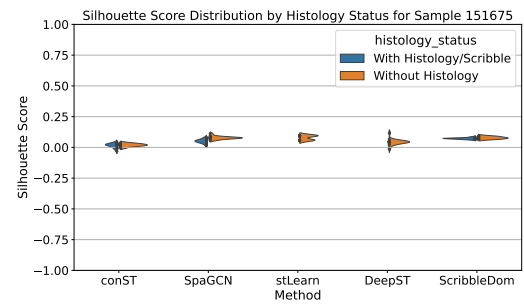


d

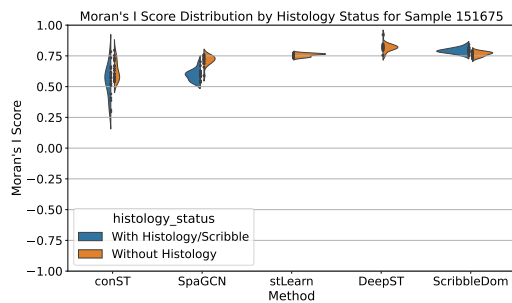
Figure S9. Metrics for sample 151674. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



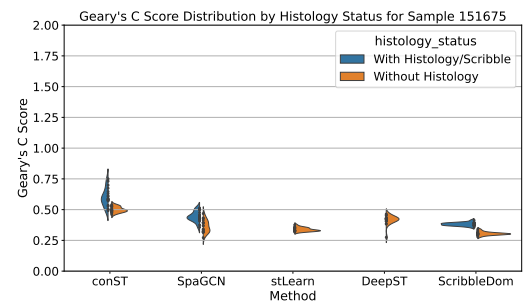
a



b

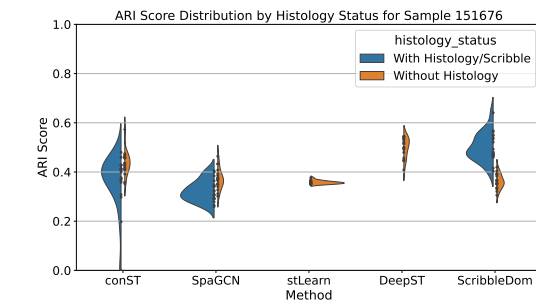


c

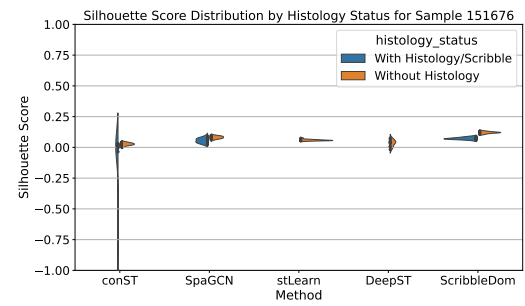


d

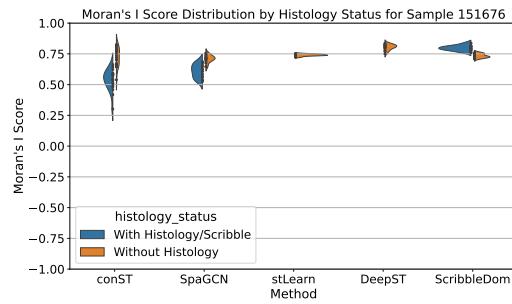
Figure S10. Metrics for sample 151675. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



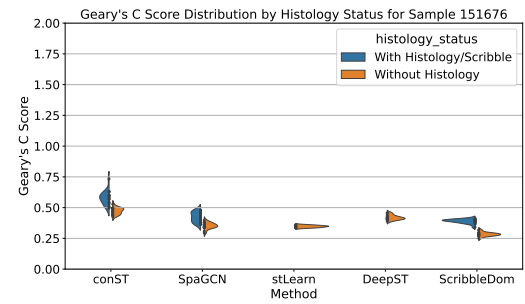
a



b

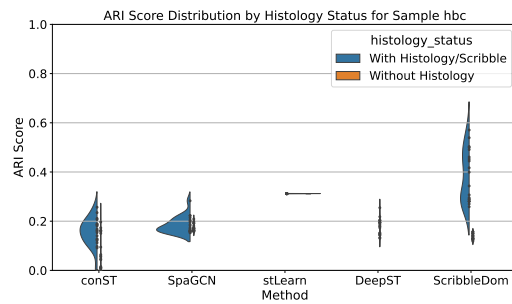


c

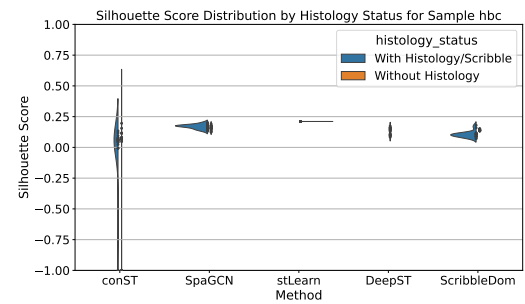


d

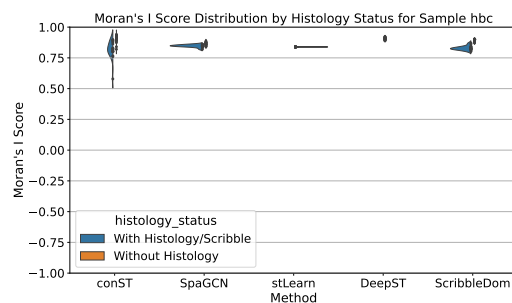
Figure S11. Metrics for sample 151676. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



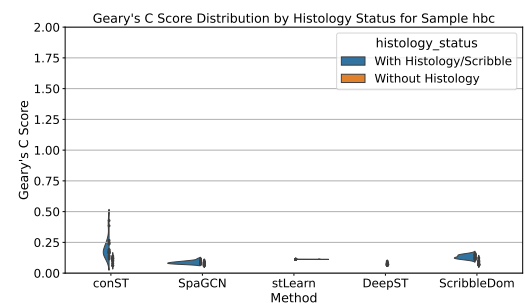
a



b

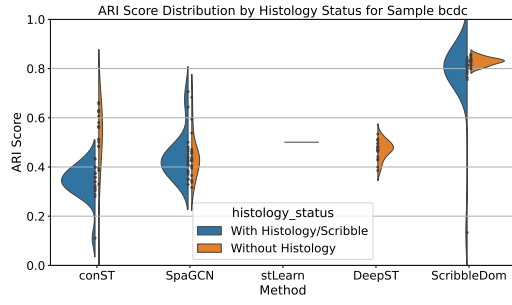


c

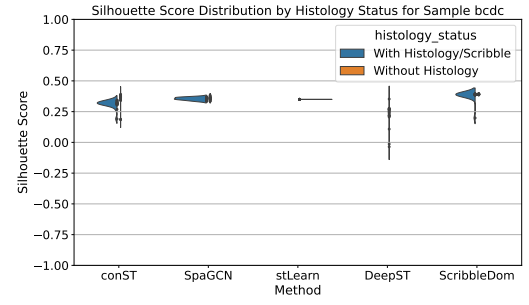


d

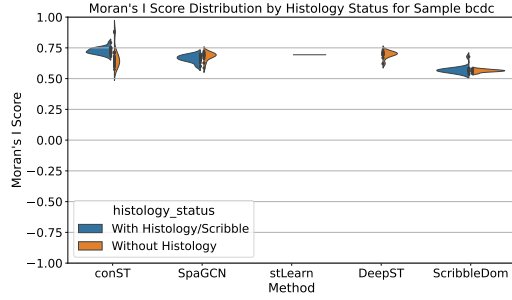
Figure S12. Metrics for sample hbc. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



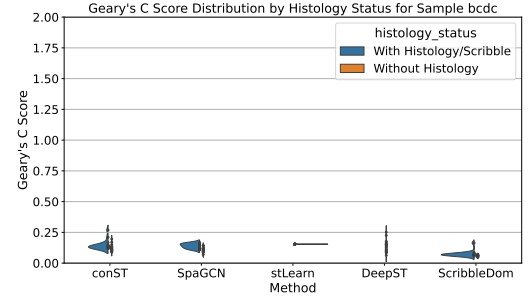
a



b

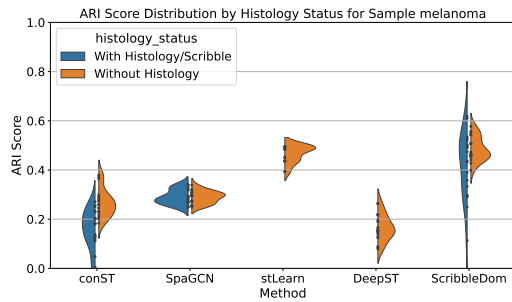


c

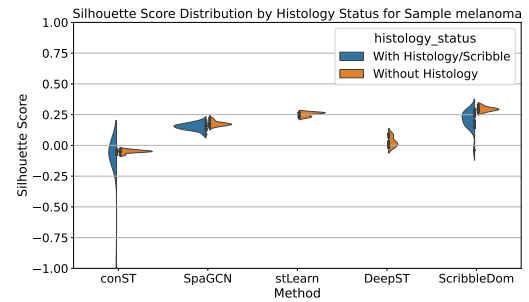


d

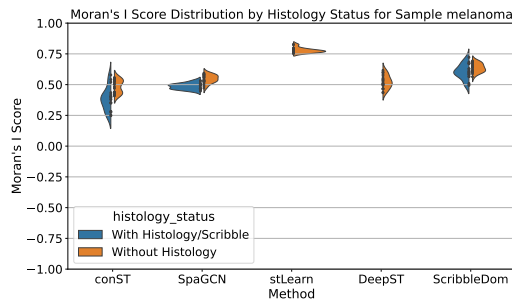
Figure S13. Metrics for sample bcdc. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



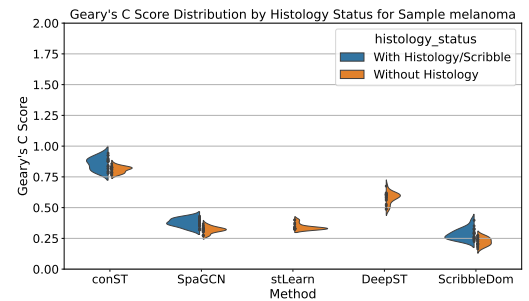
a



b

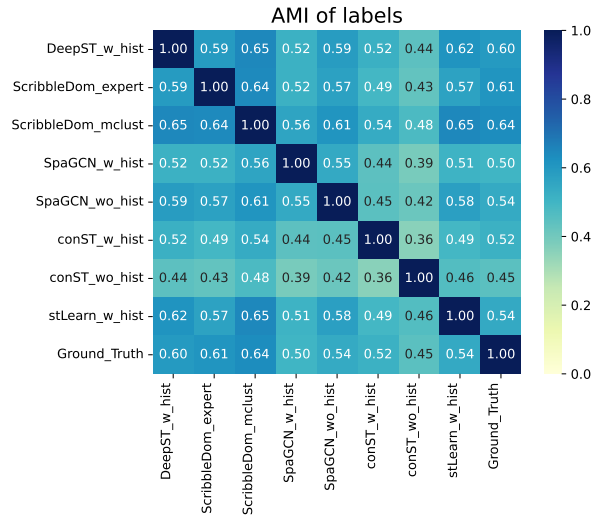


c

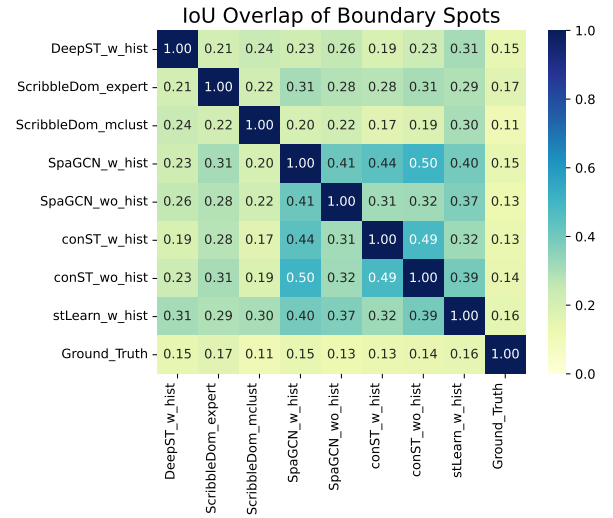


d

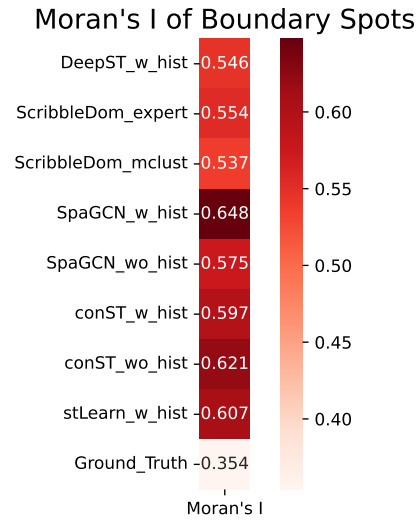
Figure S14. Metrics for sample melanoma. (A) Violin plot of ARI values across 20 random seeds. (B) Silhouette scores representing gene expression cluster density. (C) Moran's I scores indicating spatial autocorrelation. (D) Geary's C scores reflecting local spatial dissimilarity.



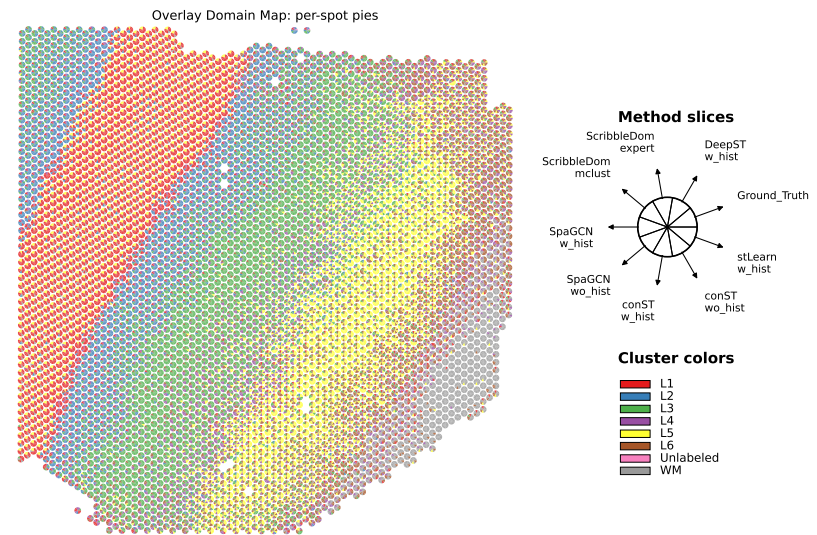
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

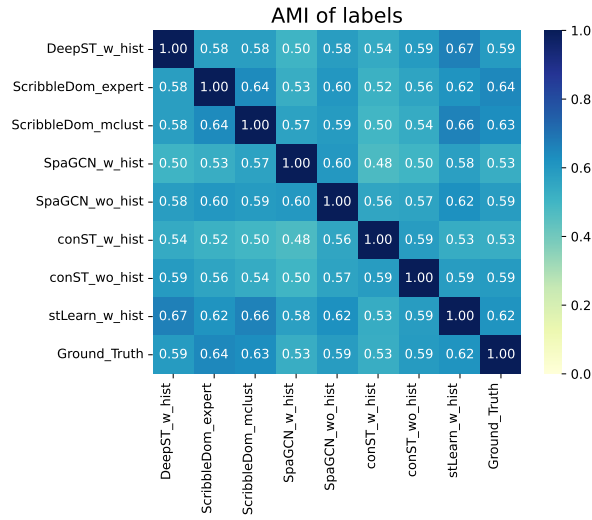


c Spatial Autocorrelation of Boundaries (Moran's I).

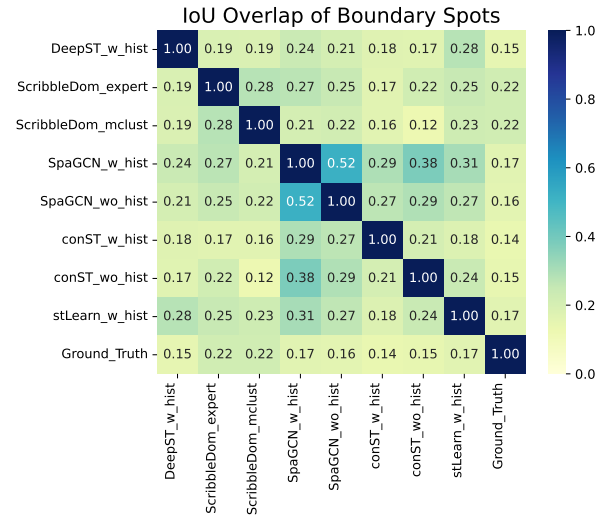


d Spatially Resolved Consensus Visualization.

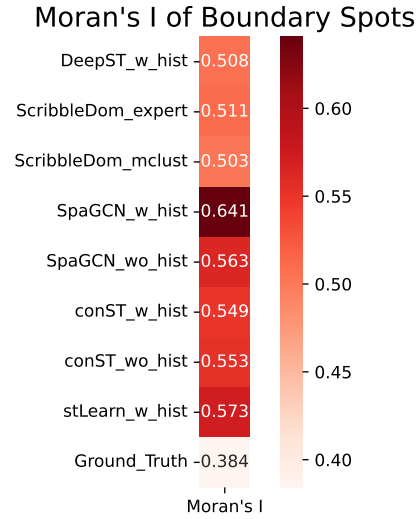
Figure S15. Multi-metric assessment of domain consistency and boundary definition. Using sample 151508, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



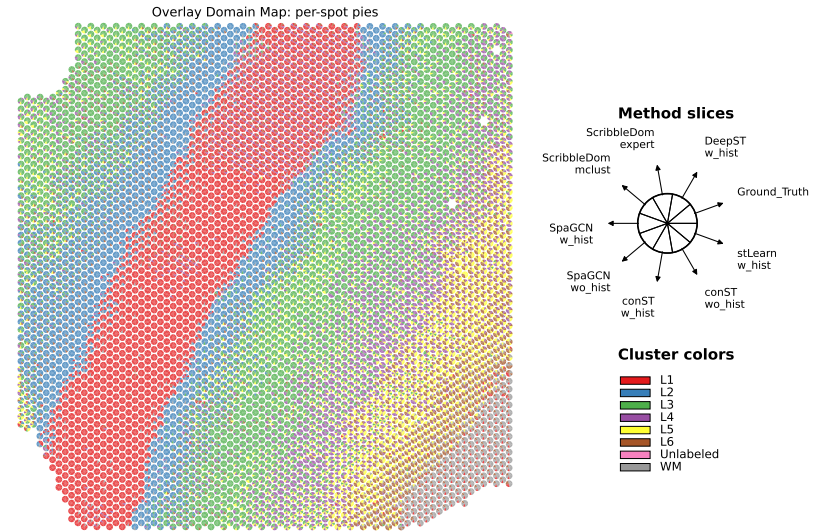
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

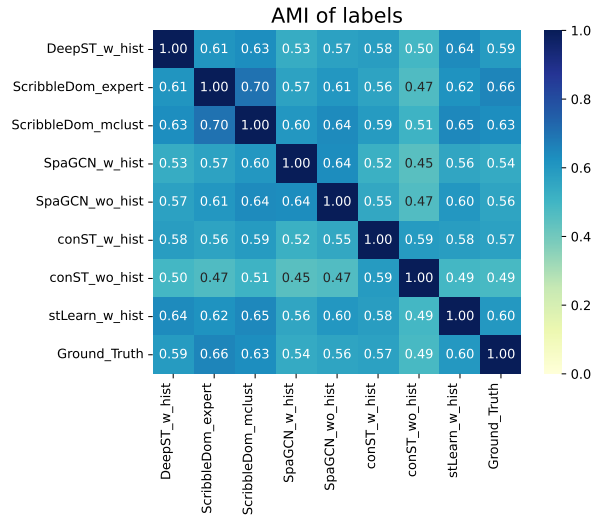


c Spatial Autocorrelation of Boundaries (Moran's I).

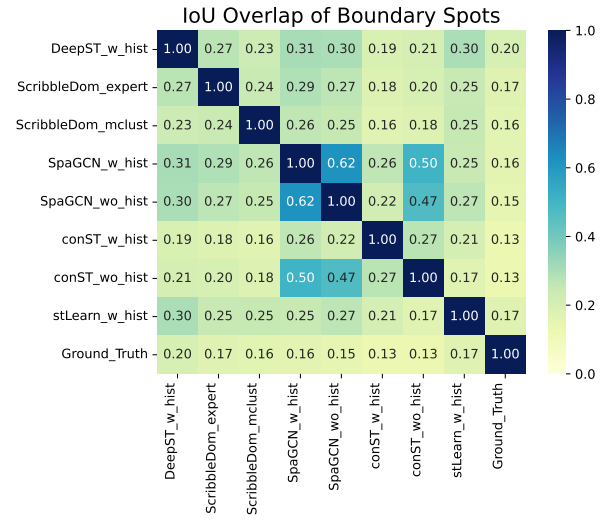


d Spatially Resolved Consensus Visualization.

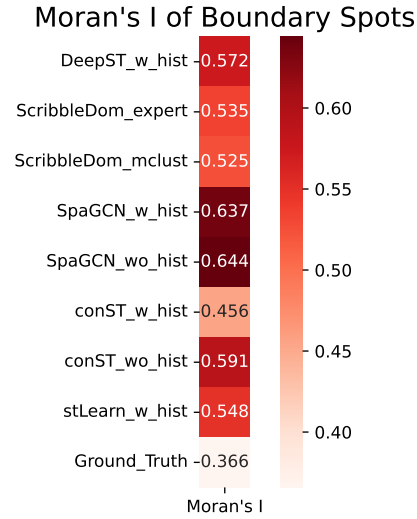
Figure S16. Multi-metric assessment of domain consistency and boundary definition. Using sample 151509, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



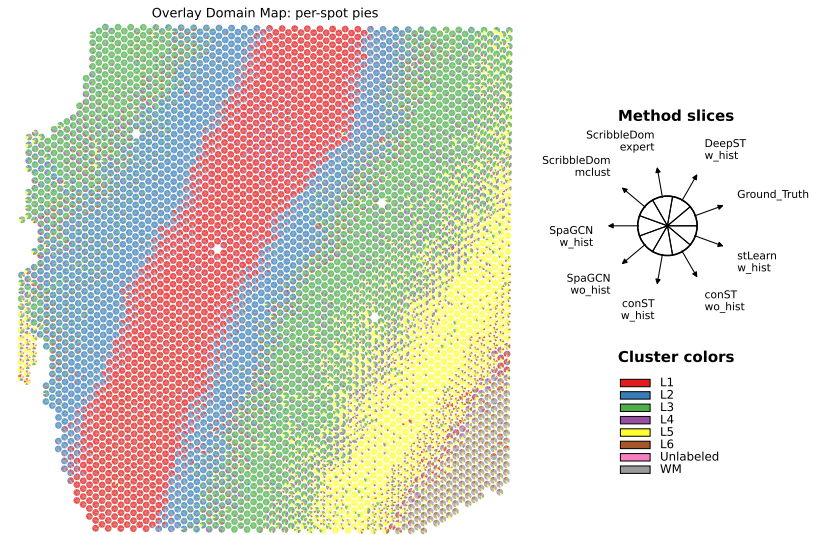
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

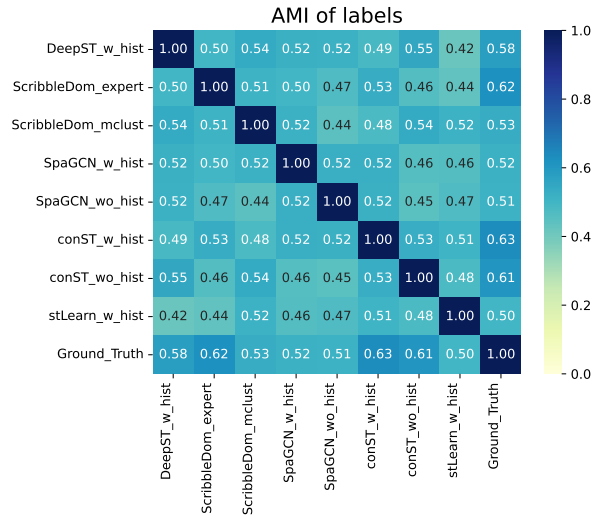


c Spatial Autocorrelation of Boundaries (Moran's I).

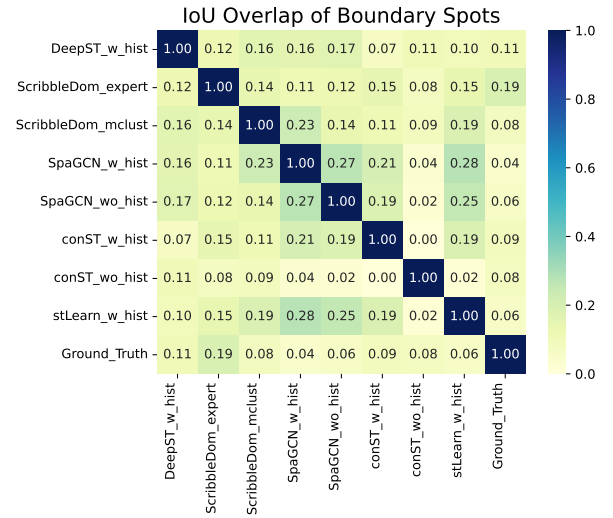


d Spatially Resolved Consensus Visualization.

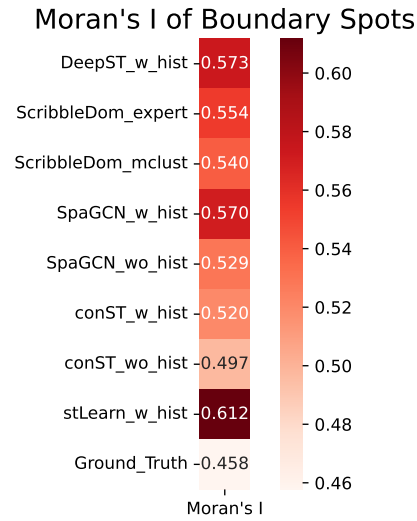
Figure S17. Multi-metric assessment of domain consistency and boundary definition. Using sample 151510, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



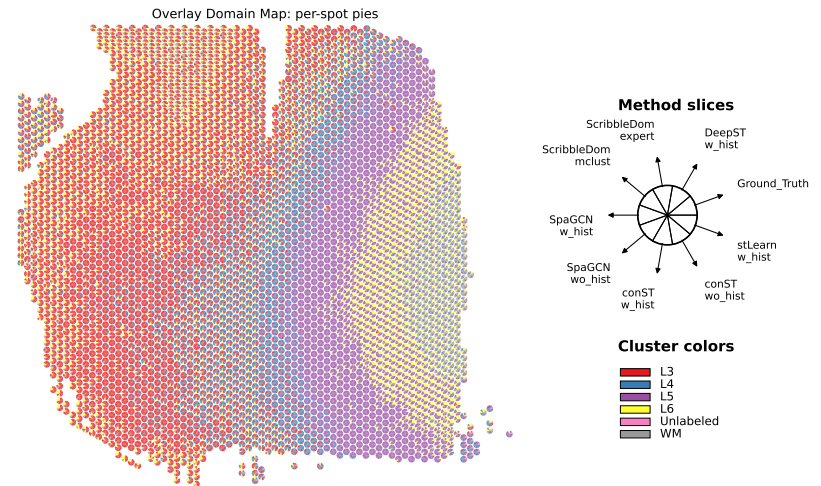
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

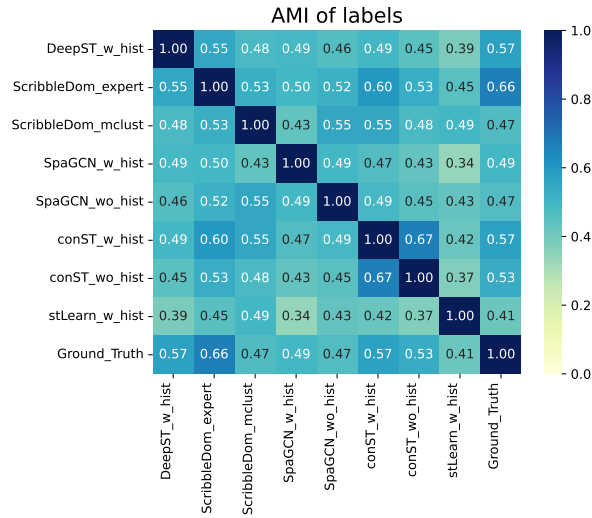


c Spatial Autocorrelation of Boundaries (Moran's I).

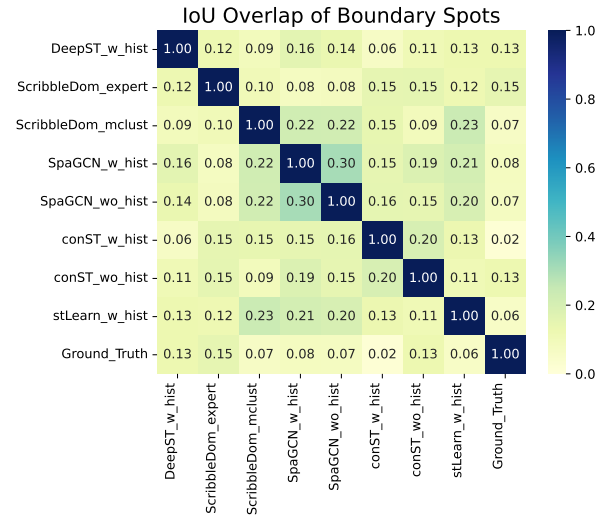


d Spatially Resolved Consensus Visualization.

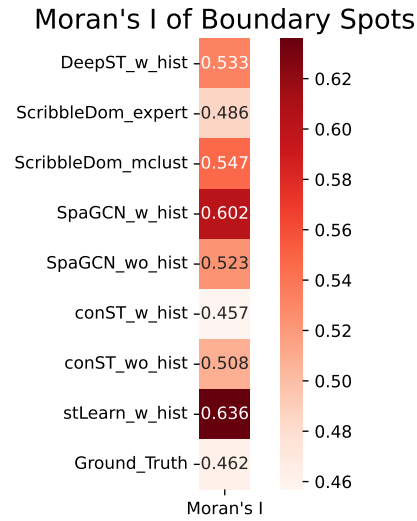
Figure S18. Multi-metric assessment of domain consistency and boundary definition. Using sample 151669, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



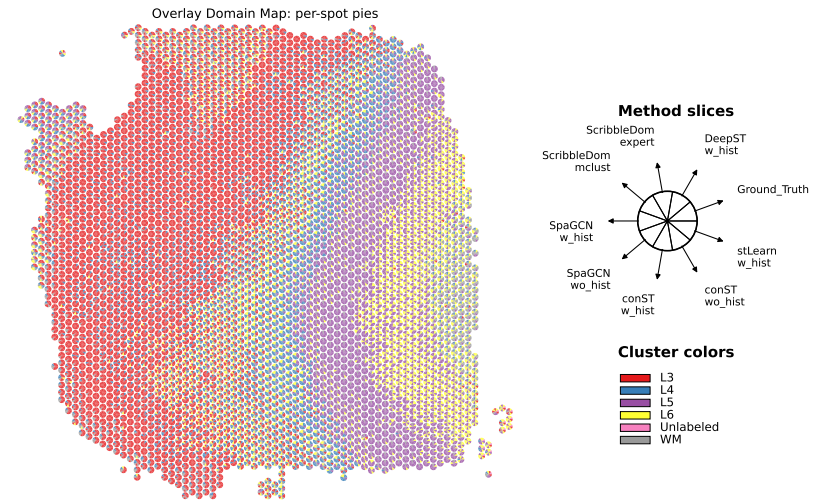
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

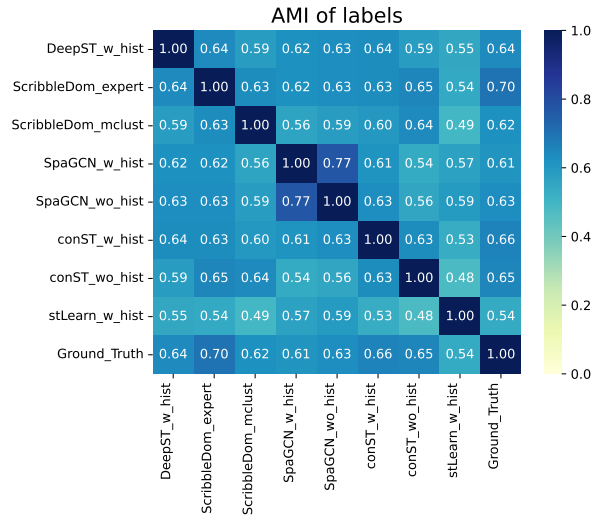


c Spatial Autocorrelation of Boundaries (Moran's I).

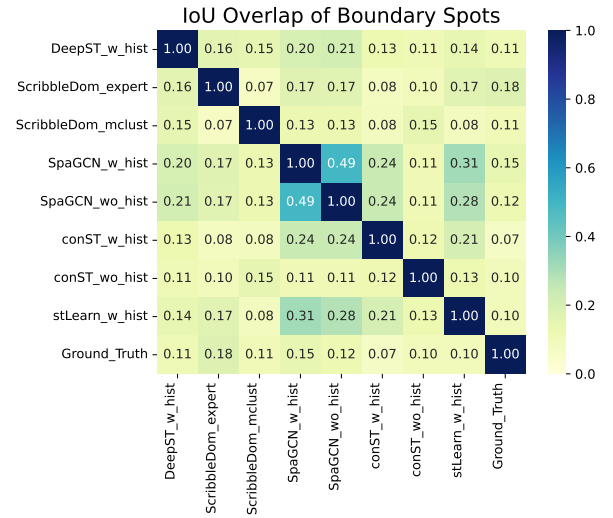


d Spatially Resolved Consensus Visualization.

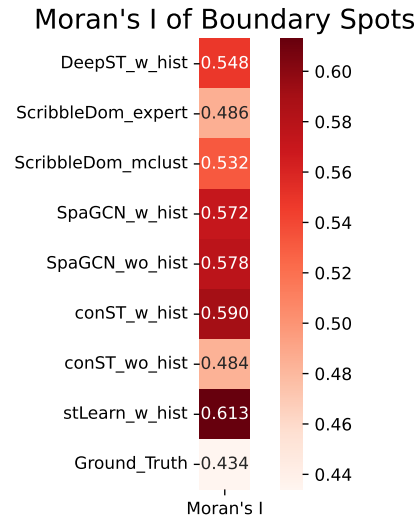
Figure S19. Multi-metric assessment of domain consistency and boundary definition. Using sample 151670, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



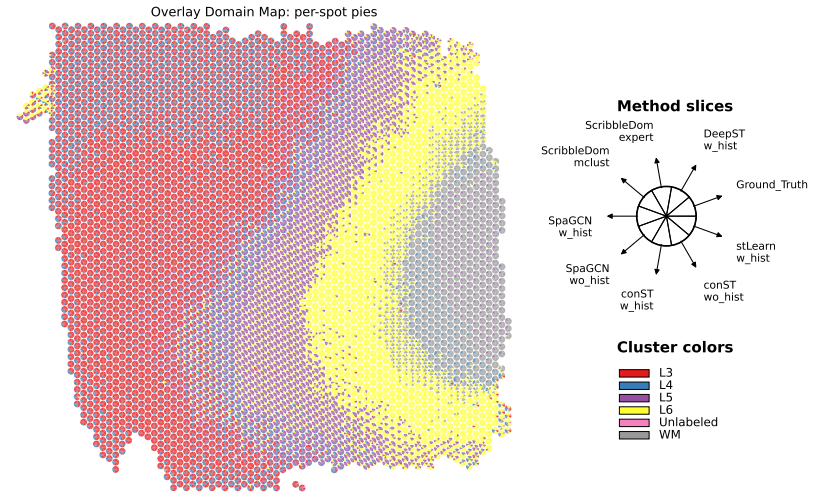
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

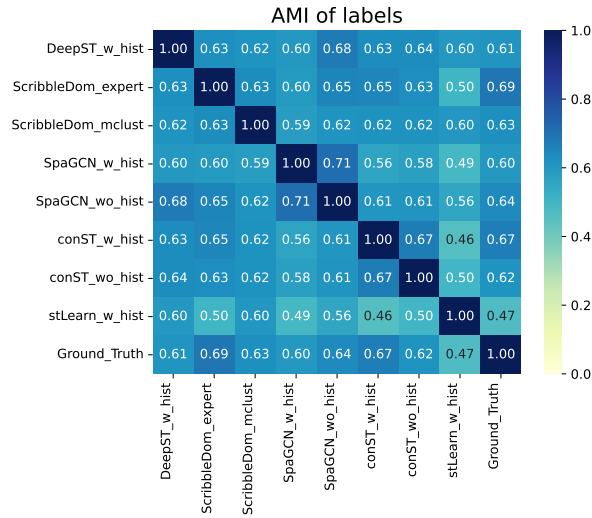


c Spatial Autocorrelation of Boundaries (Moran's I).

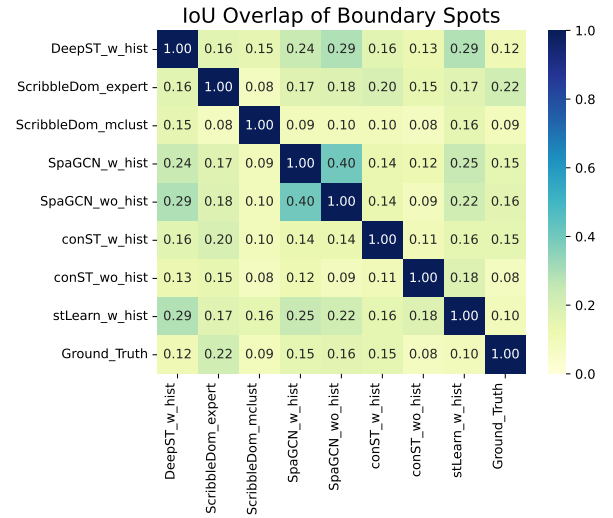


d Spatially Resolved Consensus Visualization.

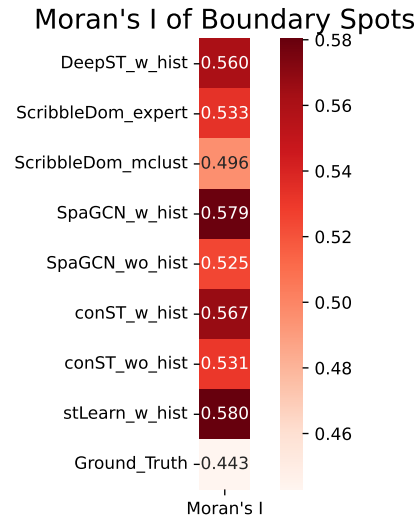
Figure S20. Multi-metric assessment of domain consistency and boundary definition. Using sample 151671, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



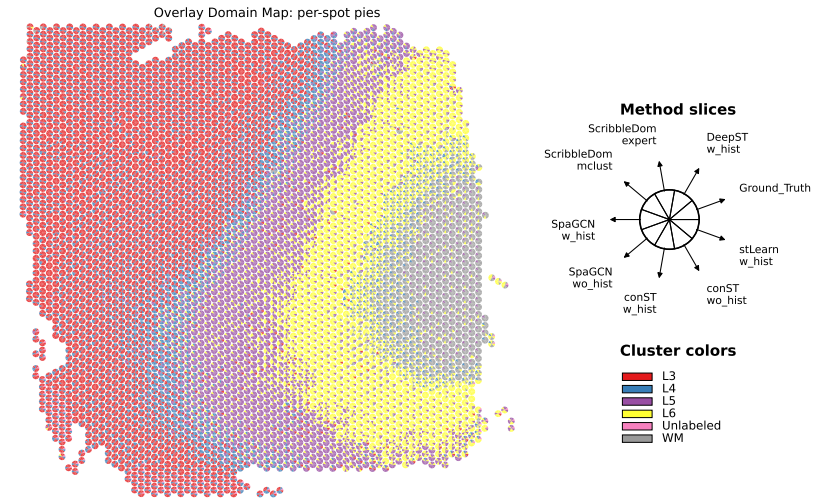
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

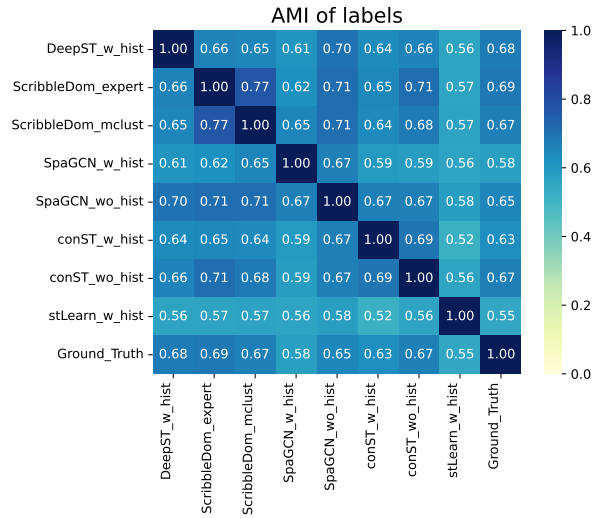


c Spatial Autocorrelation of Boundaries (Moran's I).

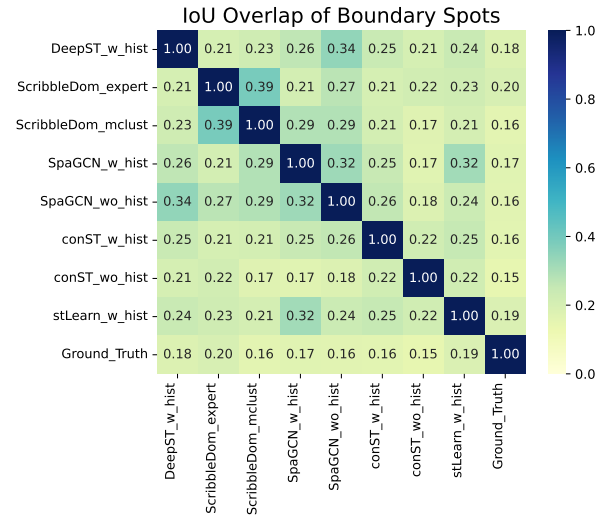


d Spatially Resolved Consensus Visualization.

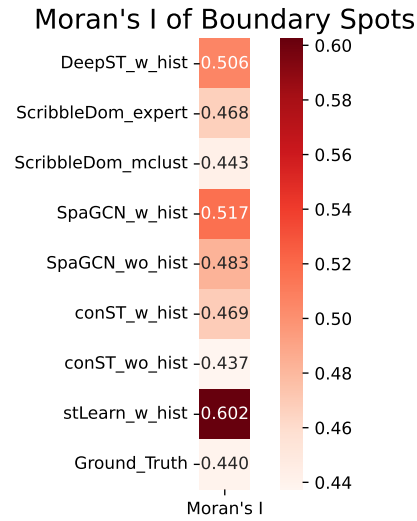
Figure S21. Multi-metric assessment of domain consistency and boundary definition. Using sample 151672, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



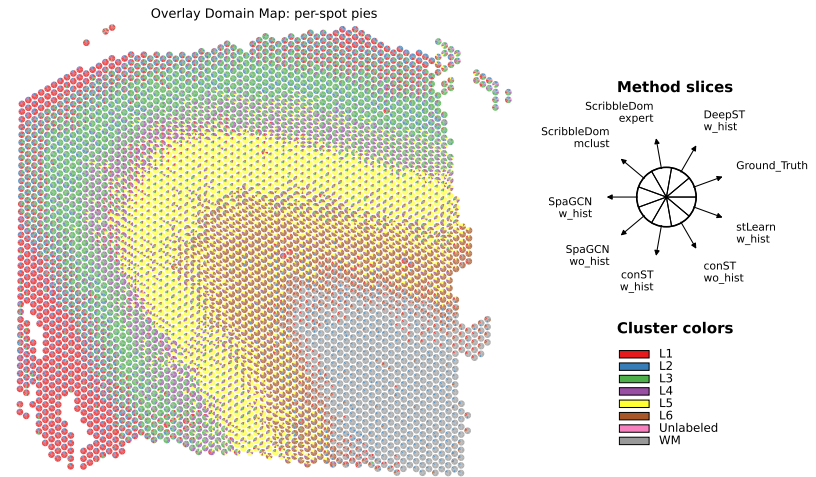
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

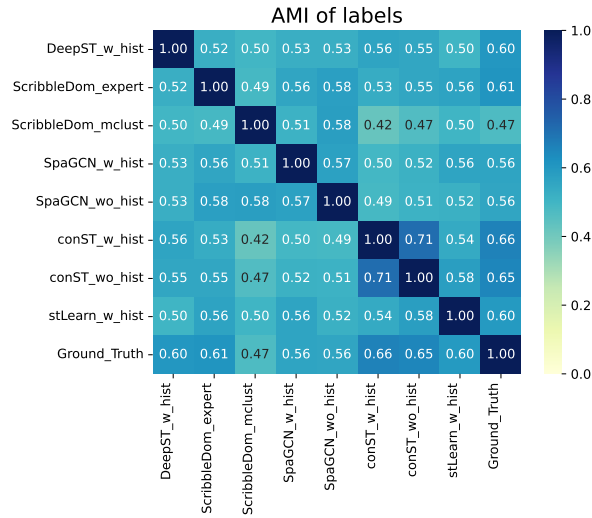


c Spatial Autocorrelation of Boundaries (Moran's I).

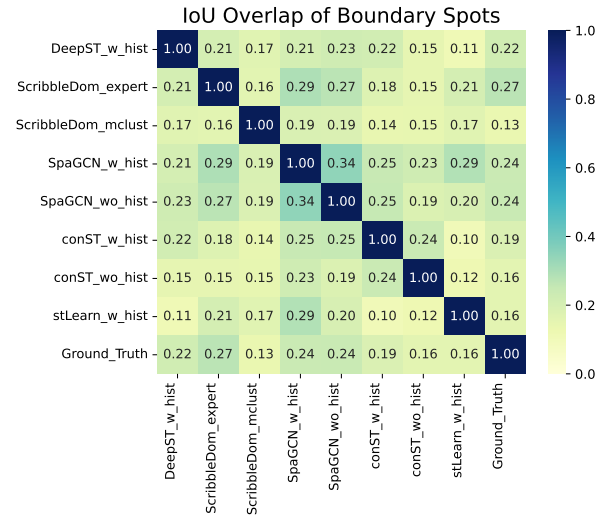


d Spatially Resolved Consensus Visualization.

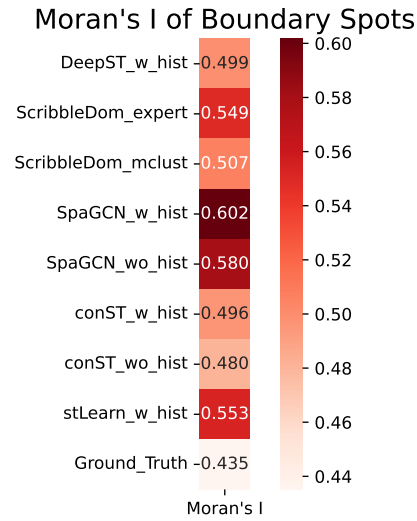
Figure S22. Multi-metric assessment of domain consistency and boundary definition. Using sample 151673, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



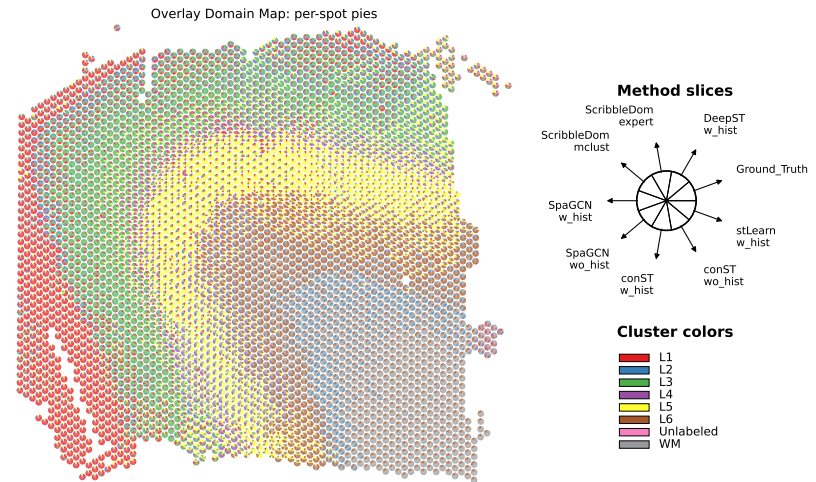
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

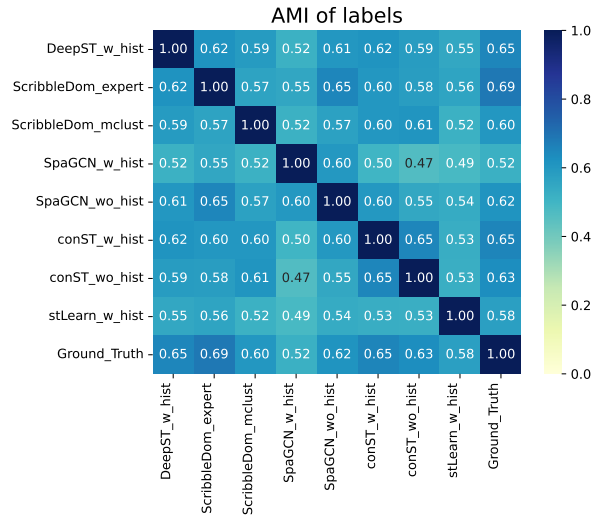


c Spatial Autocorrelation of Boundaries (Moran's I).

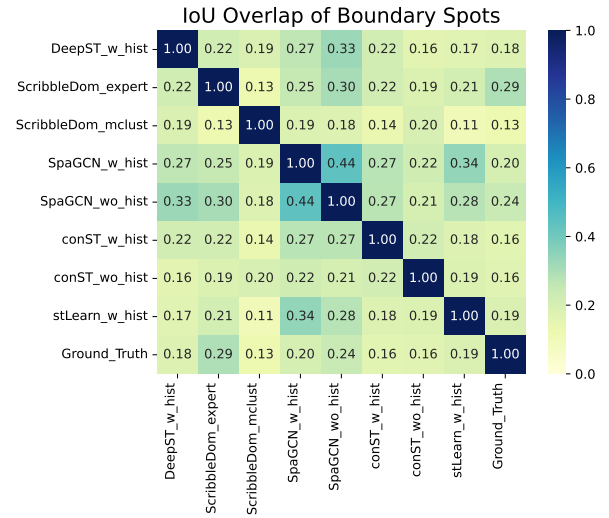


d Spatially Resolved Consensus Visualization.

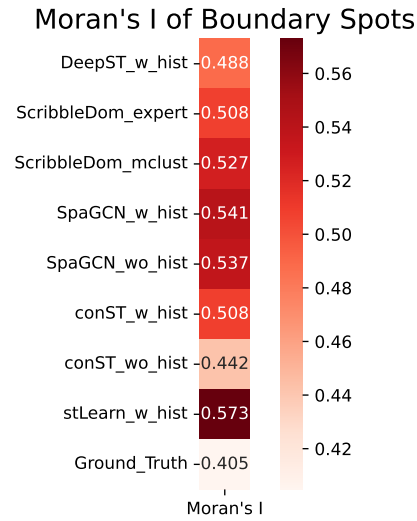
Figure S23. Multi-metric assessment of domain consistency and boundary definition. Using sample 151674, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



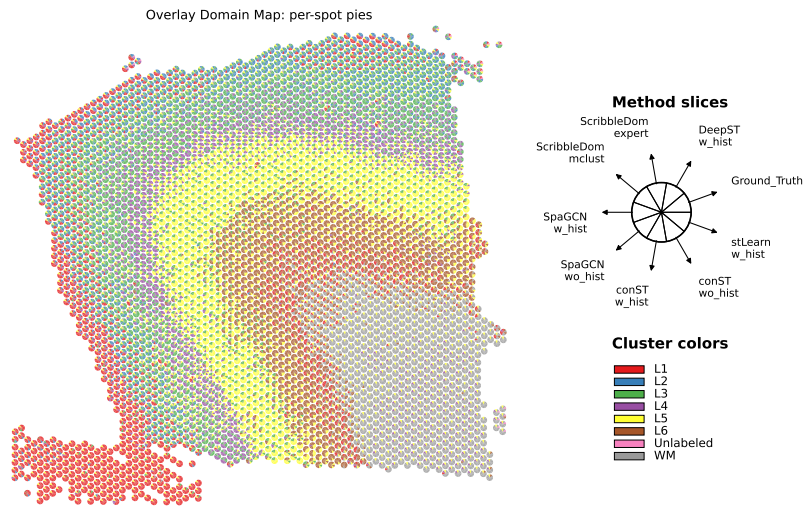
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

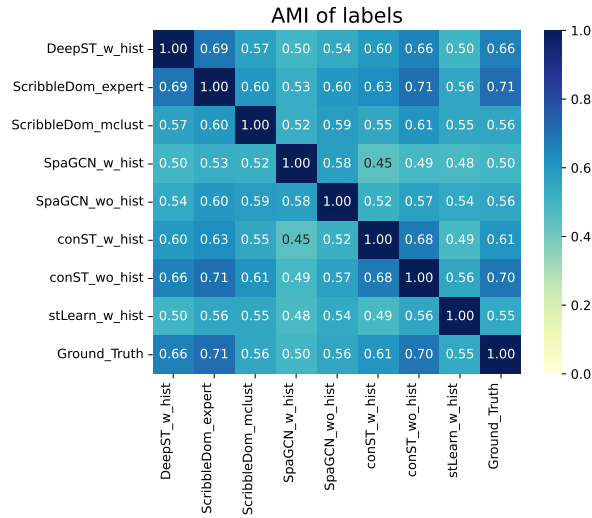


c Spatial Autocorrelation of Boundaries (Moran's I).

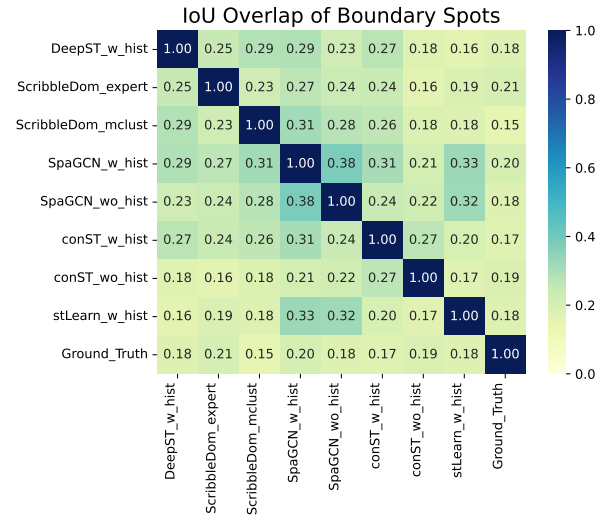


d Spatially Resolved Consensus Visualization.

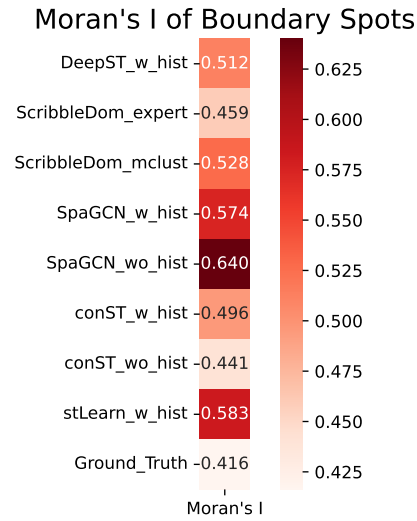
Figure S24. Multi-metric assessment of domain consistency and boundary definition. Using sample 151675, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



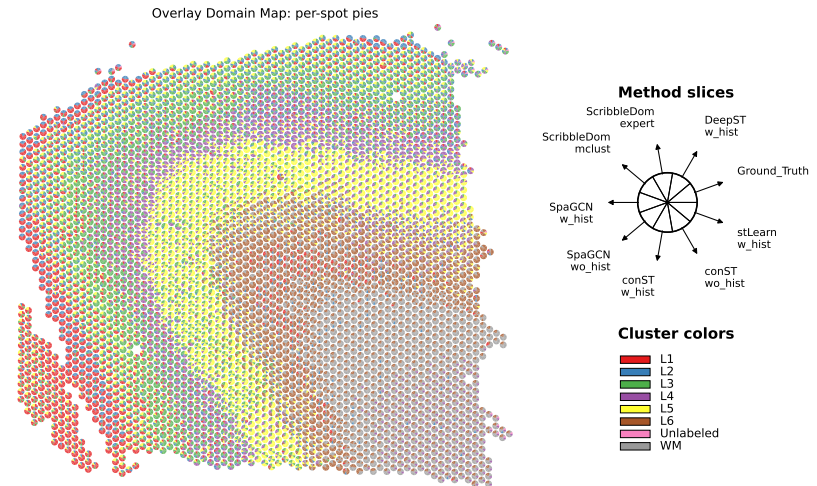
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

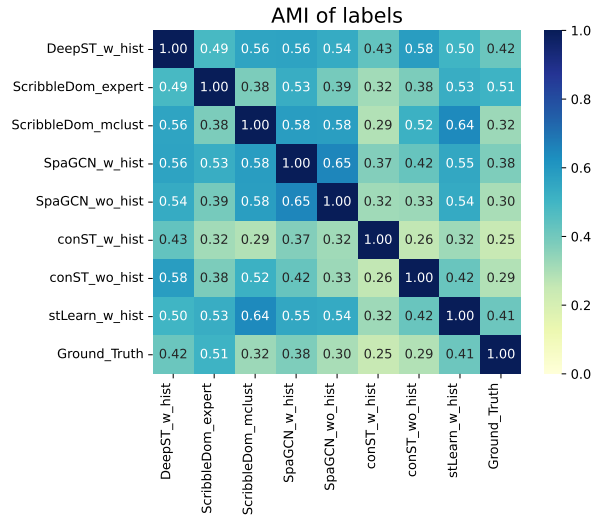


c Spatial Autocorrelation of Boundaries (Moran's I).

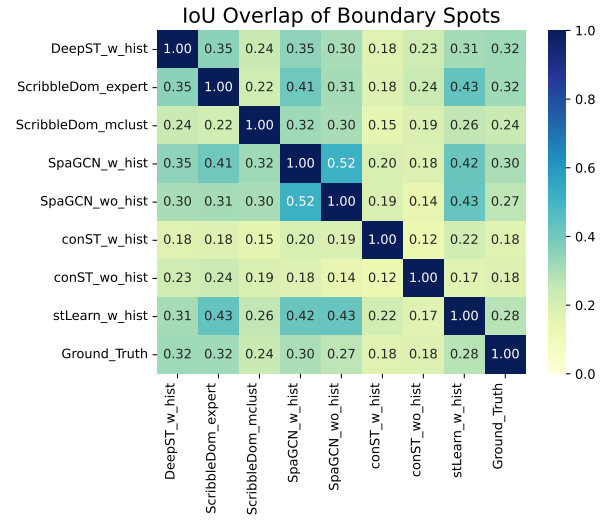


d Spatially Resolved Consensus Visualization.

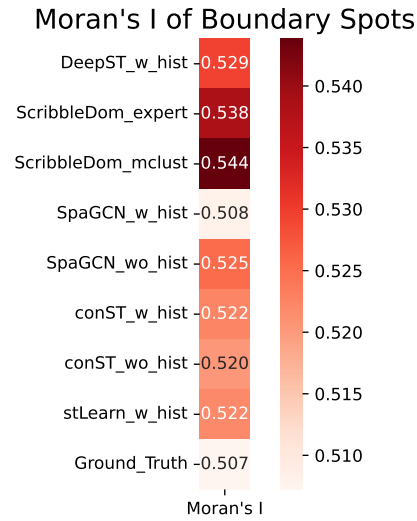
Figure S25. Multi-metric assessment of domain consistency and boundary definition. Using sample 151676, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



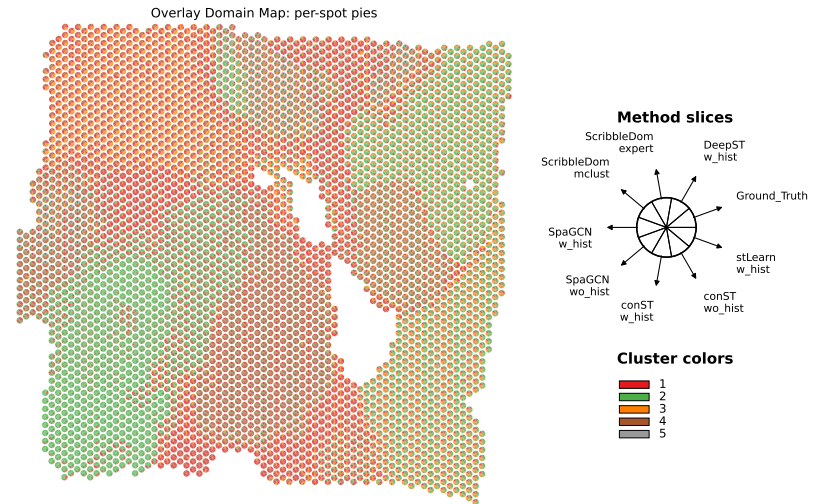
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

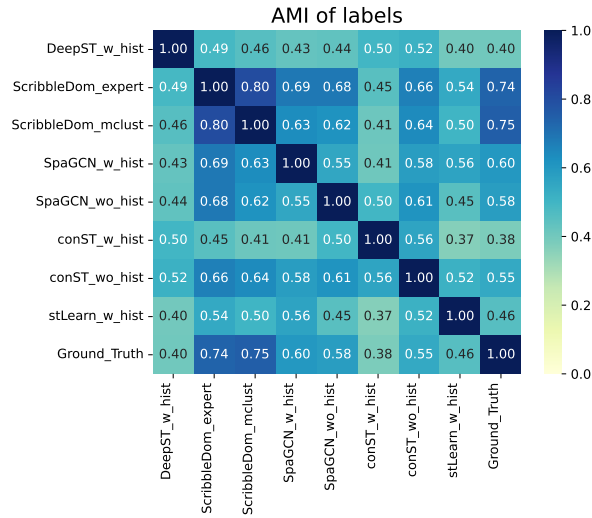


c Spatial Autocorrelation of Boundaries (Moran's I).

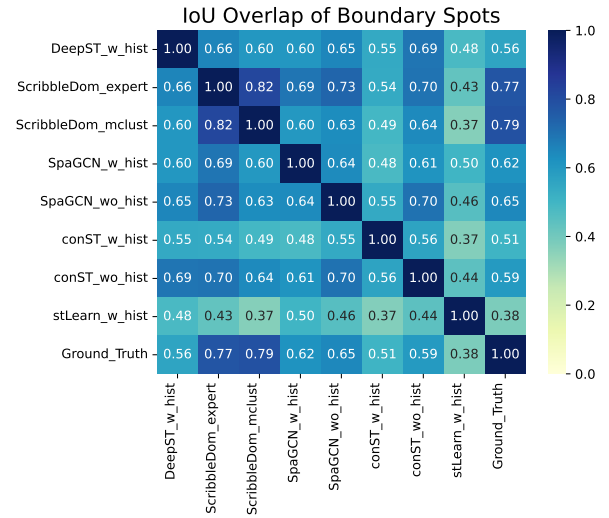


d Spatially Resolved Consensus Visualization.

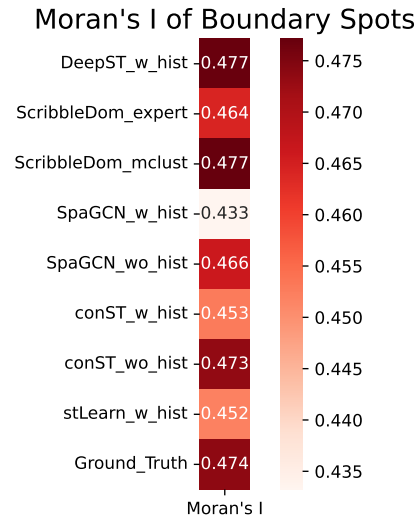
Figure S26. Multi-metric assessment of domain consistency and boundary definition. Using sample hbc, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



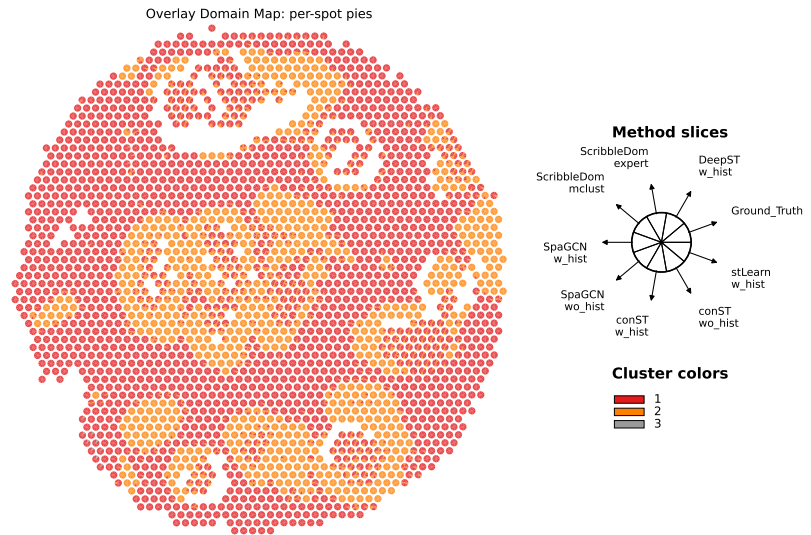
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.

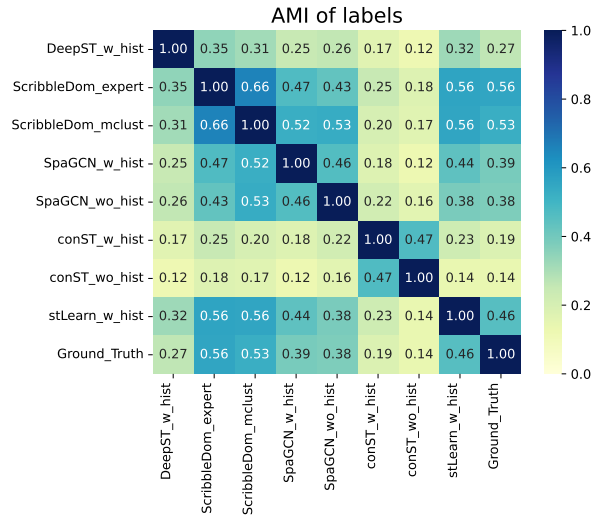


c Spatial Autocorrelation of Boundaries (Moran's I).

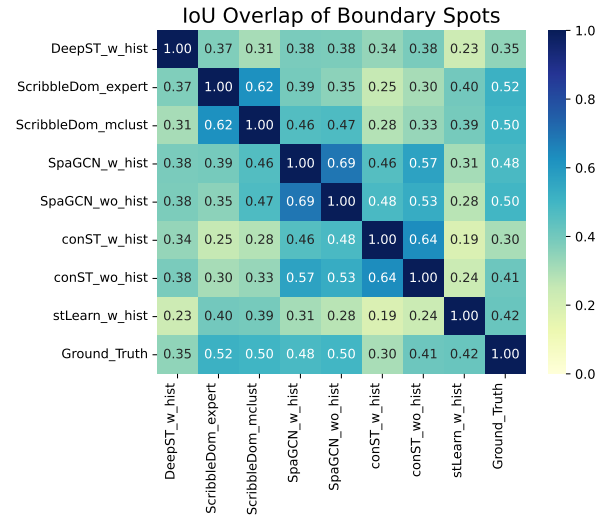


d Spatially Resolved Consensus Visualization.

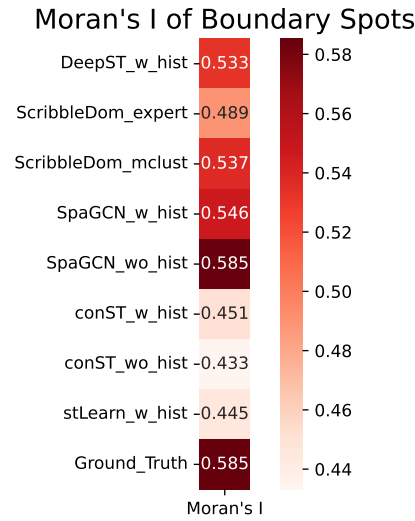
Figure S27. Multi-metric assessment of domain consistency and boundary definition. Using sample bcdc, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.



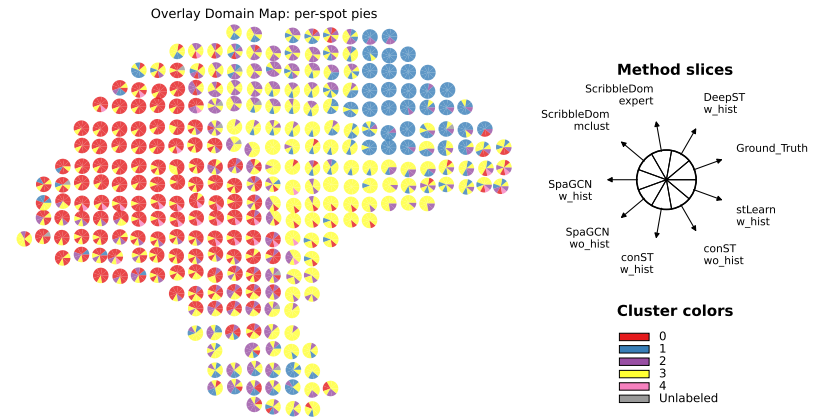
a Pairwise Adjusted Mutual Information (AMI) of labels.



b IoU Overlap of Boundary Spots.



c Spatial Autocorrelation of Boundaries (Moran's I).



d Spatially Resolved Consensus Visualization.

Figure S28. Multi-metric assessment of domain consistency and boundary definition. Using sample melanoma, we evaluate: (a) global label agreement via Adjusted Mutual Information; (b) precision of domain transitions via boundary IoU; and (c) boundary continuity via Moran's I. (d) A consensus overlay visualizes spot-level agreement, distinguishing robust domains from uncertain borders using method-specific pie slices.

References

1. Hu, J. 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, [10.1038/s41592-021-01255-8](https://doi.org/10.1038/s41592-021-01255-8) (2021).
2. Zong, Y. et al. const: an interpretable multi-modal contrastive learning framework for spatial transcriptomics. bioRxiv 2022–01, [10.1101/2022.01.14.476408](https://doi.org/10.1101/2022.01.14.476408) (2022).
3. Liang, J. et al. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision, 1833–1844 (2021).
4. Hartigan, J. A. & Wong, M. A. Algorithm as 136: A k-means clustering algorithm. J. royal statistical society. series c (applied statistics) **28**, 100–108 (1979).