Learning Spatially-Aware Representations of Transcriptomic Data via Transfer Learning

Minsheng Hao
Department of Automation, BNRIST
Tsinghua University
hms20@mails.tsinghua.edu.cn

Lei Wei
BNRIST
Tsinghua University
weilei92@tsinghua.edu.cn

Xuegong Zhang
Department of Automation, BNRIST,
School of Life Sciences
Tsinghua University
zhangxg@tsinghua.edu.cn

Abstract

Computationally integrating spatial transcriptomics (ST) and single-cell transcriptomics (SC) greatly benefits biomedical research such as cellular organization, embryogenesis and tumorigenesis, and could further facilitate therapeutic developments. We proposed a transfer learning model, STEM, to learn spatially-aware embeddings from gene expression for both ST and SC data. The embeddings satisfy both the preservation of spatial information and the elimination of the domain gap between SC and ST data. We used these embeddings to infer the SC-ST mapping and the pseudo SC spatial adjacency, and adopted the attribution function to indicate which genes dominate the spatial information. We designed a comprehensive evaluation pipeline and conducted two simulation experiments, and STEM achieved the best performance compared with previous methods. We applied STEM to human squamous cell carcinoma data and successfully uncovered the spatial localization of rare cell types. STEM is a powerful tool for building single-cell level spatial landscapes and could provide mechanistic insights of heterogeneity and microenvironments in tissues.

1 Introduction

High-resolution single-cell gene expression data with spatial information is critical for revealing the mechanisms of cellular organization, embryogenesis and tumorigenesis [1–5], and could further facilitate therapeutic developments [6, 7]. Recently, many spatial transcriptomic (ST) profiling protocols have been developed [8–11]. The most commonly-used ST protocols aggregate multiple cells into one spot, and the provided data contain both spatial coordinates and in-situ gene expressions with limited resolution and gene coverage [12]. On the contrary, single-cell RNA sequencing (SC) captures exact single-cell level transcripts with a higher throughput of gene species [13] but cannot detect any spatial information. Computationally integrating SC and ST data can build an informative single-cell level spatial landscape and benefits studies of both sides [14]. For example, inferring the pseudo spatial adjacency association of SC data helps identify spatial structures of cell types or niches [15, 16], and deconvoluting cell type proportions of spots in ST data by using annotated SC data can benefit the illustration of cell spatial niches and communication landscapes in microenvironments [17–20].

Preprint. Under review.

The core of building a single-cell level spatial landscape is to establish SC-ST and SC-SC spatial associations. The computational challenge is to recover the spatial information of SC data, which requires us to learn the relation between gene expression profiles and locations from ST data and transfer it to SC data for spatial information inference.

In this study, we proposed a deep transfer learning model, STEM, to learn SpaTially-aware EMbeddings of ST and SC data. We designed a shared encoder for SC and ST data to get unified embeddings in one latent space, and optimized the embeddings to satisfy both the preservation of spatial information and the elimination of the domain gap between SC and ST data. We further used the embeddings to infer the SC-ST mapping and the pseudo SC spatial adjacency.

Compared with existing methods [15, 16, 21–23], STEM overcomes the domain gap between SC and ST data and explicitly extracts the spatial information from gene expression as spatially-aware embeddings. Based on the embeddings, STEM provides a flexible and accurate way to infer spatial associations. Moreover, STEM identifies spatially dominant genes (SDGs) that highly dominate the inferred spatial location of a cell, which could benefit the understanding of underlying mechanisms related to cellular spatial organization or communication.

We designed a comprehensive pipeline at both single-cell and tissue levels for evaluation, including examining the correctness of the predicted SC-ST mapping matrix and the inferred SC spatial adjacency matrix. We conducted two simulation experiments, and the results shown the superiority of STEM in inferring spatial associations. We further used STEM to identify SDGs in mouse gastrulation data [24] and locate enriched regions of rare cell types in human squamous cell carcinoma [19]. These experiments highlight the power of representation learning for extracting spatial information from gene expression, and demonstrate the feasibility of transfer learning models for integrating data in different modalities and constructing cross-domain associations.

2 Task Formulation

ST data include two types of information: a gene expression matrix $X^{ST} \in \mathbb{R}^{N \times K}$ and a spatial coordinate matrix $Y^{ST} \in \mathbb{R}^{N \times 2}$, where N is the number of sequenced spots and K is the number of genes. SC data only have the gene expression matrix $X^{SC} \in \mathbb{R}^{M \times K}$, where M is the number of sequenced single cells. As SC and ST data are obtained by different techniques, distributions of gene expressions within X^{ST} and X^{SC} such as the sparsity and maximum expression values are different.

Three goals are needed for establishing SC and ST spatial associations. The first goal is to infer the spatial information of SC data. It can be formulated as a SC adjacency matrix $S_c \in \mathbb{R}^{M \times M}$ or the estimated spatial coordinate $\tilde{Y}_{SC} \in \mathbb{R}^{M \times 2}$. The second goal is to learn the spatial associations between SC and ST data, and it can be formulated as a SC-ST mapping matrix $C \in \mathbb{R}^{M \times N}$. The third goal is to know the contribution of genes to determining the spatial location of one single cell, i.e. a gene attribution vector. The attribution vector W of a single cell i is defined as

$$W_i = (a_1, a_2, ..., a_K) := \mathcal{F}(x_i^{SC}, C_i) \in \mathbb{R}^K \tag{1}$$

where $\mathcal{F}$ is an attribution function, the vector C_i is the corresponding row in the mapping matrix C, and a_j is the contribution of gene j to the mapping result C_i. The vector W can be further used to identify SDGs.

3 Methodology

STEM is in an encoder-predictor fashion (Fig. 1). The encoder represents SC and ST gene expression vectors as embeddings in a unified latent space. The embeddings are simultaneously optimized by two modules of predictor: the spatial information extracting module and the domain alignment module. These two modules predict the same ST spatial adjacency in different ways. The spatial information extracting module only uses the ST embeddings and guarantees the learned embeddings contain spatial information extracted from the gene expression. The domain alignment module uses SC and ST embeddings and eliminates the SC-ST domain gap by first minimizing the Maximum Mean Discrepancy (MMD) [25] of SC and ST embeddings and then constructing ST-SC-ST spatial associations as ST adjacency to find the optimal mapping matrix C.

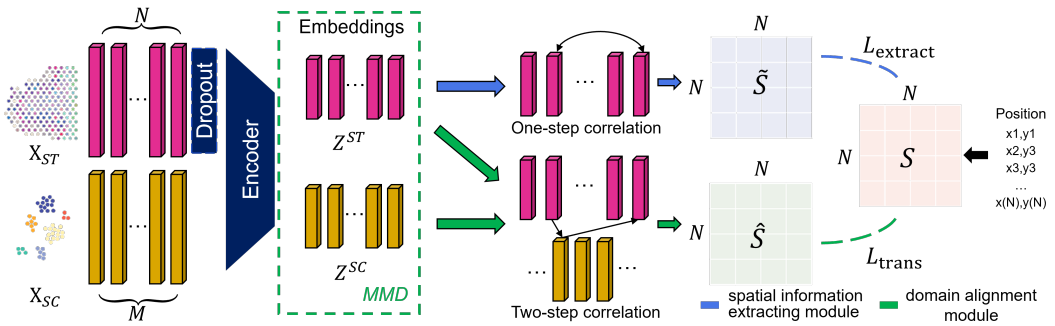

Figure 1: The schematic overview of our model.

After training, the embeddings of SC and ST are used to infer SC-ST mapping matrix and SC-SC spatial adjacency matrix. The SDGs are found by applying the integrated gradient [26] technique as an attribution function upon the trained model.

Encoder We input SC and ST gene expression vectors (rows in X^{SC} or X^{ST}) into one encoder. As one spot in ST data contains more than one cell in SC data, the number of expressed genes in one spot is usually more than that in one cell. To make our model robust to data sparsity, We add an additional dropout layer [27] for ST data before the encoder. The dropout rate d is defined as follows:

$$d = 1 - \frac{\text{Median}(n_{SC})}{\text{Median}(n_{ST})} \tag{2}$$

where n_{SC} and n_{ST} represent the number of expressed gene species in each cell or spot, respectively. We use a MLP encoder to embed the expression vectors into latent embeddings $Z^{ST} \in \mathbb{R}^h$ and $Z^{SC} \in \mathbb{R}^h$ with the same dimension size h.

Predictor part 1: spatial information extracting module We use the ST embeddings Z^{ST} to construct the ST spatial adjacency matrix $\tilde{S}$ and calculate the loss between it and the ground truth ST spatial adjacency matrix S. The ground truth matrix S is created from the spatial coordinate labels via a spatial kernel function ϕ which could be Gaussian, Mahalanobis or others. Here we use the Gaussian kernel (see Appendix 1 for details). We apply the L1 normalization over columns in S to let the sum of adjacency values of each spot be 1:

$$S_{ij} := \frac{\phi(Y_i^{ST}, Y_j^{ST})}{\sum_k S_{ik}} \tag{3}$$

Meanwhile, the predicted $\tilde{S}$ matrix is measured by the inner product of two ST embedding vectors with softmaxing over columns:

$$\tilde{S}_{ij} := \text{softmax}_{col} < Z_i^{ST}, Z_j^{ST} >= \frac{\exp(< Z_i^{ST}, Z_j^{ST} >)}{\sum_k \exp(< Z_i^{ST}, Z_k^{ST} >)}. \tag{4}$$

We use the cross entropy H to calculate the loss between $\tilde{S}$ and S:

$$L_{\text{extract}} = H(\tilde{S}, S) \tag{5}$$

Predictor part 2: domain alignment module We reduce the mean distance between ST and SC embeddings, and use these embeddings to estimate the SC-ST and ST-SC mapping matrices. Then we exploit two mapping matrices to construct another two-step ST spatial adjacency matrix $\hat{S}$. As $\hat{S}$ depends on the SC-ST cross-domain spatial relationship, we can indirectly find the optimal cross-domain relationship by directly optimizing $\hat{S}$. The similar idea is proposed in the Haeusser's work [28, 29]. In our work, we extend its applicability from classification to relation construction and fully utilize the cross domain association matrix as the SC-ST mapping matrix.

We introduce the MMD loss on SC embeddings Z^{SC} and ST embeddings Z^{ST} in a mini-batch to reduce the mean distance between them (Appendix 2):

$$L_{\mathrm{MMD}} = \mathrm{MMD}(Z^{SC}, Z^{ST}) \tag{6}$$

We get the similarity between ST and SC embeddings as $B_{ij} :=< Z_i^{SC}, Z_j^{ST} >$. We define the mapping matrix $C \in \mathbb{R}^{M \times N}$ from SC to ST by softmaxing B over columns:

$$C_{ij} := \mathrm{softmax}_{col}\, B_{ij} = \frac{\exp(B_{ij})}{\sum_k \exp(B_{ik})}. \tag{7}$$

We also define the mapping matrix from ST to SC $\hat{C} \in \mathbb{R}^{N \times M}$ by replacing B with B^T. We can get the two-step spatial adjacency matrix $\hat{S} \in \mathbb{R}^{N \times N}$ from ST to ST as:

$$\hat{S} = \hat{C} \cdot C \qquad where \quad \hat{S}_{ij} := (\hat{C} \cdot C)_{ij} = \sum_k \hat{C}_{ik} C_{kj} \tag{8}$$

We calculate the cross entropy loss between $\hat{S}$ and the ground truth spatial adjacency matrix S:

$$L_{\mathrm{trans}} = H(\hat{S}, S) \tag{9}$$

The total loss of our model consists of three parts (Equation 5, 6 and 9):

$$L = L_{\mathrm{extract}} + \alpha L_{\mathrm{MMD}} + \beta L_{\mathrm{trans}} \tag{10}$$

We set the hyper parameter $\alpha = 0.2$ and gradually increase β from 0 to 1 during training, which enforces the model first to rebuild the spatial adjacency in ST and then learn the spatial mapping between ST and SC.

Attribution function The integrated gradient (IG) requires three parts for computing attributions: the input and output of the deep learning model, and a baseline input (Appendix 3). For cell i in SC data, we set the input as the corresponding SC gene expression vector X_i^{SC}, the baseline as a zero vector, and the output as the maximum value in the corresponding row in the mapping matrix $C_{i\cdot}$:

$$W_i = \mathrm{IG}(X_i^{SC}, \mathbf{0}, \max C_{i\cdot}) \tag{11}$$

As the maximum value of $C_{i\cdot}$ indicates that this cell has the maximum probability located around some ST spot, the attribution vector shows the contribution of genes for determining this spatial location. We use the attribution function to get the gene attribution vector of each cell in SC data and group all attribution vectors by cell annotation (e.g. cell type or spatial regions). For a cell group of interest, we perform the one-versus-others Wilcoxon test [25] and define genes that have significantly higher values in that group as SDGs.

4 Experimental design

4.1 Evaluation pipeline

We design a comprehensive evaluation pipeline which includes three processes: simulation data generation, output unification, and performance evaluation (Fig. 2).

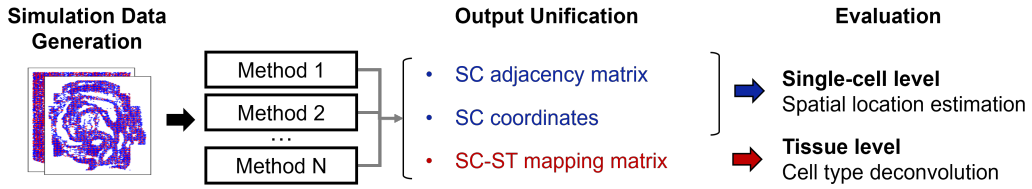

Figure 2: The evaluation pipeline.

Simulation data generation Transcriptomic data with spatial information at the single-cell resolution is required for evaluating the methods' ability of inferring spatial associations. Currently, such data can only be provided by some less popular spatial sequencing technologies [8, 30–32] which are low-throughput and require complicated operations. We take these experimental single-cell resolution spatial transcriptomic data as SC data with the ground-truth spatial information, and simulate pseudo ST data by creating a spatial grid on the spatial space of these data. The detailed simulation process can be found in Appendix 4. Each spot in the simulated pseudo ST data has the information of gene expression, spatial coordinate and cell type proportion.

Output unification To make the results comparable, we unify the outputs of different methods into three parts: the SC spatial adjacency matrix, the reconstructed SC spatial coordinates and the SC-ST mapping matrix. For methods that predict SC coordinates, the SC-ST mapping matrix and SC adjacency matrix are obtained by calculating the spatial distance between or within single cells and spots. The closer the spatial distance between the cells or spots, the higher the weight in the matrix. For other methods that provide the SC-ST mapping matrix, the SC coordinates are obtained by averaging the coordinates of spots according to the mapping weights.

Performance evaluation For the first and second goals, we validate the model performance at the single-cell level and tissue level, respectively. For the third goal, it is worth emphasizing that our model STEM is the only one that could achieve it, and thus we validate the detected SDGs by known knowledge. At the single-cell level, we focus on whether the spatial association within single cells can be inferred. We use the mean squared error (MSE) as the error metric between the predicted and true spatial coordinates. As the number of reference spots (K_{ref}) may affect the predicted coordinate results, we calculate MSEs under different K_{ref}. To measure the models' ability of constructing the spatial relationship among cells, we use the SC spatial adjacency matrix to calculate the hit number which is defined as the number of one cell's K-nearest neighbors that are successfully predicted. At the tissue level, we focus on whether the cell type information can be correctly mapped from single cells to spots through the SC-ST mapping matrix. We calculate the Pearson correlation coefficient (PCC) between the predicted and true cell type spatial distribution.

4.2 Simulation experiments

We conducted simulation experiments on two datasets with the evaluation pipeline. The first dataset is single-cell resolution mouse gastrulation data generated by seqFISH [30]. The second dataset is mouse hippocampus data generated by Slide-seq V2 [8] with a spatial resolution of 10 μm which is comparable to the size of a cell. Both of the datasets consist of gene expression profiles, spatial coordinates and cell type information, and we treated them as SC data with true spatial coordinates.

We compared STEM with five methods, including Seurat [21], Tangram [22], SpaOTsc [23], Celltrek [16] and scSpace [15]. Seurat integrates SC and ST data by performing canonical correlation analysis on gene expression data. Tangram optimizes the mapping matrix between SC and ST data by minimizing the cosine distance between predicted and true gene expression. SpaOTsc introduces structured and unbalanced optimal transport to integrate SC and ST data. Celltrek builds a graph containing both SC and ST nodes based on the random forest distance and estimates the spatial coordinates of SC data. scSpace uses transfer component analysis to directly estimate the spatial coordinates. The original outputs of these approaches are summarized in Table 1. We ran all these methods with default parameters and unified the outputs. The embedding-based calculation makes STEM the only one that satisfies both goal without any additional result transformation.

4.3 Real data application

We applied STEM to human squamous cell carcinoma (hSCC) data. The data are obtained from Ji's study [19] and contain paired SC and ST data from the same donor. The true spatial association between SC and ST data is unknown. We trained a model for each donor. To explore the results, We first verified whether STEM can reproduce the conclusion reported in the original study, and then used STEM to gain new biological insights.

Table 1: Original outputs of different approaches

	SC-ST mapping	SC-SC association	SC coordinate	SDG detection
Celltrek	✗	✗	✓	✗
scSpace	✗	✗	✓	✗
Seurat	✓	✗	✗	✗
SpaOTsc	✓	✗	✗	✗
Tangram	✓	✗	✗	✗
Ours	✓	✓	✗	✓

4.4 Implementation details

STEM is based on the publicly available Pytorch [33] platform. The encoder is a three-layer MLP with LayerNorm [34] regularization and ReLU activation. The hidden size is 512, 256 and 128, respectively. The dropout rate of the Dropout Layer is 0.6, 0.8 and 0.2 for gastrulation, hippocampus and hSCC data, respectively. During training, the AdamW [35] optimizer is applied with default parameters. We took all SC and ST data in one training step, and sampled 2,000 SC and ST data for MMD Loss. We didn't observe any performance improvement by setting a larger sample size. We used the Captum [36] package to implement the integrated gradient technique. With one NVIDIA GeForce RTX 3090 GPU, the training process takes around 5-10 minutes.

5 Results

5.1 Results on simulation data

5.1.1 At the single-cell level

Compared with other methods, STEM achieved the lowest MSE for predicting the absolute spatial coordinates (Fig. A2). Moreover, we found that STEM achieved the lowest MSE in the majority of cell groups with different cell type annotations (Fig. A3 and A4), demonstrating that cell type information does not affect the model's ability of extracting spatial information from expression values. Under different numbers of reference spots K_{ref} (Table 2 and 3 right), STEM also achieved the lowest MSE. We observed that the MSE of STEM was decreased and converged as the number of reference spots increases. The error of Tangram and Seurat was stable, and the error of SpaOTsc increased as the reference number increased.

We then assessed the methods' ability of constructing the spatial relationship among single cells. on the gastrulation data, STEM got the highest hit number, about two times as many as the second-place method (Fig. 3a). When the true neighbor number $K_{\mathrm{nbr}} = 200$, STEM found more than 80 correct neighbors. On the hippocampus data, STEM had a comparable performance with Tangram when $K_{\mathrm{nbr}} \leq 100$, and had a higher improvement as K_{nbr} increased. And other methods showed relatively low hit numbers. These results show that the spatial association inferred by STEM is more consistent with the actual spatial distribution for single-cell data.

Table 2: PCC and MSE on the gastrulation data

	Tissue-level PCC						Cell-level MSE			
	Brain	Card.	Endothelium	Gut tube	Neur.	Spin.	$K_{\mathrm{ref}} = 10$	$K_{\mathrm{ref}} = 30$	$K_{\mathrm{ref}} = 50$	ALL
Celltrek	0.896	0.931	0.682	0.881	0.821	0.820	\	\	\	24.193
scSpace	0.773	0.767	0.091	0.458	0.462	0.670	\	\	\	37.684
Seurat	0.519	0.767	0.680	0.708	0.546	0.478	28.416	30.066	30.074	30.074
SpaOTsc	0.892	0.849	0.822	0.828	0.706	0.824	21.341	29.379	35.151	59.846
Tangram	0.871	0.942	**0.885**	0.918	**0.921**	0.829	29.274	29.270	29.270	29.275
Ours	**0.969**	**0.972**	0.793	**0.954**	0.917	**0.936**	**16.146**	**16.125**	**16.125**	**16.125**

*Card. = Cardiomyocytes, Neur. = Neural crest, Spin. = Spinal cord

Table 3: PCC and MSE on the hippocampus data

	Tissue-level PCC					Cell-level MSE			
	Astr.	CA region	Dent.	Inte.	Poly.	$K_{\text{ref}} = 10$	$K_{\text{ref}} = 30$	$K_{\text{ref}} = 50$	ALL
Celltrek	0.502	0.822	0.721	0.421	0.262	\	\	\	1120.889
scSpace	0.129	0.355	0.157	0.019	0.014	\	\	\	926.345
Seurat	0.453	0.574	0.516	0.570	0.439	814.535	818.980	819.747	819.747
SpaOTsc	0.707	0.785	0.783	0.714	0.367	859.681	867.138	878.930	1084.062
Tangram	0.337	0.821	0.698	**0.724**	**0.734**	628.639	626.379	628.677	631.207
Ours	**0.728**	**0.863**	**0.794**	0.674	0.621	**604.161**	**585.110**	**579.306**	**573.174**

*Astr. = Astrocytes, Dent. = Dentate pyramids, Inte. = Interneurons, Poly. = Polydendrocytes

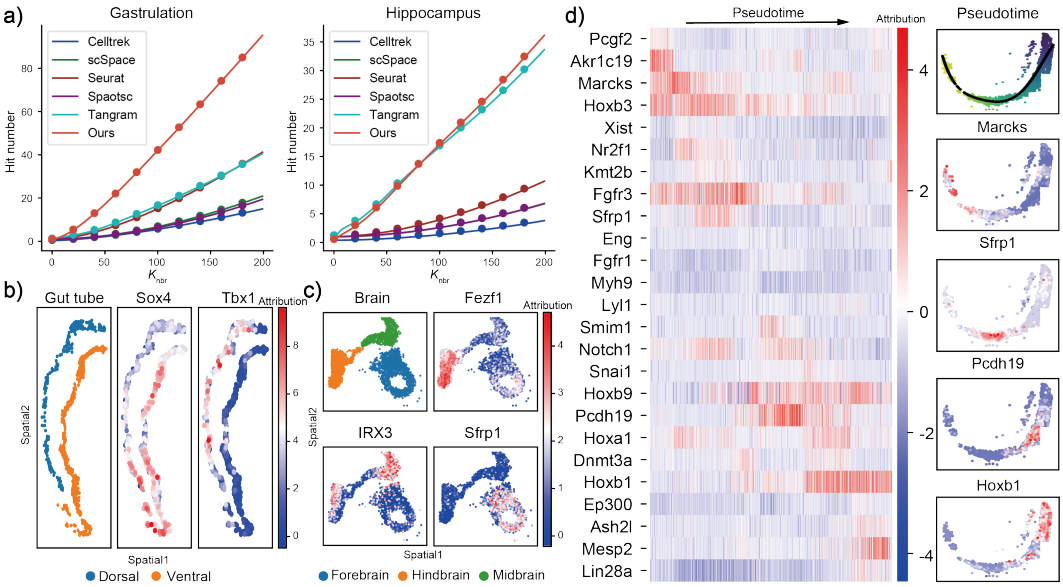

Figure 3: **a)** The hit number of different methods with different K_{nbr}. **b)** Spatial subclusters and the attribution values of representative SDGs in mouse gut tube. **c)** Spatial subclusters and the attribution value of representative SDGs in mouse brain. **d)** The attribution heatmaps and four representative SDGs in mouse spinal cord.

5.1.2 At the tissue level

As shown in Fig. A5, STEM and Tangram had the comparable performance across all cell types on the gastrulation data. All methods achieved PCCs higher than 0.6 for all cell types except for Seurat and scSpace. Seurat transfers cell type information from SC to ST data only using gene expression information, and scSpace focuses on reconstructing the spatial location of single cells in the latent space, which may lead to the low efficiency of these two methods in restoring the SC-ST spatial association. STEM got higher correlations on cardiomyocytes, gut tube, brain and spinal cord (Table 2 left). By visually checking these cell types, we found that they exhibit stronger signals in spatial distributions compared with other cell types (Fig. A6). From the biological perspective, the functions of these cells are highly correlated with the region they are located, so it is more easily to infer their spatial location information from gene expression profiles. For example, cells in forebrain, midbrain and hindbrain subregions participate in different functional processes, have different cell fates, and thus highly express different gene sets [24].

On the hippocampus data, the predicted distribution of different cell types is shown in Fig. A7. We observed that the performance of all methods was decreased (Table 3 left) in comparison with that of the gastrulation data, which may be because the SC data used here are not at real single-cell resolution. STEM and Tangram were the only two methods to achieve PCCs above 0.5 on all cell types (Fig. A8). STEM showed higher PCCs on astrocytes, Cornu Ammonis (CA) regions, dentate pyramids of

0.73, 0.86 and 0.79, respectively. These results support our intuition that by supervising the two-step spatial adjacency matrix in the association module, a good SC-ST mapping matrix could be learned.

5.1.3 Spatial dominate genes

We followed Lohoff's work [24] to spatially separate the gut tube into the dorsal side and ventral side on the gastrulation data. We found that *Tbx1*, *Smoc2* and *Kmt2b* are SDGs for the dorsal side, and *Sox4*, *Osr1* and *Podxl* are SDGs for the ventral side (Fig. 3b). Among these SDGs, the spatial patterns of *Tbx1* and *Smoc2* have been validated by the hybridization chain reaction experiments in the previous study [24]. We performed a similar procedure on the brain region of gastrulation data to spatially separate it into 3 subregions: forebrain, midbrain and hindbrain. We uncovered *Fezf1*,*Irx3* and *Sfrp1* are the representative SDGs of these three subregions, respectively (Fig. 3c). All these genes have been identified as marker genes of the corresponding brain region in previous studies [37–39].

STEM could reveal SDGs in a continuous manner. We focused on the spinal cord region and generated a spatial trajectory through its spatial distribution. Along the trajectory curve, we found several SDGs including *Hoxb1*, *Pcdh19*, *Sfrp1* and *Marcks* that showed continuous trends (Fig. 3d). These results demonstrate that STEM successfully identifies marker genes that have high contributions for determining cell location, which could provide insights into the spatial formation and evolution of cells in complex normal tissues or tumor microenvironments.

5.2 Discovery on the real dataset

We used STEM to map all the single cells in the hSCC data into the tissue spatial space (Fig. 4a). We checked whether the results obtained by STEM support the known descriptions illustrated by the original paper. As shown in Fig. 4b, we verified that a sub-cell type called tumor specific keratinocyte (TSK) is colocalized with the cancer-associated fibroblast and endothelial cell at the tumor leading edge. We compared the estimated location of TSKs with the sc-TSK score defined by Ji's study [19] (Fig. A9), and found STEM located TSKs in the high-score region. We then explored the spatial distribution of other keratinocyte (KC) subtypes, including normal, tumor basal, tumor cycling, and tumor differentiating KCs. We found that TSKs are spatially distant from them (Fig. 4c). This finding suggests that the formation of TSKs may be caused by the difference in microenvironment. We also found that pDC cells are located in the non-TSK leading edge region (Fig. 4d), which unveils the mechanism of the finding reported in the original paper that pDC-enriched gene functions are active in this region.

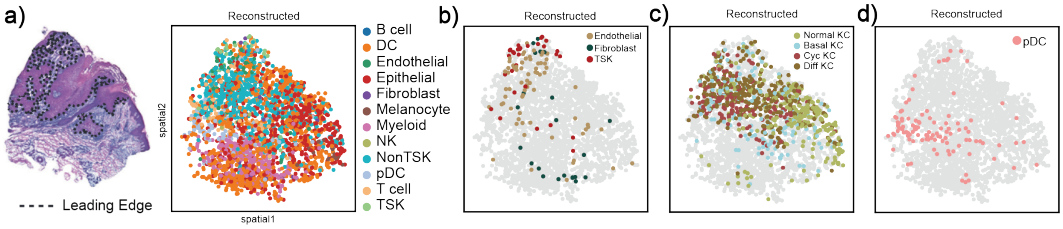

Figure 4: **a)** The hematoxylin-eosin (HE) stained image of the sequenced tissue and estimated cell locations given by our model. **b)** Spatial distribution of TSKs, endothelial cells and fibroblasts. **c)** Spatial distribution of other KC subtypes. **d)** Spatial distribution of pDC cells.

6 Discussion

Computationally building the spatial gene expression landscape at the single cell level can promote our understanding of heterogeneity and microenvironments in tissues. We proposed a model, STEM, to learn the spatially-aware representations of transcriptomic data via transfer learning. We designed a comprehensive pipeline to evaluate the different methods' ability of inferring spatial associations, which could serve as a benchmark for future work. We conducted two simulation experiments and one real dataset application. The results proved the superiority of STEM compared with other methods.

STEM is a powerful model to build up a single-cell level spatial landscape by integrating SC and ST data in a unified embedding space. Upon the model, the attribution technique is used to computationally highlight the SDGs with high contributions to cells' spatial location. These SDGs could benefit the discovery of biological mechanisms and even potential drug targets.

STEM takes the advantage of deep transfer learning to eliminate the SC-ST domain gap and infer spatial associations between and within domains. Our work demonstrates that transfer learning could not only unify data in different domains but also be feasible to establish mapping across domains, and knowledge can be discovered from the mapping via the attribution technique or other approaches.

Acknowledgments and Disclosure of Funding

The authors thank Sijie Chen, Haoxiang Gao and Yixin Chen for their helpful discussions.

This work was partially supported by National Natural Science Foundation of China(NSFC) (grants 62250005, 61721003 and 62103227), National Key R&D Program of China (grant 2021YFF1200901) and Tsinghua-Fuzhou Institute for Data Technology (TFIDT2021005).

References

1. Larsson, L., Frisén, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nature Methods* **18,** 15–18 (2021).
2. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nature Biotechnology* **40,** 308–318 (2022).
3. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nature Methods* **18,** 9–14 (2021).
4. Zhang, L. *et al.* Clinical and translational values of spatial transcriptomics. *Signal Transduction and Targeted Therapy* **7,** 111 (2022).
5. Walker, B. L., Cang, Z., Ren, H., Bourgain-Chang, E. & Nie, Q. Deciphering tissue structure and function using spatial transcriptomics. *Communications biology* **5,** 1–10 (2022).
6. Tavares-Ferreira, D. *et al.* Spatial transcriptomics of dorsal root ganglia identifies molecular signatures of human nociceptors. *Science translational medicine* **14,** eabj8186 (2022).
7. Anderson, A. C. *et al.* Spatial transcriptomics. *Cancer Cell* **40,** 895–900 (2022).
8. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology* **39,** 313–319 (2021).
9. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353,** 78–82 (2016).
10. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods* **16,** 987–990 (2019).
11. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363,** 1463–1467 (2019).
12. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* **24,** 425–436 (2021).
13. Chen, S. *et al.* hECA: The cell-centric assembly of a cell atlas. *Iscience* **25,** 104318 (2022).
14. Wang, R., Peng, G., Tam, P. P. & Jing, N. Integration of computational analysis and spatial transcriptomics in single-cell study. *Genomics, Proteomics & Bioinformatics* (2022).
15. Liao, J. *et al. Reconstruction of the cell pseudo-space from single-cell RNA sequencing data with scSpace* tech. rep. (2022).
16. Wei, R. *et al.* Spatial charting of single-cell transcriptomes in tissues. *Nature biotechnology,* 1–10 (2022).
17. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic acids research* **49,** e50–e50 (2021).
18. Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature,* 1–12 (2022).
19. Ji, A. L. *et al.* Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182,** 497–514 (2020).

20. Madissoon, E. *et al. A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche* tech. rep. (2021).
21. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184,** 3573–3587 (2021).
22. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature Methods* **18,** 1352–1362 (2021).
23. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nature communications* **11,** 1–13 (2020).
24. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature Biotechnology* **40,** 74–85 (2022).
25. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. & Smola, A. A kernel two-sample test. *The Journal of Machine Learning Research* **13,** 723–773 (2012).
26. Sundararajan, M., Taly, A. & Yan, Q. *Axiomatic attribution for deep networks* in *International conference on machine learning* (2017), 3319–3328.
27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15,** 1929–1958 (2014).
28. Haeusser, P., Frerix, T., Mordvintsev, A. & Cremers, D. *Associative domain adaptation* in *Proceedings of the IEEE international conference on computer vision* (2017), 2765–2773.
29. Haeusser, P., Mordvintsev, A. & Cremers, D. *Learning by association–A versatile semi-supervised training method for neural networks* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 89–98.
30. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods* **11,** 360–361 (2014).
31. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348,** aaa6090 (2015).
32. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185,** 1777–1792 (2022).
33. Paszke, A. *et al.* Automatic differentiation in pytorch (2017).
34. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).
35. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
36. Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896* (2020).
37. Eckler, M. J. & Chen, B. Fez family transcription factors: controlling neurogenesis and cell fate in the developing mammalian nervous system. *BioEssays* **36,** 788–797 (2014).
38. Robertshaw, E., Matsumoto, K., Lumsden, A. & Kiecker, C. Irx3 and Pax6 establish differential competence for Shh-mediated induction of GABAergic and glutamatergic neurons of the thalamus. *Proceedings of the National Academy of Sciences* **110,** E3919–E3926 (2013).
39. Trevant, B. *et al.* Expression of secreted frizzled related protein 1, a Wnt antagonist, in brain, kidney, and skeleton is dispensable for normal embryonic development. *Journal of cellular physiology* **217,** 113–126 (2008).

Appendix

1 Spatial adjacency matrix

Giving the ground truth ST data spatial coordinate $Y^{ST} \in \mathbb{R}^{N\times 2}$, the pairwise spatial association strength S_{ij} in the ST spatial adjacency matrix S is calculated by the kernel function between coordinate Y_i^{ST} of cell i and Y_j^{ST} of cell j. In our work we use the Gaussian kernel function:

$$\phi(Y_i^{ST}, Y_j^{ST}) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{1}$$

where σ is the standard deviation and controls the width of the Gaussian 'bell'. To select a proper size of σ, on the two simulation data we traversed the values from 0 to 3 times of distance between two adjacent spots, and use the hit number as the evaluation metric. We found that by setting σ as the half of the adjacent distance ($\sigma = 3$ on the gastrulation data and $\sigma = 55$ on the hippocampus data), our model achieved the highest hit number.

2 Maximum mean discrepancy

Maximum mean discrepancy (MMD) is a distance between the mean embeddings of distributions defined in a reproducing kernel Hilbert space $\mathcal{H}_k$ with a characteristic kernel k. It has widely applied in Transfer Learning models. In general, the data X in source domain follows the distribution P_x, and the data Y in target domain follow the distribution P_y. And the mean embeddings of these distributions P_x and P_y in $\mathcal{H}$ are unique elements $\mu_k(P_x)$ and $\mu_k(P_y)$. Then the MMD is

$$\mathrm{MMD}(P_x, P_y) = \|\mu_k(P_x) - \mu_k(P_y)\|_{\mathcal{H}_k} \tag{2}$$

In practice, this distance has its empirical estimation via the kernel trick. In our work, we implemented the multi-kernel MMD, and each kernel has is the Gaussian kernel with different bandwidth.

3 Integrated gradient

A common way for humans to perform attribution relies on counterfactual intuition. To assign blame to a certain cause, the Integrated Gradient technology implicitly consider the absence of the cause as a baseline for comparing outcomes. Specifically, the absence is modeled as a baseline input vector. For a deep neural network $F : R^n \to [0, 1]$, let x be the input and x' be the baseline. The integrated gradient (IG) is the gradient at all points along the path from the baseline x' to the input x. Specifically, For i^{th} dimension for an input, the integrated gradient is defined as follows:

$$\mathrm{IG} := (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha \times (x_i - x_i'))}{\partial x_i} \tag{3}$$

Compared with other attribution methods, the IG satisfies the axiom of sensitivity while the others like gradients-based methods violate sensitivity. The sensitivity is defined as: If for every input and baseline that differ in one feature and have different predictions, the differing feature should be given a non-zero attribution.

4 Simulation data generation

We placed the pseudo ST spots on the crossing point of the grid. We generated a 30×40 grid and 50×30 grid for gastrulation and hippocampus data, respectively. The gene expression profile of each spot is obtained by summing the expression of its surrounding single cells. We kept the spots which contains more than 3 cells. It is noticeable that in real spatial data the tissue cannot be fully covered by spots, so the transcripts of some single cells cannot be captured. We also take this into account in the simulation process. The gene expression of each spot aggregates only about 50-70% of the local surrounding single cells. In other words, one-third of single cells gene expression profiles are not included in the ST data. As shown in Fig. A1, only the cells in red compose the spot data.

Then the cell type proportion of each spot is calculated based on its contained single cells. So the generated pseudo ST data contain gene expression profile, spatial coordinate and cell type proportion information. For gene selection, we kept all 351 genes on the gastrulation data and selected the top 5,000 highly variable genes for SC and ST data on the hippocampus data.

5 Supplemental Figures

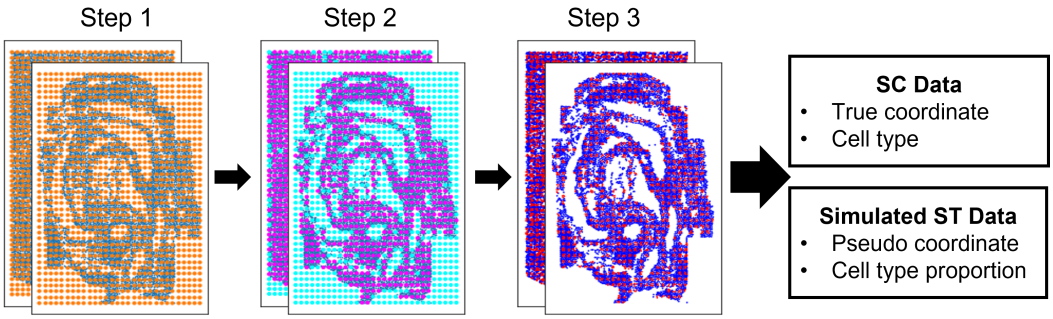

Figure A1: the pipeline of pseudo ST data simulation. **Step1**: Creating spatial grid. **Step2**: Filtering out spots covering no more than 3 cells. Only purple spots are kept. **Step3**: Aggregating partial cell gene expression. Only cells in red contribute the spot gene expression.

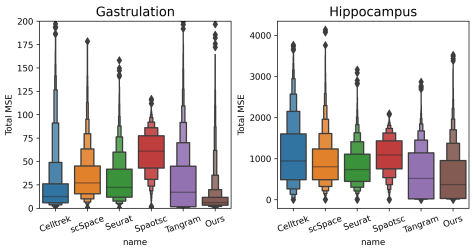

Figure A2: The boxplot of total MSE on gastrulation and hippocampus data. The lower the better.

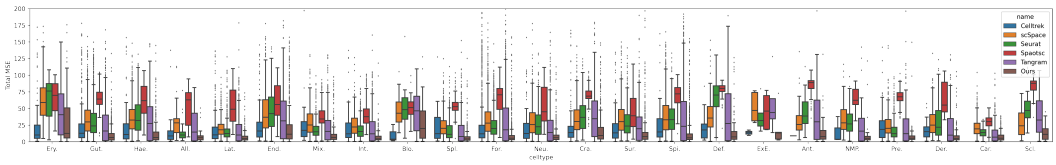

Figure A3: MSE of all approaches acorss different cell types on the gastrulation data. The lower the better. Full name in X axis: Erythroid, Gut tube, Haematoendothelial progenitors, Allantois, Lateral plate mesoderm, Endothelium, Mixed mesenchymal mesoderm, Intermediate mesoderm, Blood progenitors, Splanchnic mesoderm, Forebrain/Midbrain/Hindbrain, Neural crest, Cranial mesoderm, Surface ectoderm, Spinal cord, Definitive endoderm, ExE endoderm, Anterior somitic tissues, NMP, Presomitic mesoderm, Dermomyotome, Cardiomyocytes, Sclerotome

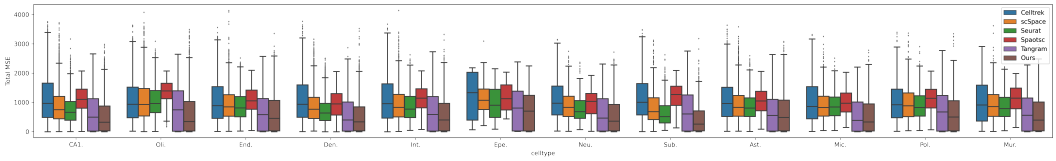

Figure A4: MSE of all approaches acorss different cell types on the hippocampus data. The lower the better. Full name in X axis: CA1/CA2/CA3 Subiculum, Oligodendrocytes, Endothelial Tip, DentatePyramids, Interneurons, Ependymal, Neurogenesis, Subiculum Entorhinal cl3, Astrocytes, Microglia, Endothelial Stalk, Polydendrocytes, Subiculum Entorhinal cl2, Mural

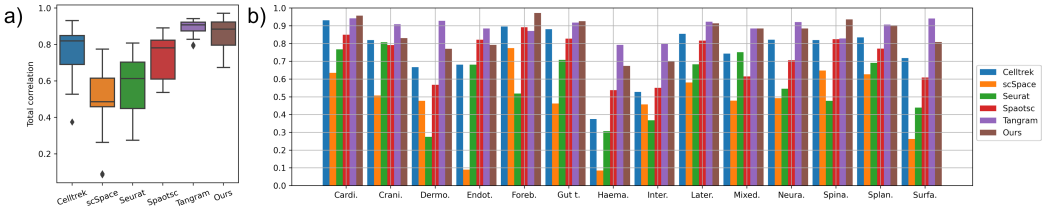

Figure A5: **a)** Overall PCC of all approaches on the gastrulation data. **b)** PCC of all approaches acorss different cell types on the gastrulation data. Name in X axis: Cardiomyocytes, Cranial mesoderm, Dermomyotome, Endothelium, Forebrain/Midbrain/Hindbrain, Gut tube, Haematoendothelial progenitors, Intermediate mesoderm, Lateral plate mesoderm, Mixed mesenchymal mesoderm, Neural crest, Spinal cord, Splanchnic mesoderm, Surface ectoderm

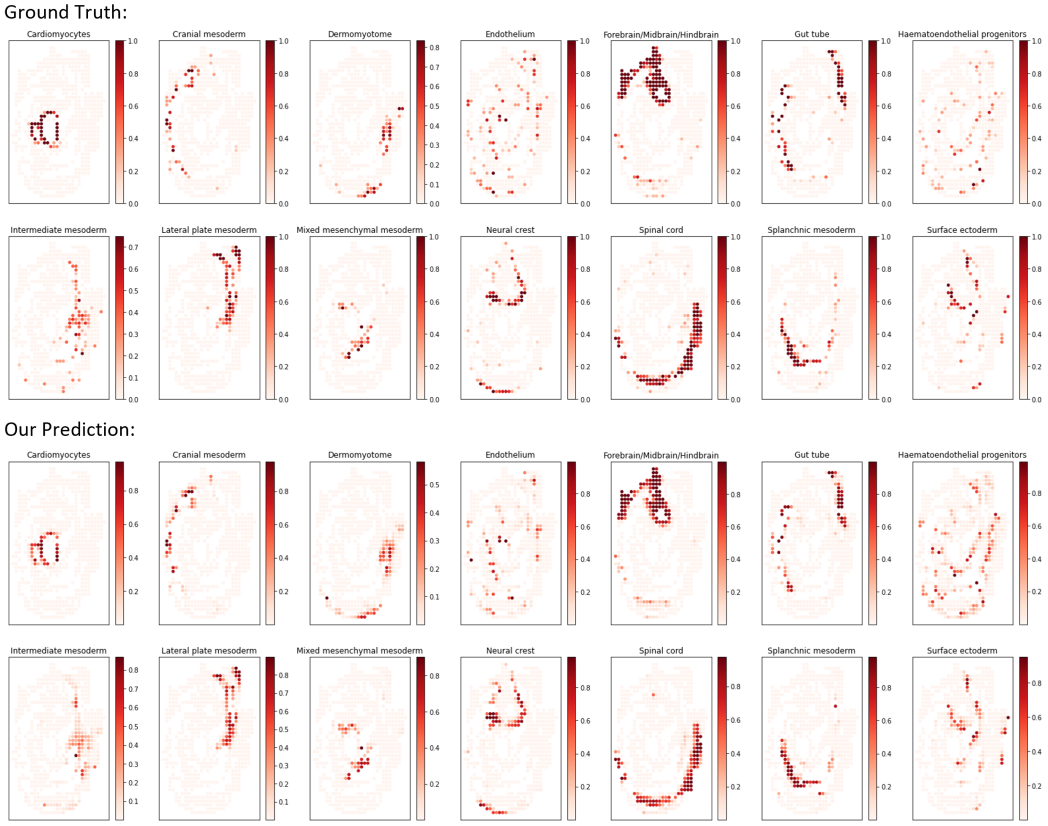

Figure A6: Visualization of cell type spatial distribution on the gastrulation data. Above: Ground truth. Bottom: Our prediction

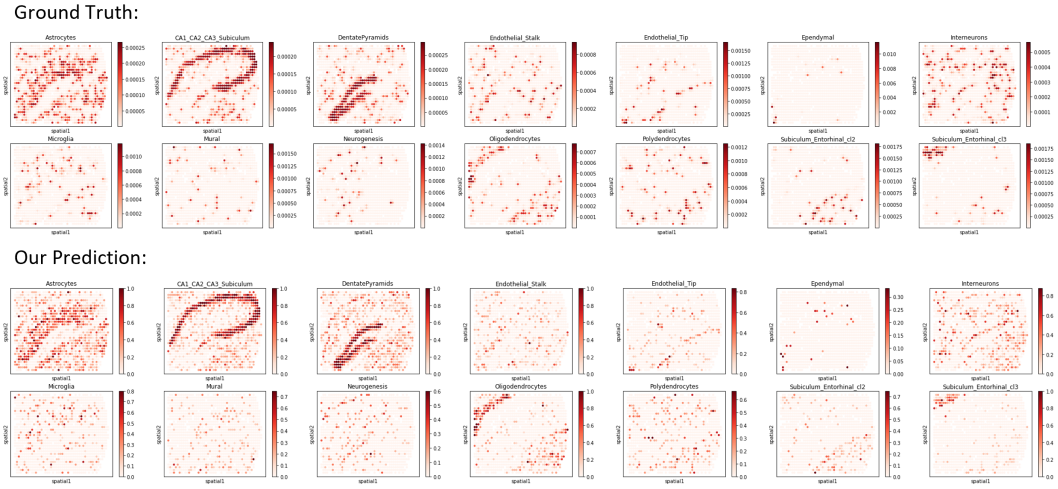

Figure A7: Visualization of cell type spatial distribution on the hippocampus data. Above: Ground truth. Bottom: Our prediction

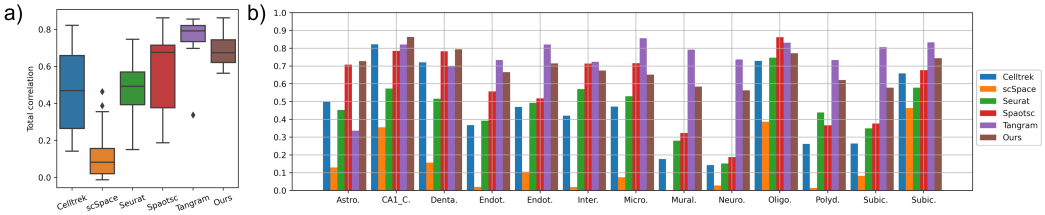

Figure A8: **a)** Overall PCC of all approaches on the hippocampus data. **b)** PCC of all approaches acorss different cell types on the hippocampus data. Name in X axis: Astrocytes, CA1/CA2/CA3 Subiculum, DentatePyramids, Endothelial Stalk, Endothelial Tip, Interneurons, Microglia, Mural, Neurogenesis, Oligodendrocytes, Polydendrocytes, Subiculum Entorhinal cl2, Subiculum Entorhinal cl3

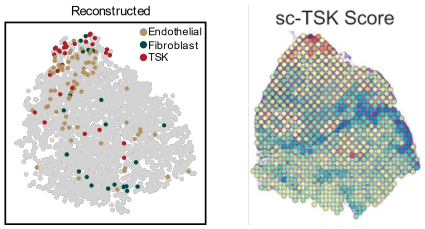

Figure A9: The predicted spatial distribution of TSK, endothelial and firbroblast cells (left) and the sc-TSK score reported on Ji's study.