Cross-Slice Knowledge Transfer via Masked Multi-Modal Heterogeneous Graph Contrastive Learning for Spatial Gene Expression Inference

Supplementary Material

S1. Related Work

Image-to-expression mapping Early approaches mainly focused on directly mapping pathology images to spatial gene expression profiles through supervised regression models. Representative works such as STNet [3] train convolutional neural networks (CNNs) to predict the expression of individual genes from local tissue patches. While these methods are computationally efficient and intuitive, they primarily capture local morphological features and ignore spatial context or relationships between neighboring spots.

Intra-Slide Spatial Modeling To overcome the locality limitation of pure image-to-expression mappings, intra-slide spatial modeling methods have been developed to capture spatial dependencies among neighboring spots within the same tissue section. For instance, HisToGene [11], Hist2ST [17], and THItoGene [8] incorporate spot-level spatial adjacency into graph-based architectures, enabling message passing between neighboring spots. HGGEP [9] and M2OST [13] further enhance spatial modeling by integrating multi-scale contextual features from pathology images, mimicking the hierarchical reasoning process of pathologists.

These intra-slide approaches demonstrate substantial performance improvements over simple CNN baselines, as they exploit the geometric structure of ST data. However, they remain constrained to single-slide analysis and cannot effectively leverage cross-slice similarities or shared patterns across different tissue samples.

Multimodal alignment A complementary line of work aims to align pathology images and gene-expression profiles within a shared latent space, enabling expression prediction through cross-modal retrieval rather than direct regression. BLEEP [15] jointly encodes H&E patches and expression vectors using a CLIP-style [12] contrastive objective and infers expression by retrieving and aggregating the nearest reference profiles. mclSTExp [10] similarly learns a joint embedding by integrating spatial context through a Transformer encoder and aligning patch–spot pairs via contrastive learning, allowing expression estimation through nearest-neighbor aggregation. EGGN [16] further incorporates exemplar retrieval into a graph framework, linking windows to matched exemplars and propagating molecular information through message passing.

S2. Experimental Datasets

In our experiments, we evaluated the performance of the proposed SpaHGC across seven publicly available matched image-ST datasets covering a variety of cancer types and tissue contexts (Table S1). These include one human skin squamous cell carcinoma (cSCC) dataset [7] from the traditional ST platform with 12 samples; three breast cancer datasets, comprising the HER2+ subtype from the ST platform with 36 samples [1], and two Visium platform datasets—Alex (subtype: TNBC) [14] with 4 samples and Visium BC (subtype: HER2+) [5] with 3 samples. Additionally, we incorporated the HEST-1k [6] dataset to further evaluate the generalizability of our model across diverse tissue types. We selected 4 lymph node samples and two pancreatic cancer datasets (Pancreas1 and Pancreas2), containing 4 and 3 samples respectively.

The traditional ST platform typically samples several hundred spatially resolved spots per tissue section, each measuring expression levels for approximately 20,000 genes. In contrast, the Visium platform offers enhanced spatial resolution and sampling density, enabling the capture of thousands of spots per section while maintaining comparable gene coverage per spot.

S2.1. Data Preprocessing

Following Li et al. [11], we adopted the following preprocessing pipeline for gene selection. For each tissue slice, we first applied the Scanpy toolkit to identify the top 1,000 highly variable genes (HVGs) using the function *scanpy.pp.highly_variable_genes* on log-normalized counts. To obtain a consistent feature space across all slices within the same dataset, we then computed the intersection of the HVG sets from all slices and restricted subsequent analyses to this shared HVG set. As a result, the size of the shared HVG set varied across datasets, ranging from approximately 200 to 800 genes.

Subsequently, for each spot in all datasets, we normalized the gene expression counts by dividing each gene's raw count by the total expression within that spot, followed by multiplication with a scaling factor of 1,000,000. The normalized values are subsequently transformed using the natural logarithm in the form of $\log(1+x)$, where x denotes the normalized expression value. For image preprocessing, we extract 224×224 pixel patches from the pathology slides based on the centroid coordinates of each spot, and use a pretrained pathology image model to extract their visual feature representations.

S3. Baselines

To evaluate the performance of SpaHGC, we compared it with nine state-of-the-art methods methods as follows:

STNet [3] is one of the earliest methods to explore image-driven expression prediction. It utilizes DenseNet-121 [4] to extract features from image patches corresponding to each spot and employs fully connected layers to regress the expression levels of multiple target genes.

HisToGene [11] introduces the Vision Transformer (ViT) [2], leveraging self-attention to capture global dependencies between image patches and enhance spatial structure representation.

Hist2ST [17] crops an image patch around each sequenced spot and uses a Convmixer module to capture intra-patch 2D visual features. Spot coordinates are encoded and fused with the features, and a Transformer models global spatial dependencies across the slide. A GraphSAGE layer then aggregates local neighborhood information on a 4-nearest-neighbor graph, with LSTM-based layer aggregation to combine multi-layer representations. Gene expression is modeled via a ZINB likelihood with an additional MSE loss, and a data-augmentation-driven self-distillation strategy is employed to improve data efficiency.

EGGN [16] predicts spot-level gene expression from pathology images by coupling exemplar retrieval with graph neural inference. Each slide is partitioned into windows, which are embedded with a ResNet-18 backbone to define a retrieval space; for every window, its nearest exemplars—with known expression profiles—are retrieved from a reference bank. A graph is then constructed whose nodes include both windows and exemplars, with four edge types encoding window–window spatial proximity, exemplar–exemplar similarity, and the bidirectional links between windows and their matched exemplars. Over this graph, a GraphSAGE backbone performs message passing within the window and exemplar subgraphs, while a Graph Exemplar Bridging block explicitly injects exemplar knowledge into window features using the exemplars' expression signals. Finally, an attention-based prediction head weights each window's exemplars and aggregates them with the updated window representation to produce gene-expression vectors for all windows.

THItoGene [8] first partitions each pathology images slide into spot-centered image patches and uses dynamic convolution to adaptively extract fine-grained, multi-scale visual representations. It then introduces an Efficient-CapsNet with attention-based routing to capture the hierarchical spatial relationships of cellular morphology and tissue architecture, thereby strengthening the modeling of the linkage between image phenotypes and gene expression. Next, spot coordinate embeddings are fused with deep visual features, and a ViT is employed to model global, cross-region dependencies. Finally, a 4-nearest-neighbor graph is constructed based on Euclidean distances between spots, over which a GAT performs attention-based aggregation to explicitly learn the correspondence between spatial neighborhoods and gene expression, yielding spot-level expression predictions.

HGGEP [9] is a hypergraph neural network for predicting gene expression from pathology images. It first partitions each whole-slide image into spot-centered patches and enhances morphological cues with a Gradient Enhancement Module based on difference convolution. A lightweight ShuffleNet-V2 backbone then extracts multi-stage features, which are refined at each stage by channel and spatial attention and further processed with a Vision Transformer to capture long-range, cross-spot dependencies. To model higher-order relationships across stages, a Hypergraph Association Module constructs hyperedges using both Euclidean distance and local positional metrics, enabling joint reasoning over groups of spots. The resulting hypergraph representations are concatenated with ViT features and passed to an LSTM to promote inter-stage information flow, and a final MLP outputs spot-level gene expression predictions.

BLEEP [15] formulates pathology images-to-expression prediction as bi-modal representation learning followed by retrieval-based imputation: paired H&E patches and spot-level expression profiles are encoded by an image encoder and an expression encoder into a shared low-dimensional space using a CLIP-style [12] contrastive objective that is smoothed by incorporating intra-batch image–image and expression–expression similarities, which prevents near-duplicate spots from being pushed apart in training; at inference, each query patch is projected into the joint space and its k nearest reference expression profiles are retrieved and linearly combined to yield the predicted expression vector, thereby shifting the task from per-gene regression to joint profile interpolation while leveraging morphological–molecular alignment learned from the reference.

mclSTExp [10] predicts gene expression from pathology images by learning a joint embedding between image patches and spatially contextualized spot expression. The model first encodes each spot-centered H&E patch with an image encoder and projects it into a latent space, while treating spots as "tokens" and the whole slide as a "sentence" so that a Transformer with self-attention produces semantically consistent spatial representations. It then aligns the two modalities, bringing matched patch–spot pairs closer and pushing mismatches apart in cosine space, thereby yielding a multimodal representation in which morphology and expression are co-registered. At inference time, a query patch is mapped into this shared space, its top-k nearest reference spots are retrieved, and their ground-truth expression profiles are aggregated with weights to produce the gene-expression prediction for the corresponding slide.

M2OST [13] formulates pathology images to expression prediction as a many-to-one regression problem in which multi-level pathology images patches that share a common target spot are processed jointly to produce a single gene-expression vector. The model employs a decoupled multi-scale encoder: deformable patch embedding emphasizes the spot's central region while generating fine-grained in-spot tokens and coarser surrounding tokens; intra-level token mixing performs ViT-style self-attention within each scale; cross-level token mixing introduces fully connected cross-scale attention to exchange information across scales with different sequence lengths; and cross-level channel mixing applies squeeze–excitation to fuse channels in a length-agnostic manner. The [CLS] tokens from all levels are finally concatenated and fed to a linear head to output spot-level expressions, and the architecture naturally accommodates variable numbers of inputs while leveraging nearby inter-spot cues inherent in multi-resolution pathology images.

S4. Evaluation Metrics

We use Pearson Correlation Coefficient (PCC) and Root Mean Square Error (RMSE) to evaluate the predictive performance of the model. Mathematically, PCC and RMSE can be described as:

$$\mathrm{PCC} = \frac{\mathrm{Cov}(\mathbf{y}_t, \hat{\mathbf{y}}_t)}{\mathrm{Var}(\mathbf{y}_t) \cdot \mathrm{Var}(\hat{\mathbf{y}}_t)} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_t - \hat{\mathbf{y}}_t)^2} \tag{2}$$

where $\mathbf{y}_t$ denotes the ground truth values, $\hat{\mathbf{y}}_t$represents the predicted values, $Cov(\cdot,\cdot)$ is the covariance, $Var(\cdot)$ denotes the variance.

In addition, we employ the Adjusted Rand Index (ARI) to evaluate the accuracy of the clustering results:

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \frac{\left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}, \tag{3}$$

where a_i and b_j represent the number of samples assigned to the i-th predicted cluster and the j-th ground truth cluster, respectively, while n_{ij} denotes the number of samples that are shared between the i-th predicted cluster and the j-th ground truth cluster.

S5. Implementation Details

The SpaHGC model consists of 4 layers of GraphSAGE and CNDA modules, along with one CNAP module employing 4-head multi-head attention. The model was trained for 200 epochs on all datasets, with a learning rate of 1×10^{-4} and weight decay of 1×10^{-4}. The number of neighbor nodes was set to 5 (Q), and the number of reference nodes was set to 7 (K). In the augmented views, the feature masking ratios were set to 0.8 (α) for target nodes and 0.9 (β) for reference nodes, respectively. All baseline methods used the default hyperparameters reported in their respective papers. All models were trained on a Windows operating system with Python 3.9 and PyTorch 2.1.2, using an NVIDIA RTX 4090 GPU with 24GB of memory.

S6. Computational Time vs. Model Complexity

We evaluated the training and inference times of various models on the cSCC, Visium BC, and Pancreas2 datasets. To ensure a fair comparison, all models were run under the same hardware environment (NVIDIA RTX 4090 GPU) using the default hyperparameter settings recommended in their respective publications. As shown in Supplementary Figure S12, SpaHGC achieves the fastest computational time across all datasets. In terms of model complexity, it ranks second in the number of trainable parameters, slightly higher than M2OST [13].

References

[1] Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, et al. Spatial deconvolution of her2-positive breast tumors reveals novel intercellular relationships. *bioRxiv*, pages 2020–2084, 2020. 1, 5

[2] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, pages 1–22, 2021. 2

[3] Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4(8):827–834, 2020. 1, 2

[4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, et al. Densely connected convolutional networks. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. 2

[5] 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 analy-

sis. *Nature Communications*, 14(1):8353–8368, 2023. 1, 5

[6] Guillaume Jaume, Paul Doucet, Andrew Song, et al. Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems*, 37:53798–53833, 2024. 1, 5

[7] Andrew L Ji, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 182(2):497–514, 2020. 1, 5

[8] Yuran Jia, Junliang Liu, Li Chen, et al. THItoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1):464–474, 2024. 1, 2

[9] Bo Li, Yong Zhang, Qing Wang, et al. Gene expression prediction from histology images via hypergraph neural networks. *Briefings in Bioinformatics*, 25(6): 500–511, 2024. 1, 2

[10] Wenwen Min, Zhiceng Shi, Jun Zhang, et al. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Briefings in Bioinformatics*, 25(6):551–563, 2024. 1, 2

[11] Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *BioRxiv*, pages 1–31, 2021. 1, 2

[12] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PmLR, 2021. 1, 2

[13] Hongyi Wang, Xiuju Du, Jing Liu, et al. M2ost: Many-to-one regression for predicting spatial transcriptomics from digital pathology images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7709–7717, 2025. 1, 3

[14] Sunny Z Wu, Ghamdan Al-Eryani, Daniel Lee Roden, Simon Junankar, Kate Harvey, Alma Andersson, Aatish Thennavan, Chenfei Wang, James R Torpy, Nenad Bartonicek, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics*, 53(9):1334–1347, 2021. 1, 5

[15] Ronald Xie, Kuan Pang, Sai Chung, et al. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems*, 36:70626–70637, 2023. 1, 2

[16] Yan Yang, Md Zakir Hossain, Eric Stone, et al. Spatial transcriptomics analysis of gene expression prediction using exemplar guided graph neural network. *Pattern Recognition*, 145:1–10, 2024. 1, 2

[17] Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):297–309, 2022. 1, 2

Table S1. Summary of seven matched image-ST datasets used in this study.

Dataset Name	Reference	Cancer	Subtype	Platform	Slice Num
cSCC	[7]	Squamous Cell Carcinoma	-	ST	12
HER2+	[1]	Breast Cancer	HER2+	ST	36
Alex	[14]	Breast Cancer	TNBC	Visium	4
Visium BC	[5]	Breast Cancer	HER2+	Visium	3
Lymph Node	[6]	Lymph Node	-	Visium	4
Pancreas1	[6]	Pancreatic Cancer	-	Visium	3
Pancreas2	[6]	Pancreatic Cancer	-	Visium	4

Table S2. Wilcoxon signed-rank test p-values comparing SpaHGC with competing methods across different datasets. Each p-value is computed based on paired comparisons of PCC scores across all tissue slices within the dataset.

P-value	HER2+	cSCC	Alex	Visium BC	Lymph Node	Pancreas1	Pancreas2
SpaHGC vs. STNet	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. HisToGene	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. His2ST	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. EGGN	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. THItoGene	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. HGGEP	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. Bleep	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. mclSTExp	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SpaHGC vs. M2OST	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table S3. Ablation study on the number of neighbor nodes Q.

Q Value	Lymph Node		Pancreas1		Pancreas2	
	PCC	RMSE	PCC	RMSE	PCC	RMSE
Q=1	33.81±0.002	0.244±0.004	22.44±0.001	0.185±0.001	36.67±0.003	0.237±0.002
Q=2	34.56±0.001	0.235±0.008	22.94±0.005	0.185±0.002	40.41±0.002	0.225±0.013
Q=3	34.89±0.001	0.239±0.001	23.62±0.003	0.176±0.001	40.99±0.002	0.219±0.005
Q=4	34.97±0.003	0.227±0.002	23.51±0.002	**0.174±0.001**	41.21±0.001	0.219±0.003
Q=5	**35.02±0.003**	0.225±0.004	**24.48±0.005**	0.179±0.003	**41.76±0.003**	**0.213±0.006**
Q=6	34.78±0.002	0.227±0.002	24.02±0.003	0.175±0.001	41.23±0.001	0.215±0.002
Q=7	34.22±0.002	0.224±0.007	24.26±0.001	0.172±0.002	41.04±0.002	0.216±0.001
Q=8	34.40±0.002	0.226±0.001	23.59±0.003	0.175±0.003	41.26±0.001	0.215±0.001
Q=9	34.86±0.001	**0.221±0.008**	23.80±0.003	0.176±0.001	41.32±0.002	0.219±0.003

Table S4. Ablation study on the number of reference nodes K.

K Value	Lymph Node		Pancreas1		Pancreas2	
	PCC	RMSE	PCC	RMSE	PCC	RMSE
K=1	33.76±0.002	0.235±0.002	23.49±0.001	0.180±0.018	41.19±0.001	0.214±0.002
K=2	34.24±0.001	0.230±0.004	23.50±0.001	0.182±0.002	41.39±0.003	0.214±0.001
K=3	34.33±0.002	0.231±0.004	23.09±0.002	0.184±0.001	41.54±0.001	0.213±0.001
K=4	34.21±0.001	0.228±0.003	23.51±0.001	0.183±0.002	41.73±0.003	0.216±0.001
K=5	34.43±0.001	**0.222±0.005**	24.29±0.003	0.182±0.003	41.76±0.002	0.212±0.009
K=6	34.27±0.003	0.226±0.001	24.15±0.003	0.185±0.002	41.75±0.004	0.219±0.007
K=7	**35.02±0.003**	0.225±0.004	**24.48±0.005**	**0.179±0.003**	**41.76±0.003**	**0.213±0.006**
K=8	34.49±0.002	0.228±0.001	24.05±0.002	0.186±0.001	41.61±0.001	0.213±0.011
K=9	34.86±0.001	0.226±0.008	24.30±0.004	0.188±0.017	41.59±0.002	0.191±0.001

Table S5. Ablation study on the masking ratios α (for neighbor nodes) and β (for reference nodes).

α Value	Lymph Node		Pancreas1		Pancreas2	
	PCC	RMSE	PCC	RMSE	PCC	RMSE
α=0.1	34.17±0.002	**0.220±0.002**	23.20±0.003	0.182±0.002	41.62±0.001	0.219±0.003
α=0.2	34.22±0.003	0.223±0.002	23.19±0.004	0.180±0.001	41.33±0.001	0.219±0.003
α=0.3	34.04±0.002	0.228±0.002	23.00±0.005	0.179±0.001	41.18±0.002	0.217±0.004
α=0.4	34.08±0.002	0.227±0.001	24.82±0.003	0.181±0.001	41.30±0.002	0.214±0.003
α=0.5	34.19±0.003	0.226±0.001	23.95±0.003	0.181±0.002	41.30±0.003	0.215±0.003
α=0.6	34.42±0.003	0.226±0.002	24.31±0.005	0.179±0.001	41.03±0.001	0.219±0.004
α=0.7	34.73±0.002	0.233±0.004	24.01±0.001	0.179±0.001	41.19±0.002	0.221±0.001
α=0.8	**35.02±0.003**	0.225±0.004	**24.48±0.005**	**0.179±0.003**	**41.76±0.003**	**0.213±0.006**
α=0.9	34.70±0.001	0.225±0.001	23.65±0.004	0.183±0.004	41.42±0.001	0.222±0.005
β Value	Lymph Node		Pancreas1		Pancreas2	
	PCC	RMSE	PCC	RMSE	PCC	RMSE
β=0.1	34.98±0.002	0.227±0.001	23.19±0.004	0.170±0.001	40.89±0.001	0.221±0.002
β=0.2	34.73±0.001	0.225±0.007	23.17±0.001	0.169±0.001	40.79±0.002	0.222±0.001
β=0.3	34.82±0.002	0.223±0.003	23.20±0.003	0.170±0.004	40.70±0.001	0.226±0.006
β=0.4	34.81±0.001	0.228±0.002	23.41±0.003	0.168±0.002	**41.88±0.002**	**0.213±0.005**
β=0.5	34.69±0.003	**0.221±0.005**	24.29±0.001	0.169±0.002	41.66±0.001	0.214±0.007
β=0.6	34.46±0.001	0.223±0.002	24.28±0.003	0.171±0.001	41.52±0.001	0.216±0.003
β=0.7	34.33±0.001	0.229±0.006	24.43±0.001	0.168±0.001	41.59±0.001	0.216±0.002
β=0.8	34.91±0.001	0.229±0.003	24.43±0.003	0.169±0.001	41.47±0.002	0.218±0.005
β=0.9	**35.02±0.003**	0.225±0.004	**24.48±0.005**	**0.179±0.003**	41.76±0.003	0.213±0.006

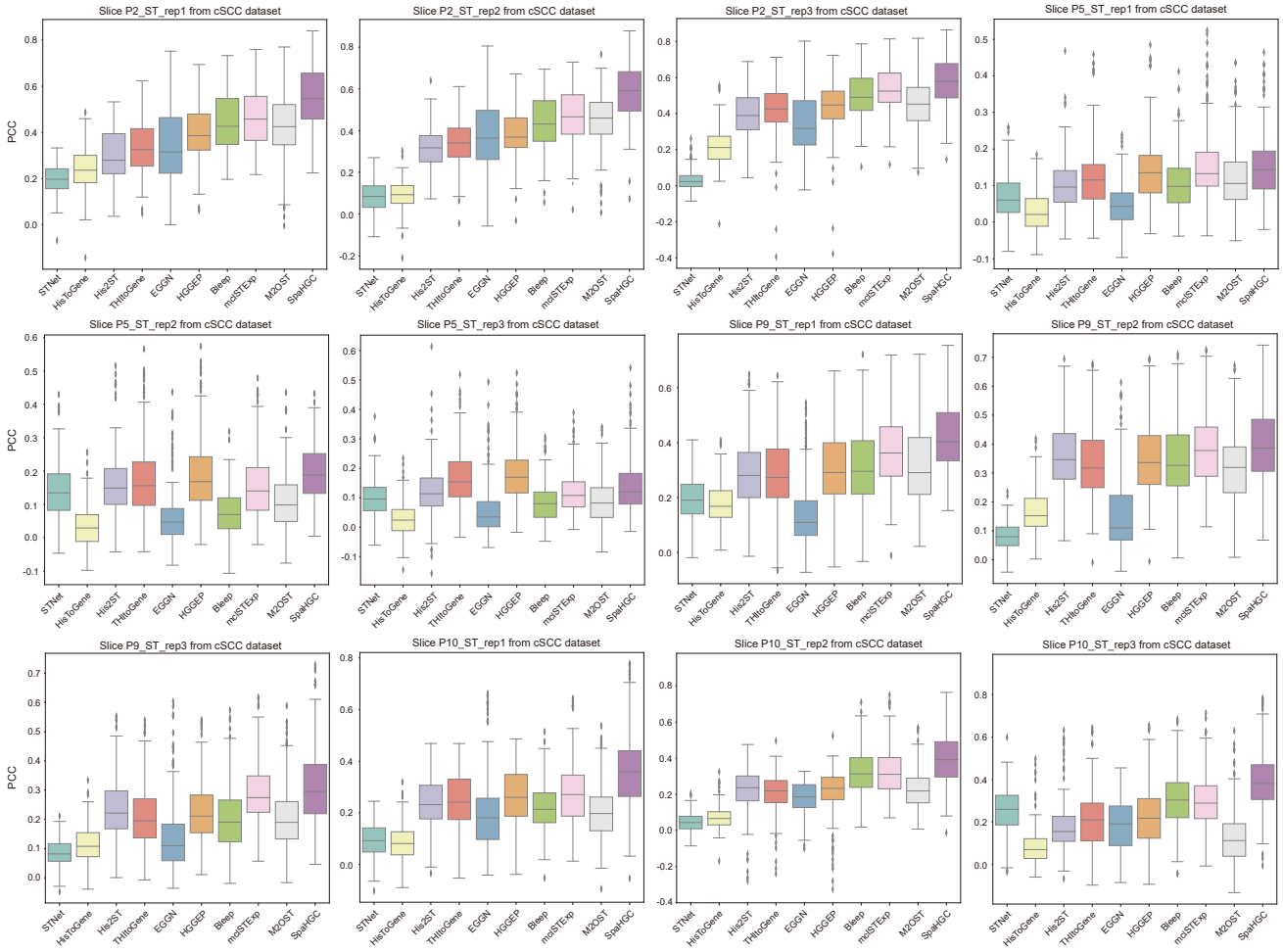


Figure S1. Per-slice PCC scores of predicted gene expression on the cSCC datasets.

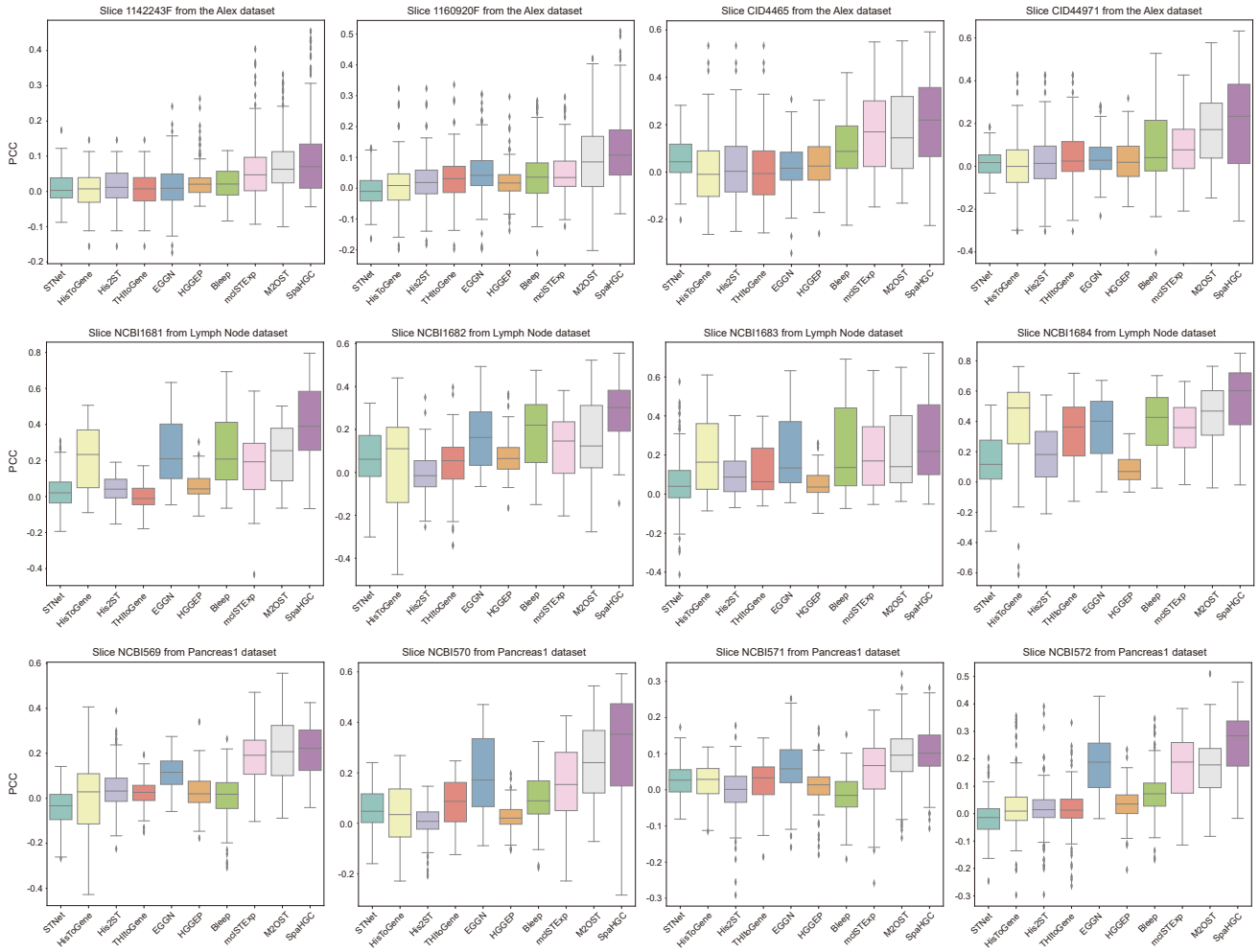


Figure S2. Per-slice PCC scores of predicted gene expression across the Alex, Lymph Node, and Pancreas1 datasets.

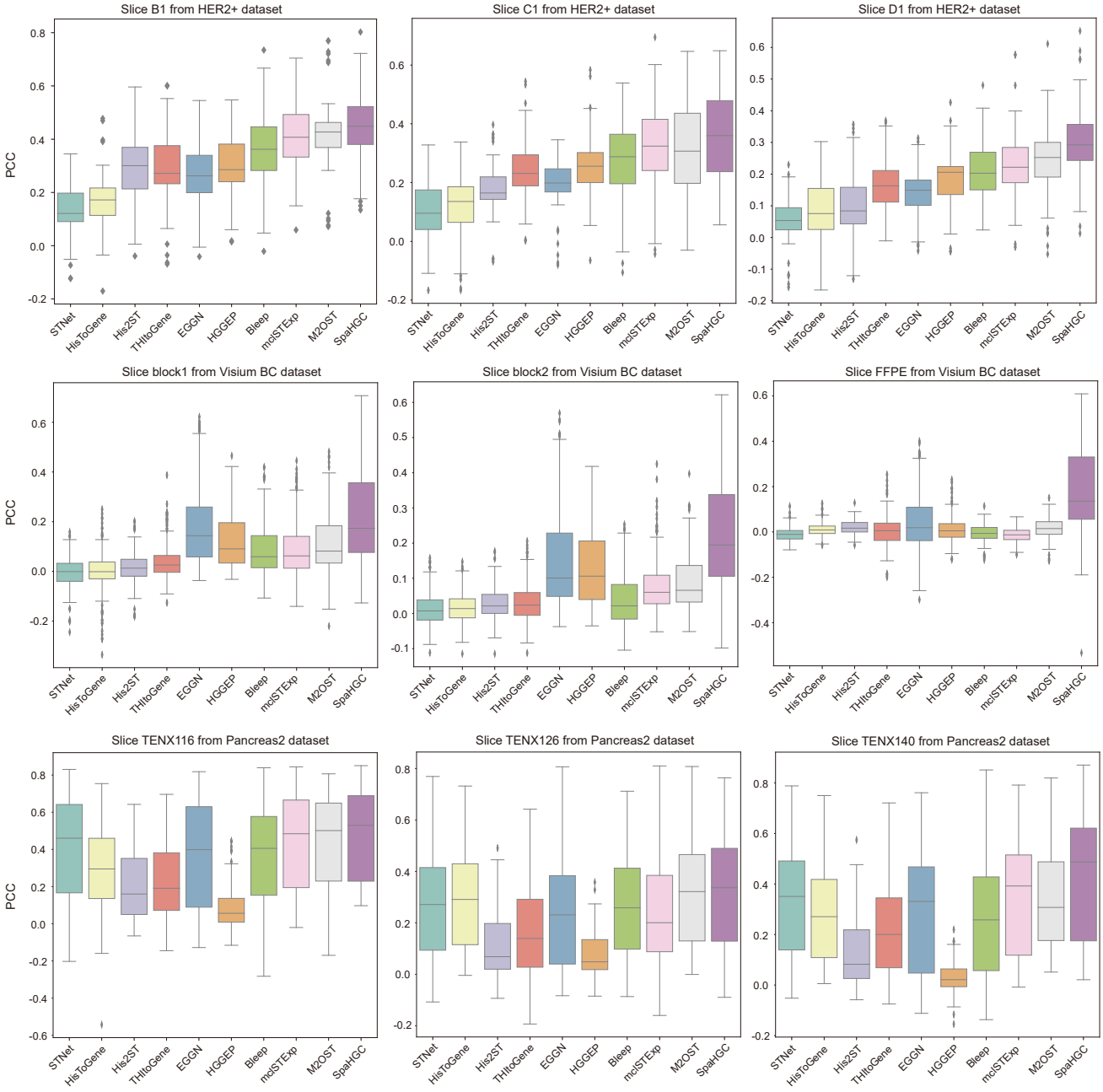


Figure S3. Per-slice PCC scores of predicted gene expression across the HER2+, Visium BC, and Pancreas2 datasets.

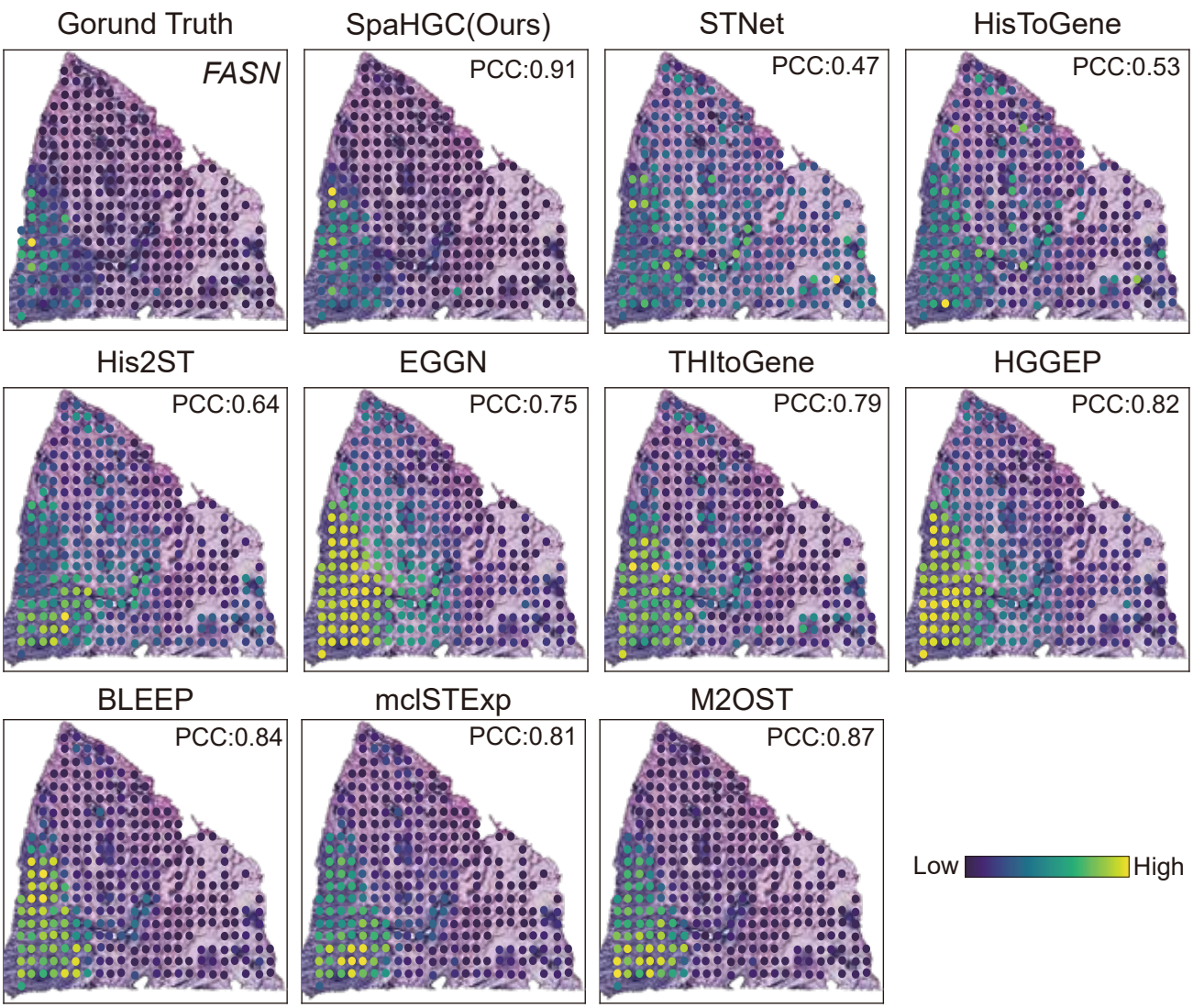


Figure S4. Visualization of predicted expression patterns for marker genes *FASN* on the HER2+ dataset.

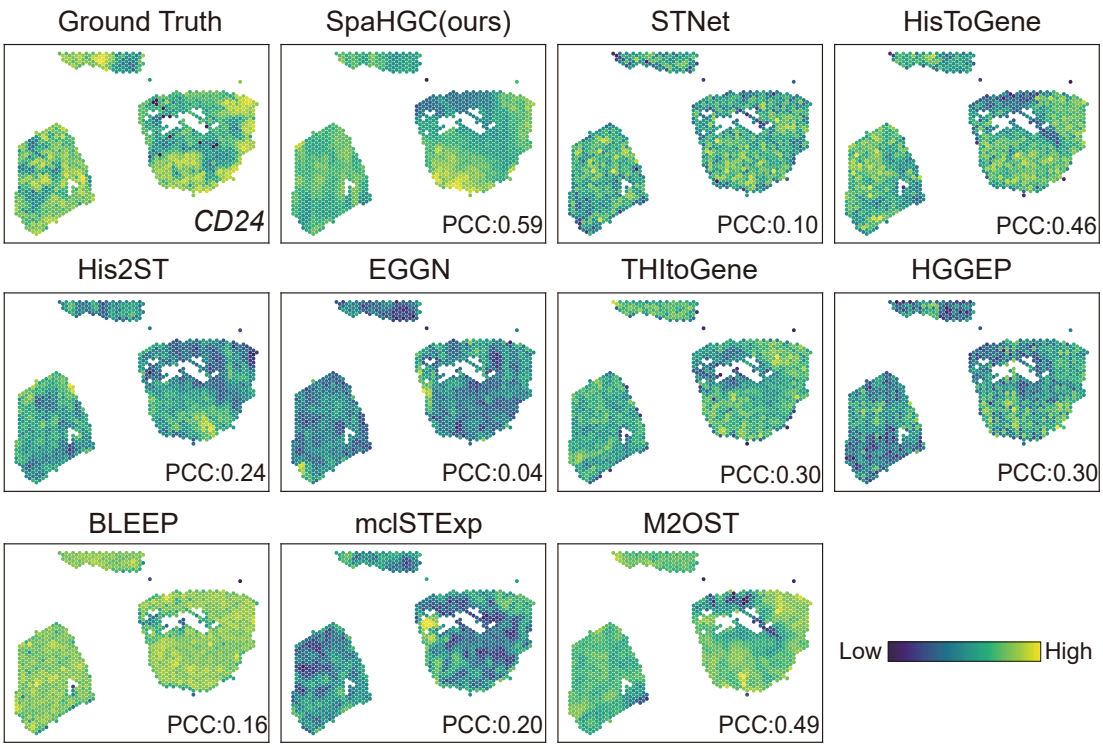


Figure S5. Visualization of predicted expression patterns for marker genes *CD24* on the Alex dataset.

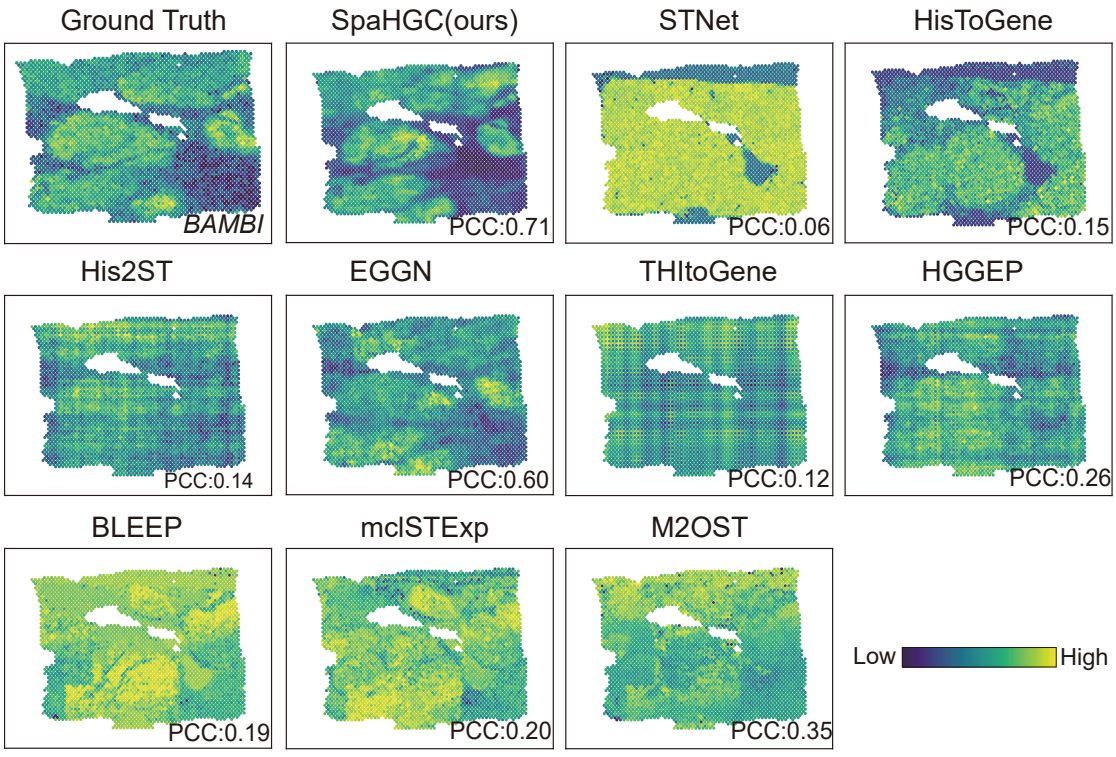


Figure S6. Visualization of predicted expression patterns for marker genes *BAMBI* on the Visium BC dataset.

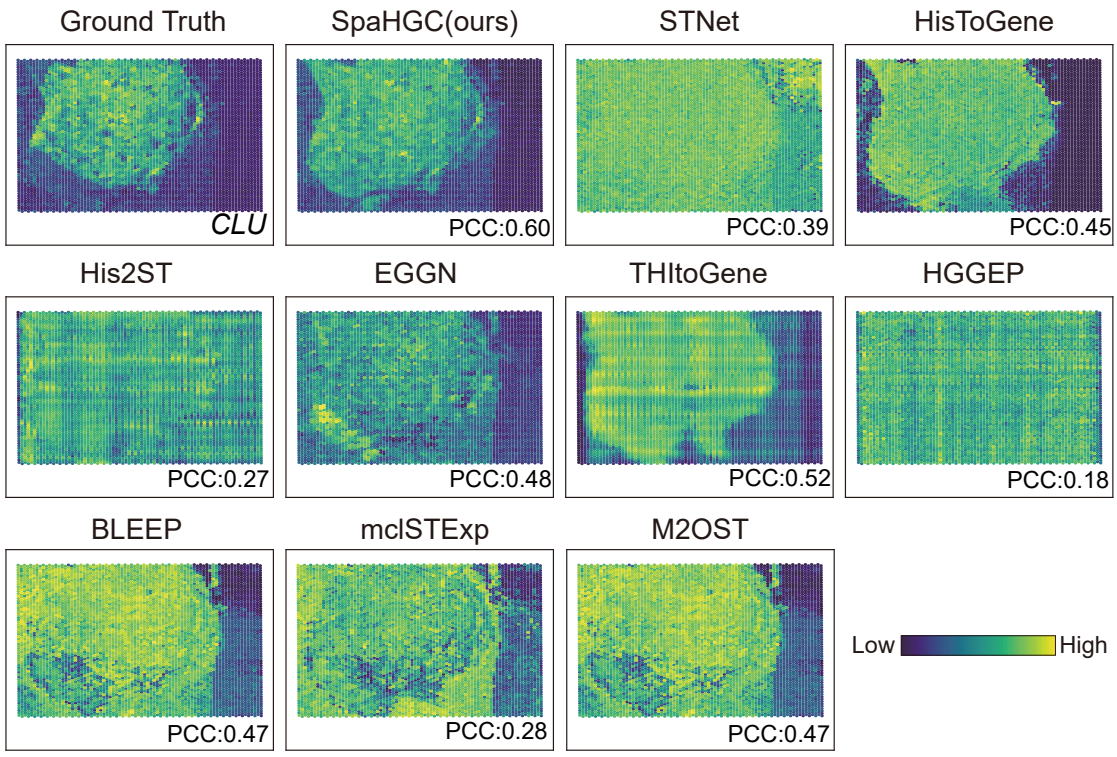


Figure S7. Visualization of predicted expression patterns for marker genes *CLU* on the Lymph Node dataset.

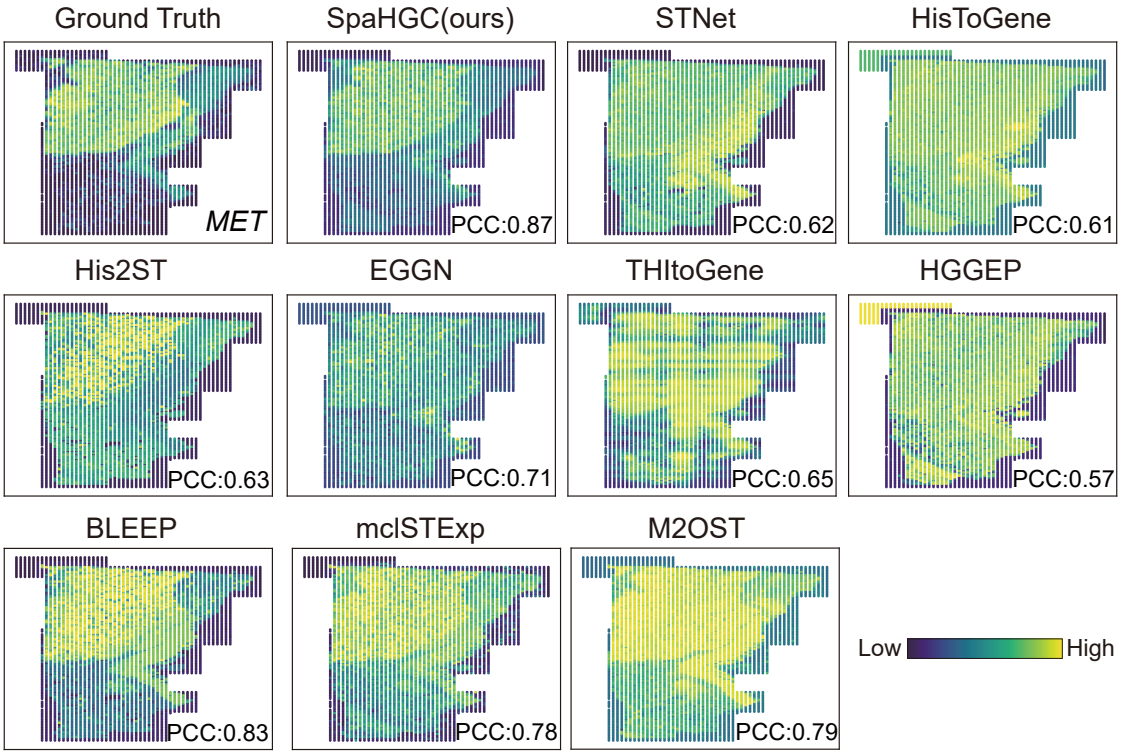


Figure S8. Visualization of predicted expression patterns for marker genes *MET* on the Pancreas2 dataset.

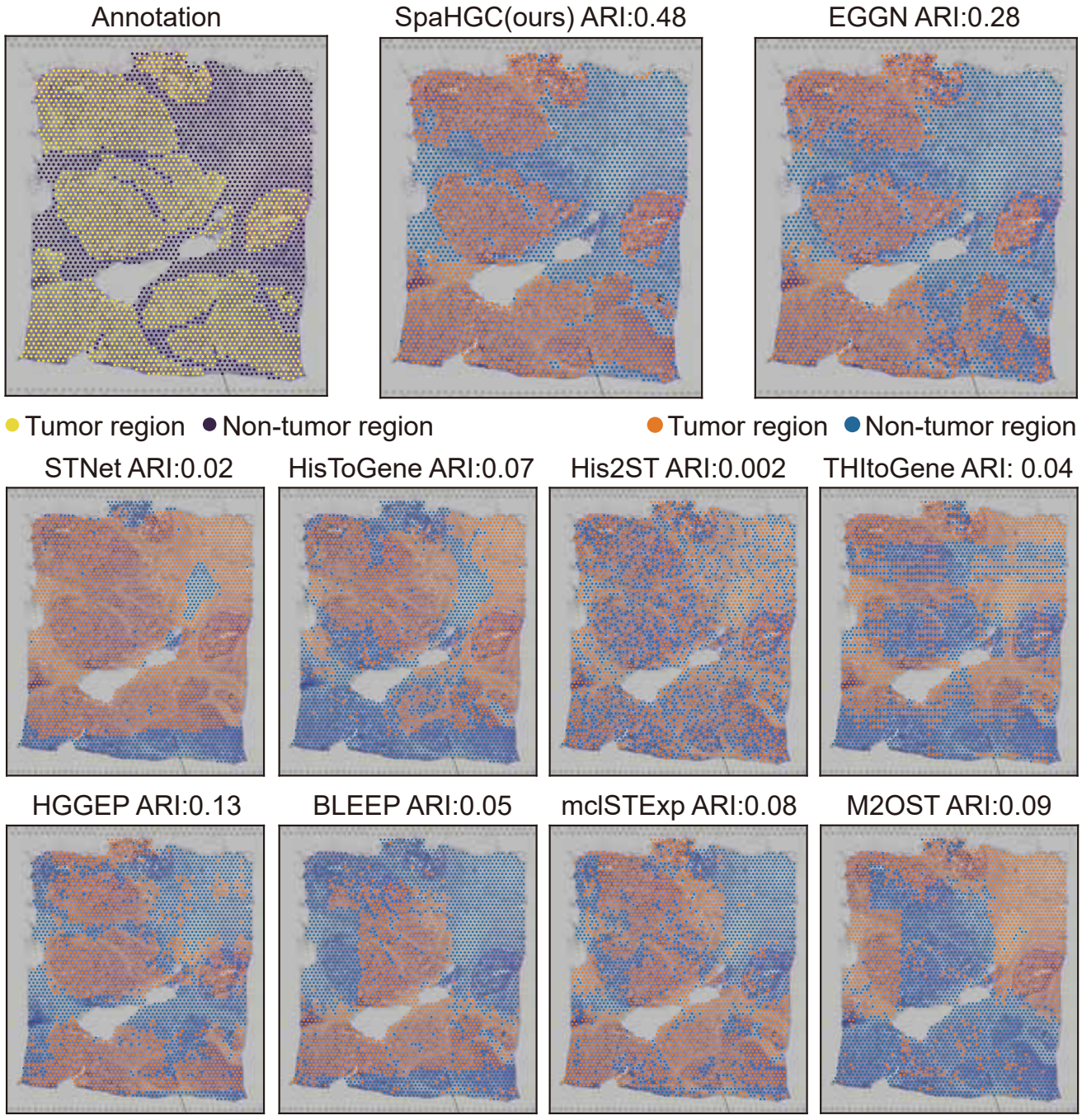


Figure S9. Clustering analysis of predicted gene expression profiles on the Visium BC dataset. Spots were grouped into cancerous and normal tissue regions based on pathologist annotations.

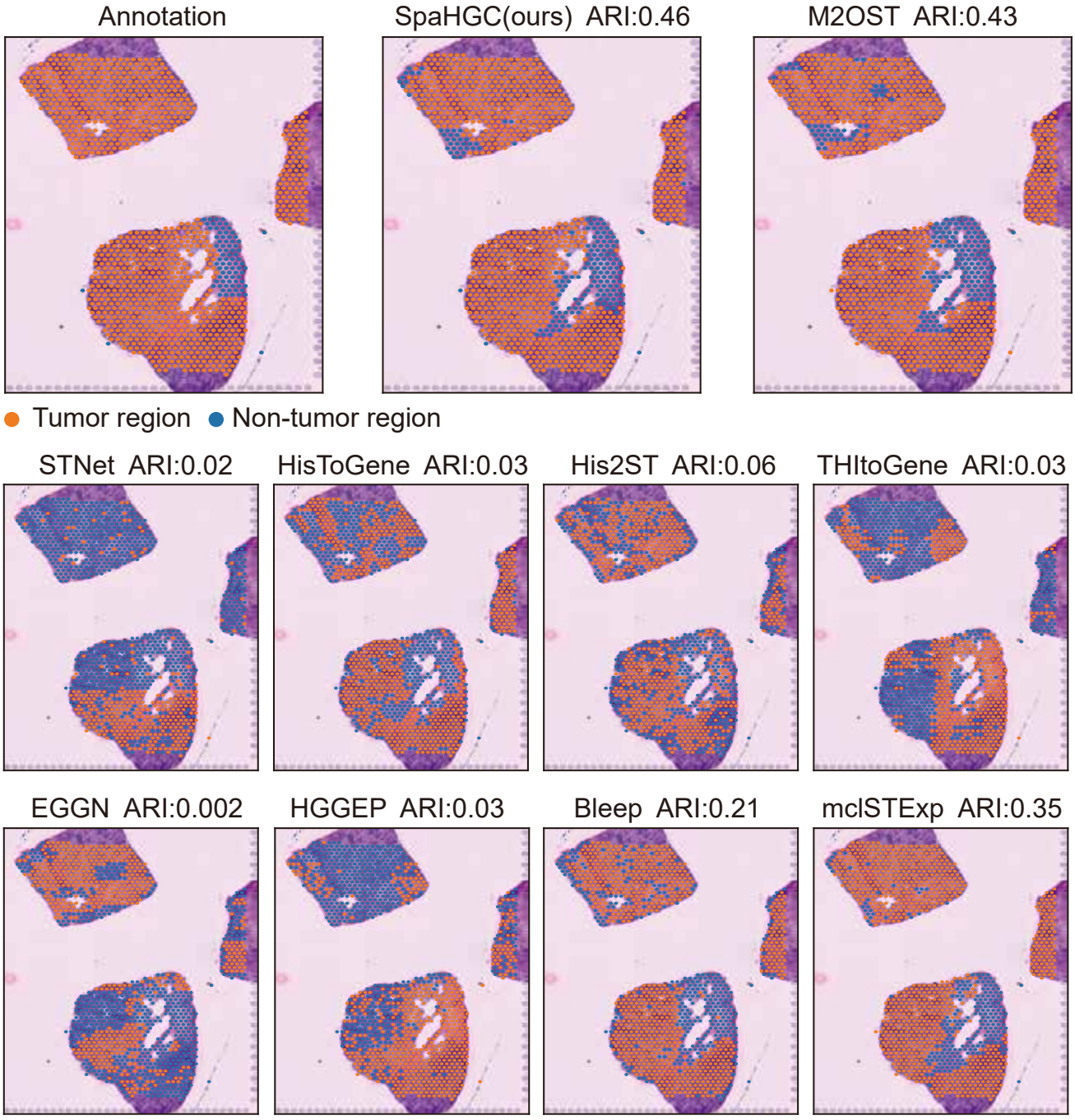
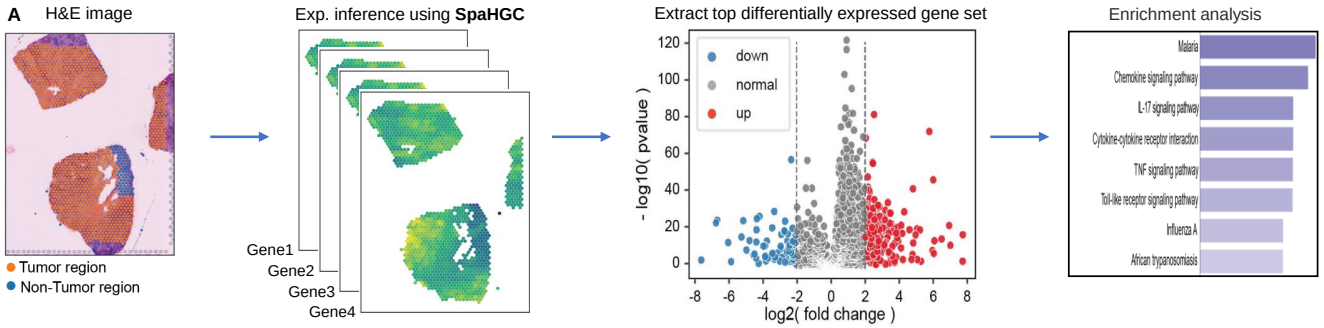
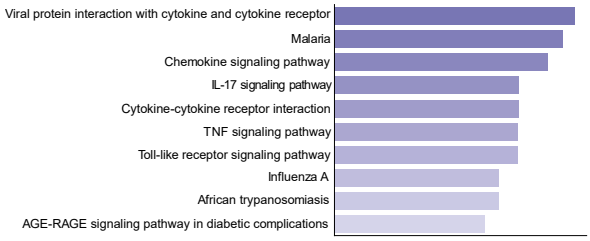


Figure S10. Clustering analysis of predicted gene expression profiles on the Alex dataset. Spots were grouped into cancerous and other tissue regions based on pathologist annotations.



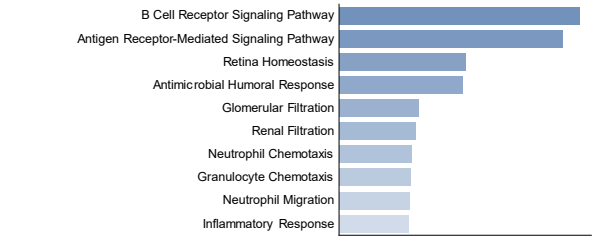
B Top 10 significant KEGG pathways with **SpaHGC**



C *P*-values of significantly enriched KEGG pathways with **SpaHGC** compared to raw data

Pathway	Raw(P-value)	SpaHGC(P-value)
Viral protein interaction with cytokine and cytokine receptor	9.279683E-01	3.298301E-06
Malaria	3.797396E-02	1.894202E-05
Chemokine signaling pathway	4.714789E-05	4.334966E-05
IL-17 signaling pathway	8.485278E-02	1.257550E-04
Cytokine-cytokine receptor interaction	9.999757E-01	2.276861E-04
TNF signaling pathway	1.037439E-05	2.112447E-04
Toll-like receptor signaling pathway	1.808792E-03	1.696942E-04
Influenza A	8.111346E-08	7.417312E-04
African trypanosomiasis	7.612932E-01	6.839904E-04
AGE-RAGE signaling pathway in diabetic complications	4.539283E-06	4.885387E-03

D Top 10 significant Gene Ontology Biological Process with **SpaHGC**



E *P*-values of significantly enriched GO:BP pathways with **SpaHGC** compared to raw data

Pathway	Raw(P-value)	SpaHGC(P-value)
B Cell Receptor Signaling Pathway	7.847457E-01	1.056927E-03
Antigen Receptor-Mediated Signaling Pathway	3.510977E-03	3.580439E-04
Retina Homeostasis	4.442353E-01	4.018074E-02
Antimicrobial Humoral Response	8.440635E-01	3.298301E-06
Glomerular Filtration	1.788733E-01	1.356826E-02
Renal Filtration	3.135022E-01	1.564006E-02
Neutrophil Chemotaxis	3.708513E-01	7.871436E-07
Granulocyte Chemotaxis	1.159051E-01	9.324995E-07
Neutrophil Migration	2.659021E-01	1.156252E-06
Inflammatory Response	3.889616E-01	3.823976E-06

Figure S11. KEGG and GO:BP Pathway Enrichment Analysis of Differentially Expressed Genes in Tumor Regions Compared to Other Tissues, Based on Gene Expression Predicted by SpaHGC on the Alex Dataset. (**A**) Workflow diagram for identifying differentially expressed genes and conducting enrichment analysis based on gene expression predictions made by SpaHGC. (**B**) and (**C**) Display the P-values of significantly enriched KEGG pathways identified by SpaHGC compared to the raw dataset. (**D**) and (**E**) Show the P-values of significantly enriched GO:BP pathways identified by SpaHGC compared to the raw data.

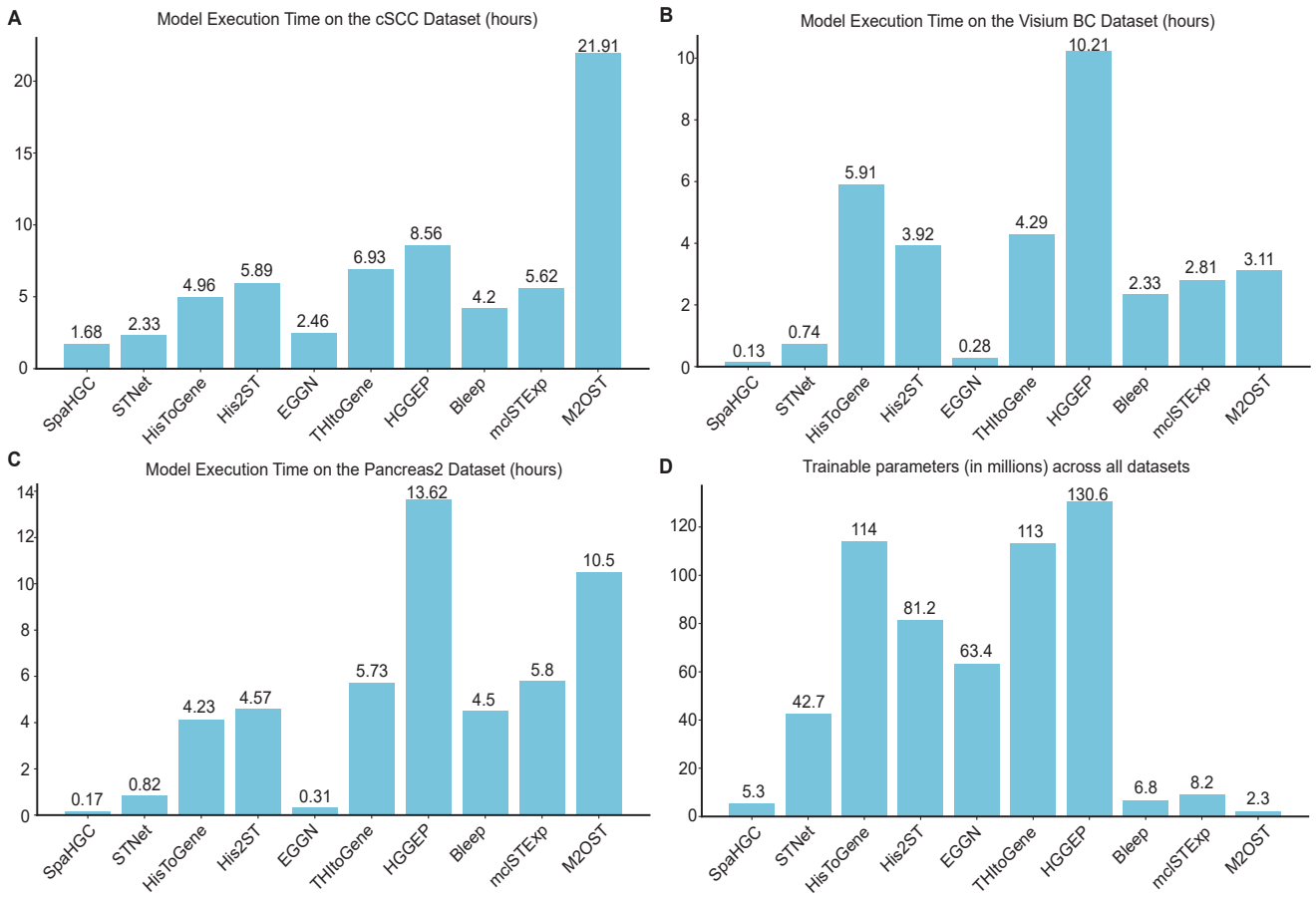


Figure S12. Comparison of computational time and trainable parameters between our proposed method and baseline approaches on the same datasets. (**A**) Runtime on the cSCC dataset (in hours); (**B**) Runtime on the Pancreas2 dataset (in hours); (**C**) Inference time on the Pancreas2 dataset (in hours); (**D**) Number of trainable parameters (in millions) across all datasets.