

## Supplementary Materials for

# Inferring single-cell resolution spatial gene expression with histological images using graph convolutional network

## Supplementary Notes

### 1. Baseline methods

We selected several state-of-the-art methods in super-resolution task that are representative of the following:

- **iStar** [1] models super-resolution gene expression from hierarchical histological features using a feedforward neural network. This method is divided into two parts: the HIPT model [2] is used to extract hierarchical histological features, and then a feedforward neural network is used to predict super-resolution gene expression.
- **XFuse** [3] integrates Spatial transcriptomics (ST) data and histology images using a deep generative model to infer super-resolution gene expression profiles. This method considers spatial gene expression and histological image data as observable effects of potential tissue states, and maps image data to potential states through a recognition neural network.
- **TESLA** [4] generates high-resolution gene expression profiles based on Euclidean distance metric, which considers the similarity in physical locations and histological image features between superpixels and measured spots.
- **STAGE** [5] to generate gene expression data for unmeasured spots or points from Spatial Transcriptomics with a spatial location-supervised Auto-encoder GEnetor by integrating spatial information and gene expression data. STAGE was originally designed to predict gene expression at spot gaps, but we were also able to obtain super-resolution gene expression using STAGE by converting spatial coordinates from the spot level to the superpixel level.

We did not compare scstGCN with BayesSpace [6] and HisToGene [7] because BayesSpace separates a spot into several sub-spots and cannot impute gene expression for unmeasured locations; HisToGene merely interpolated the gaps between spots and did not enhance the gene expression levels to single-cell resolution.

### 2. Implementation Details

For all baselines, default parameters from the original papers were used, and all experiments were executed on a single NVIDIA RTX 3090 GPU using Python 3.11.5 and Pytorch (version 2.1.1). The training protocol was established for 500 epochs and a learning rate set at 0.0001. The batch size is set to the minimum value between 128 and the integer division of number of spot by 16. In the multimodal feature mapping extractor of scstGCN, the ViT architecture utilizes a self-pretrained model, UNI [8], which uses DINov2 to pretrain on over 100 million images from diagnostic H&E-stained WSIs spanning 20 major tissue types. In our GCN module, the number of nodes in each subgraph is determined based on specific features of the dataset. For example, Visium and ST data have spot sizes of 55  $\mu\text{m}$  and 200  $\mu\text{m}$ , respectively, while the size of a single cell is approximately 8  $\mu\text{m}$ . Therefore, the number of nodes in the subgraph is set to 7 $\times$ 7 and 29 $\times$ 29, which exactly covers the entire spot.

### 3. Evaluation metrics

In this study, we employ the Root Mean Square Error (RMSE) and Structural Similarity Index (SSIM) to compare the predicted super-resolution gene expression with the observed ground truth from Xenium datasets. To calculate the RMSE, the ground truth and the predicted gene expression profiles were flattened into vectors, with the RMSE was equal to the Euclidean distance between the two vectors. The calculation formula of RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (u_i - v_i)^2}, \quad (1)$$

where  $u_i$  and  $v_i$  denote the flattened super-resolution gene expression obtained by prediction and the flattened ground truth, respectively.  $n$  denotes the length of flattened vectors. RMSE is a direct and quick metric for assessing the accuracy of predictions for any vectorizable outcome. However, for image data, RMSE ignores spatial context within the image. Therefore, we also calculate SSIM, a measure of image similarity widely used in tasks such as super-resolution and medical imaging. A higher SSIM indicates greater similarity between two images. The SSIM can be expressed as follows:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\text{conv}(x, y) + 2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2)$$

where  $x$  and  $y$  represent the ground truth and the predicted super-resolution gene expression, respectively.  $\mu_x$ ,  $\mu_y$  denote the mean of  $x$ ,  $y$ .  $\sigma_x$ ,  $\sigma_y$  denote the standard deviation of  $x$ ,  $y$ .  $\text{conv}$  denotes covariance.  $C_1$  and  $C_2$  are set to 0.01 and 0.03, respectively.

When evaluating metrics on Visium or from other platforms data without super-resolution gene expression as a label, after a post-processing step, we used the RMSE, Pearson correlation coefficient (PCC), and Mean Absolute Error (MAE) to evaluate the performance of super-resolution gene expression predictions. The PCC has a range of values  $[-1, 1]$ . It measures the strength of the linear relationship between two variables by dividing the covariance of the two variables by the product of their respective

standard deviations:

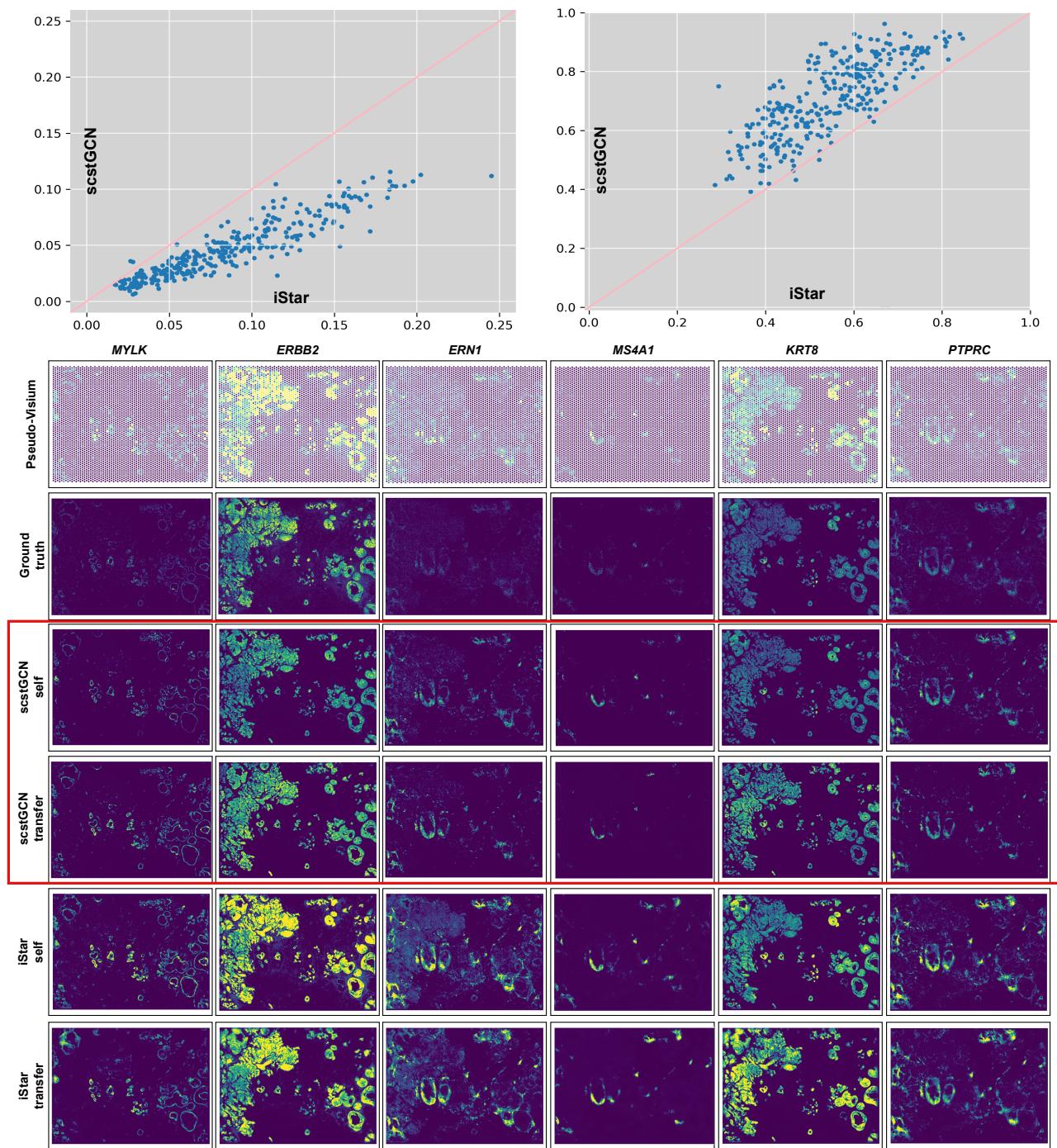
$$\mathbf{PCC} = \frac{\mathbf{conv}(x, y)}{\sigma(x)\sigma(y)}. \quad (3)$$

MAE represents the simple average of the absolute errors between predicted and observed values, exhibits lower sensitivity to outliers, and equally considers the magnitude of all errors. Similar to RMSE, the observed values and the predicted values were flattened into vectors before calculating the MAE. The MAE can be expressed as follows:

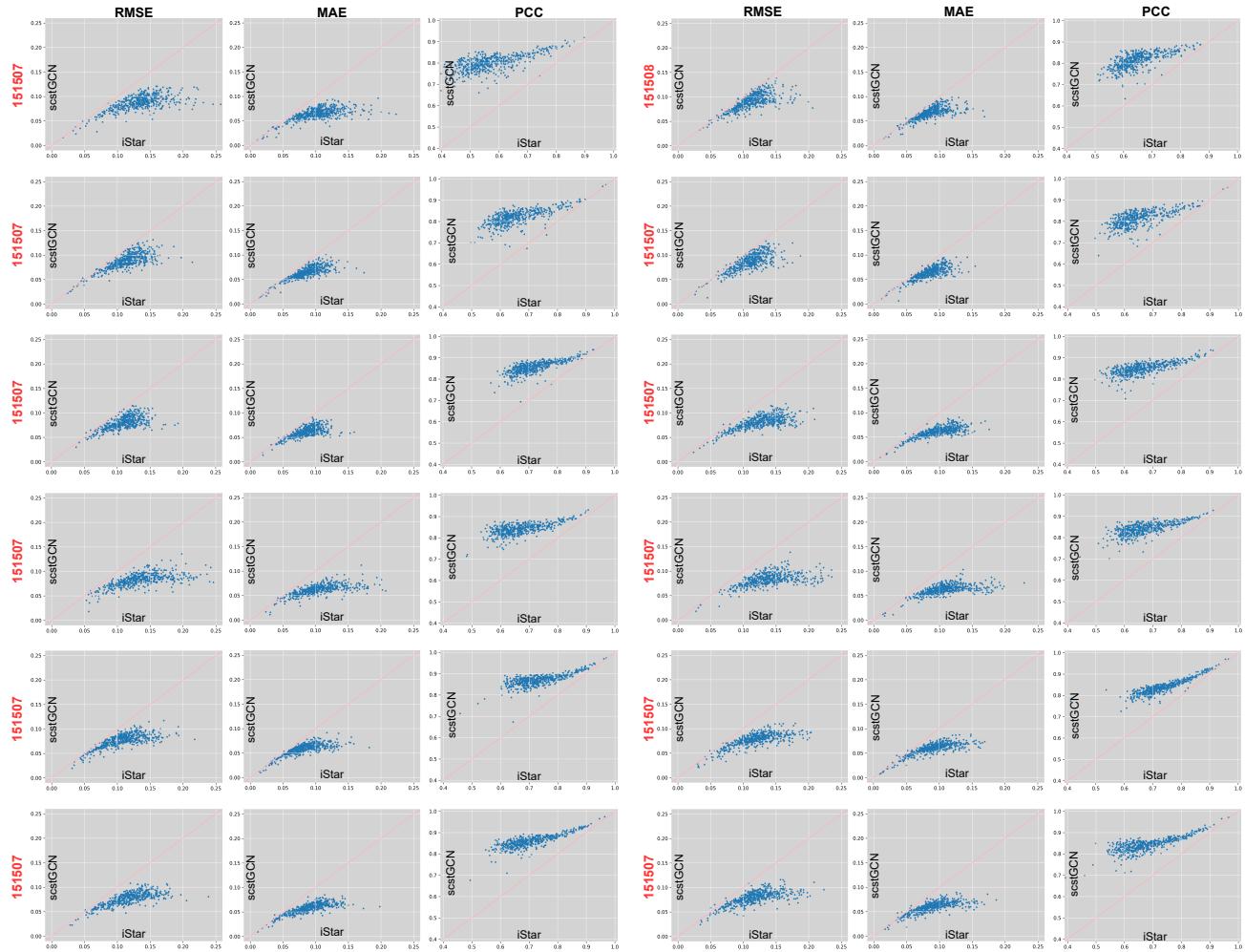
$$\mathbf{MAE} = \frac{1}{n} \sum_{i=1}^n |u_i - v_i|. \quad (4)$$

In super-resolution tasks, PCC is not a common metric because it is sensitive to abnormal noise anomalies. PCC is a common evaluation criterion in spot-based ST data, as the number of spots in low resolution spot-based ST data typically ranges from hundreds to thousands. In super-resolution tasks, the number of superpixels can reach hundreds of thousands or even millions, PCC is not suitable in high noise magnitude in the super-resolution gene expression. Therefore, in the experimental evaluation of post-based ST data such as DLPFC, we used PCC as the Evaluation metric, whereas in Xenium data, we did not use PCC to measure performance.

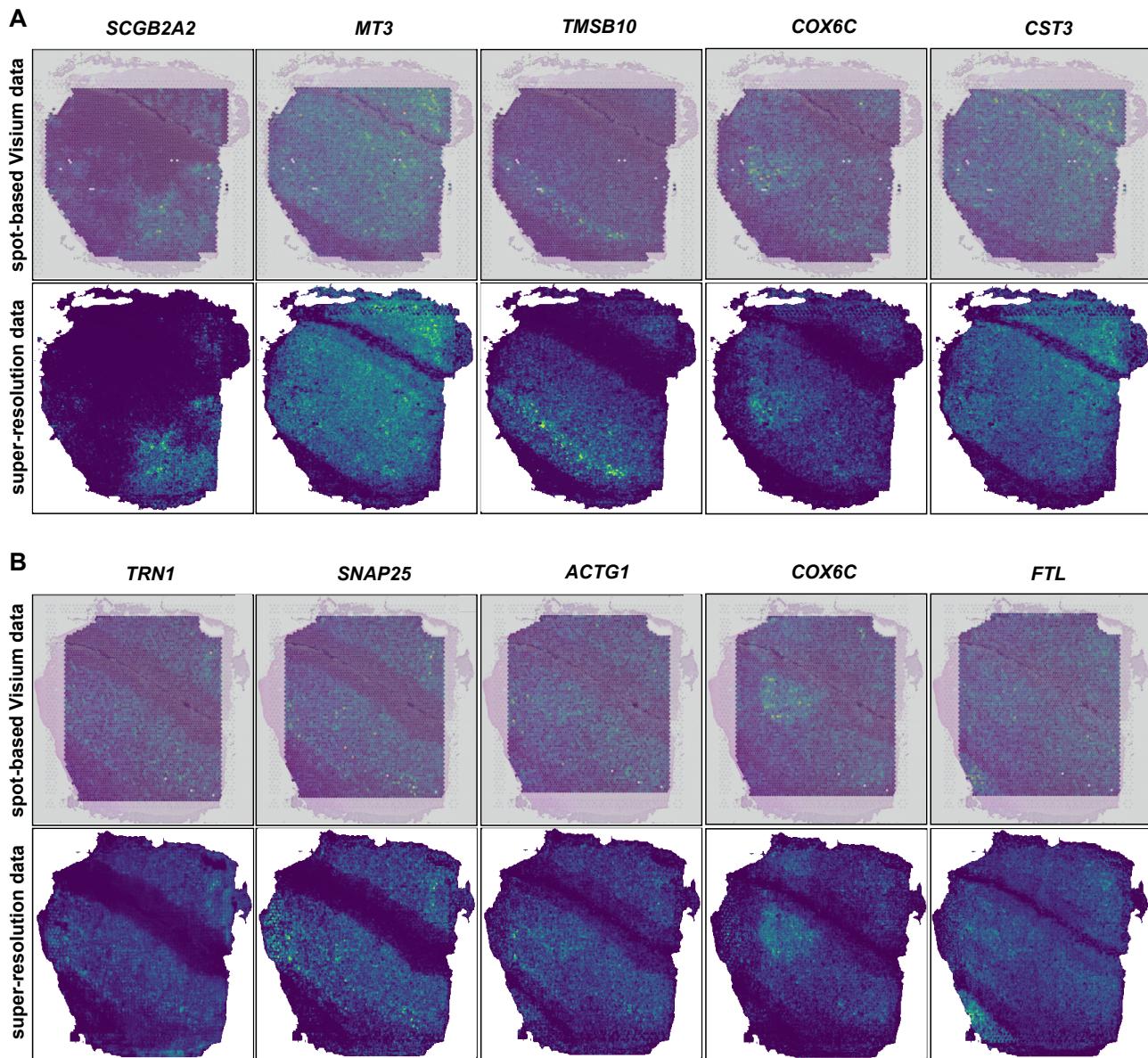
## Supplementary Figures



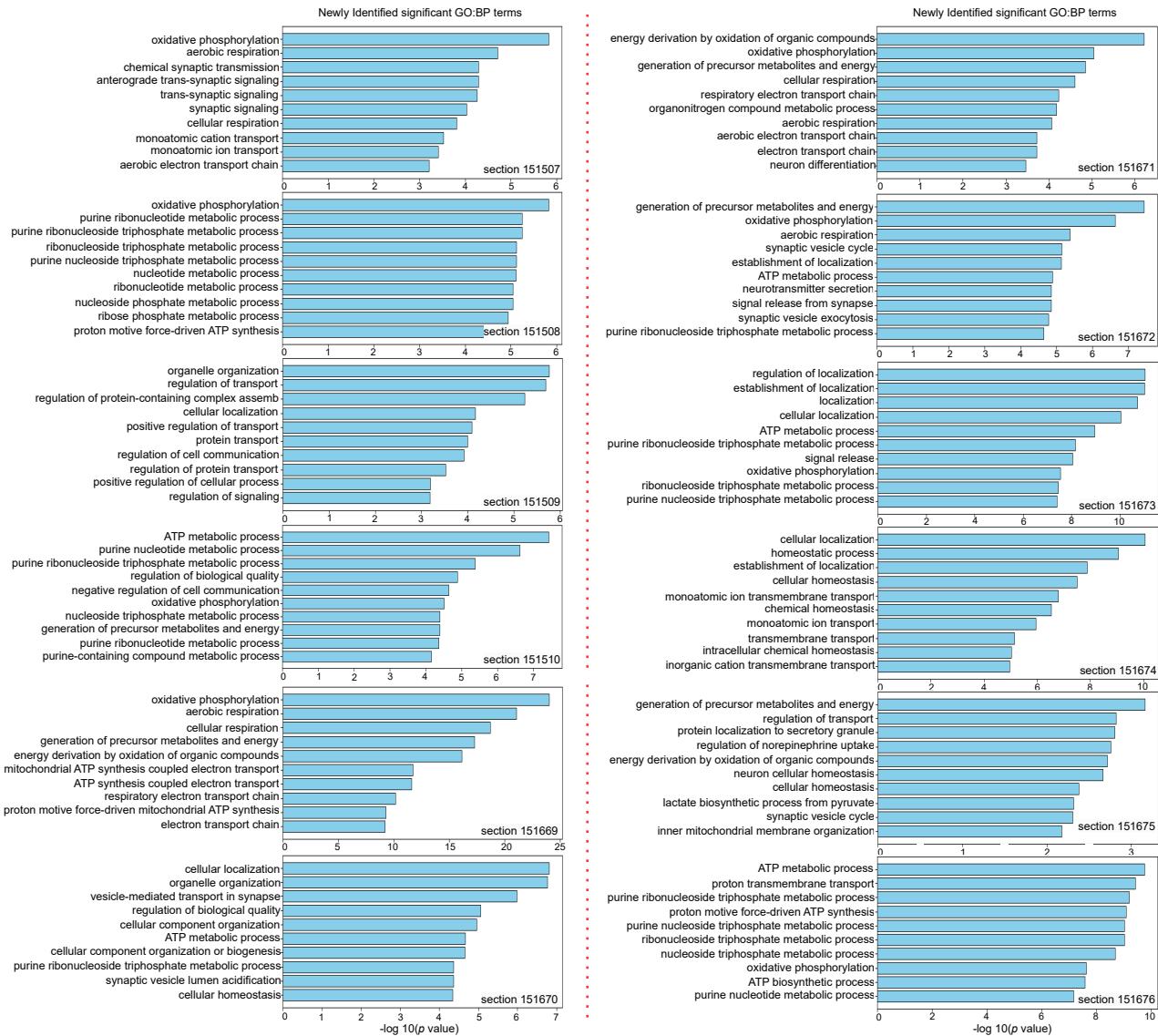
**Figure S1.** Numerical evaluation and spatial gene visualization results of scstGCN and state-of-the-art method iStar in transfer learning. For transfer learning, the pseudo-Visium data of HBC\_S1R2 from human breast cancer (HBC) Xenium data were used as the training data, and super-resolution gene expression profiles was obtained on HBC\_S1R1 using only its histological image as the input. **(A)** Scatter plots of Pearson correlations between the ground truth and the super-resolution gene expression predicted by the scstGCN and iStar for the all 313 genes. Each dot represents one of the 313 genes. **(B)** Spatial expression analysis of multiple groups of genes with different spatial patterns in HBC\_S1R1 data. The analysis are based on pseudo-Visium data, super-resolution ground truth, and predicted data using scstGCN and iStar by weakly-supervised learning and transfer learning, respectively. Each column corresponds to a gene, with the first two rows from the top displays the pseudo-Visium and ground truth, while the subsequent rows show the predicted data using different methods.



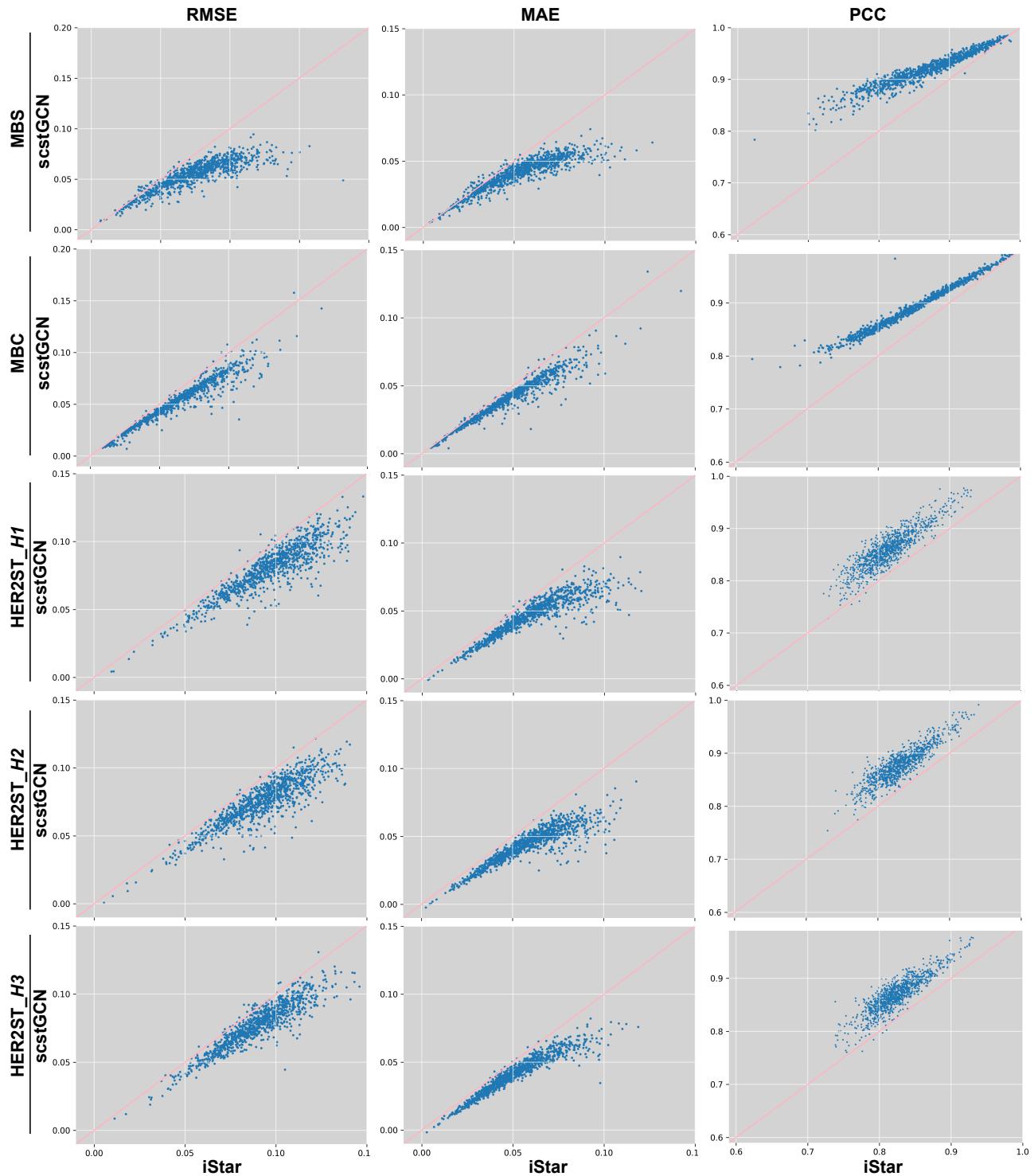
**Figure S2.** Numerical evaluation of prediction accuracy between the original spot-level gene expression and “spot-level” gene expression obtained from the enhanced expression generated by scstGCN and state-of-the-art method iStar. The prediction accuracy of scstGCN and iStar was measured using root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Pearson correlation coefficient (PCC) measurements for 1000 highly variable genes in each of all 12 sections from the human dorsolateral prefrontal cortex (DLPFC) tissue data. In each scatter plot, a dot represents a gene. The results show that scstGCN achieved better performance in all sections.



**Figure S3.** scstGCN can not only improve the resolution within the measured spot, but also predict single-cell resolution gene expression in non measured spot areas of histological images. As long as the model is trained at all measured spots, the super-resolution gene expression of all tissue regions can be predicted based solely on histological imag. (A) and (B) represent the spatial visualization of several genes having different spatial patterns for the spot-based Visium data and predicted super-resolution data by scstGCN in section 151507 and 151510, respectively.



**Figure S4.** Top-10 significant GO:BP terms were only identified in the super-resolution data predicted by scstGCN in all sections from DLPFC tissue.



**Figure S5.** Numerical evaluation of predictive performance of scstGCN and state-of-the-art method iStar on the human HER2 positive breast cancer (HER2ST), mouse brain sagittal cut (MBS) and mouse brain coronal cut (MBC) datasets. For HER2ST data, we considered three consecutively cut sections from sample  $H$ . Each row corresponds to a data, and each column corresponds to one of the evaluation metrics of RMSE, MAE, and PCC. In each scatter plot, a dot represents one of the 1000 highly variable genes. The results show that scstGCN outperforms the state-of-the-art method iStar across all datasets.

## Supplementary Tables

**Table S1.** Summary of ST datasets used in this study.

Datasets	Species	Tissue	Size	Sections	Spots/Cells	Genes	Platform	Data source
HBC	Human	Breast cancer	single-cell	3	167,780 cells (Section 1) 118,752 cells (Section 2) 142,272 cells (Section 3)	313 genes (Section 1,2) 288 genes (Sample 3)	Xenium	<a href="https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast">https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast</a>
HP	Human	Pancreas	single-cell	1	140,702 cells	377 genes	Xenium	<a href="https://www.10xgenomics.com/datasets/ffpe-human-pancreas-with-xenium-multimodal-cell-segmentation-1-standard">https://www.10xgenomics.com/datasets/ffpe-human-pancreas-with-xenium-multimodal-cell-segmentation-1-standard</a>
HN	Human	Heart	single-cell	1	26,366 cells	377 genes	Xenium	<a href="https://www.10xgenomics.com/datasets/human-heart-data-xenium-human-multi-tissue-and-cancer-panel-1-standard">https://www.10xgenomics.com/datasets/human-heart-data-xenium-human-multi-tissue-and-cancer-panel-1-standard</a>
DLPFC	Human	Cortex	55µm	12	3,460-4,789 spots	33,538 genes	Visium	<a href="http://spatial.libd.org/spatialLIBD">http://spatial.libd.org/spatialLIBD</a>
HER2ST	Human	breast cancer	100µm	32	613 spots (Section H1) 441 spots (Section G1)	15,029 genes (Section H1) 14,992 genes (Section G1)	ST	<a href="https://github.com/almaan/her2st">https://github.com/almaan/her2st</a>
MBC	Mouse	Brain (coronal)	55µm	1	2,235 spots	19,465 genes	Visium	<a href="https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-2-ffpe-2-standard">https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-2-ffpe-2-standard</a>
MBS	Mouse	Brain (sagittal)	55µm	1	3,289 spots	32,285 genes	Visium	<a href="https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard">https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard</a>

**Table S2.** TLS (Tertiary lymphoid structure) marker genes from [1].

TLS marker genes	Celltype labels
CD4	Endothelial
CD8A	T cells
CD74	Myeloid
CD79A	Plasmablasts
IL7R	T cells
ITGAE	N/A
CD1D	N/A
CD3D	T cells
CD3E	T cells
CD8B	T cells
CD19	B cells
CD22	B cells
CD52	T cells
CD79B	B cells
CR2	Myeloid
CXCL13	N/A
CXCR5	B cells
FCER2	B cells
MS4A1	B cells
PDCD1	T cells
PTGDS	CAFs
TRBC2	T cells

## References

1. Daiwei Zhang, Amelia Schroeder, et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, pages 1–6, 2024.
2. Richard J Chen, Chengkuan Chen, et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16144–16155, 2022.
3. Ludvig Bergensträhle, Bryan He, and Bothers. Super-resolved spatial transcriptomics by deep data fusion. *Nature Biotechnology*, 40(4):476–479, 2022.
4. Jian Hu, Kyle Coleman, et al. Deciphering tumor ecosystems at super resolution from spatial transcriptomics with tesla. *Cell Systems*, 14(5):404–417, 2023.
5. Shang Li, Kuo Gai, Kangning Dong, Yiyang Zhang, and Shihua Zhang. High-density generation of spatial transcriptomics with stage. *Nucleic Acids Research*, 52(9):4843–4856, 2024.
6. Edward Zhao, Matthew R Stone, et al. Spatial transcriptomics at subspot resolution with bayesspace. *Nature Biotechnology*, 39(11):1375–1384, 2021.
7. 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 2021–11, 2021.
8. Richard J Chen, Tong Ding, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.