Evidence-aware Integration and Domain Identification of Spatial Transcriptomics Data

Wei Zhang[1], Siyu Yi[2*], Lezhi Chen[3], Yifan Wang[4], Ziyue Qiao[5], Yongdao Zhou[6], Wei Ju[1*]

[1] College of Computer Science, Sichuan University, Chengdu, China
[2] College of Mathematics, Sichuan University, Chengdu, China
[3] College of Electrical Engineering, Sichuan University, Chengdu, China
[4] School of Information Technology & Management, University of International Business and Economics, Beijing, China
[5] School of Computing and Information Technology, Great Bay University, Dongguan, China
[6] NITFID, School of Statistics and Data Science, Nankai University, China
wayc04@gmail.com, {siyuyi, juwei}@scu.edu.cn, chenlezhi@stu.scu.edn.cn, yifanwang@uibe.edu.cn, zyqiao@gbu.edu.cn, ydzhou@nankai.edu.cn

Abstract

Spatial transcriptomics (ST) enables joint profiling of gene expression and spatial positions, thereby revealing spatially resolved biological functions. However, many existing ST analysis methods often fail to explicitly quantify the belief and uncertainty in decisions caused by noisy ST data, making it difficult to handle spots of varying quality in a fine-grained manner. In addition, domain identification is a fundamental and critical task in ST, but commonly used models that separate expression learning and clustering often struggle to learn cluster-friendly latent representations effectively. To address these issues, we propose PREST, a prototype-based evidence-aware integration framework for ST data. PREST performs multi-scale representation learning with fine-grained attention fusion and introduces learnable class prototypes to quantify belief and uncertainty in model decisions. We aim to align overall belief scores with latent semantic information to enhance uncertainty quantification and prototype learning, thereby promoting the learning of clustering-friendly representations. PREST further integrates an uncertainty-aware reconstruction module and spatial regularization to reduce overfitting to unreliable spots and promote denoised, discriminative representations. Extensive experiments on several benchmark datasets validate the effectiveness and superiority of our proposed PREST across various downstream tasks.

Code — https://github.com/wayc04/PREST
Appendix — https://github.com/wayc04/PREST

Introduction

Spatial transcriptomics (ST) captures gene expression while preserving the spatial context of spots within tissue sections (Williams et al. 2022), enabling comprehensive analysis of tissue architecture and cellular organization (Rao et al. 2021; Asp, Bergenstråhle, and Lundeberg 2020). By jointly profiling molecular and spatial information, ST provides unprecedented insights into how cells interact and function within their native environments. Current ST technologies can be broadly categorized into imaging-based in situ methods, such as MERFISH (Moffitt et al. 2018), osmFISH (Codeluppi et al. 2018) and seqFISH (Lubeck et al. 2014; Shah et al. 2016), and barcode-based sequencing approaches, including 10x Visium (Ji et al. 2020), SLIDE-seq (Rodriques et al. 2019), and Stereo-seq (Chen et al. 2022). Since the resolution of most platforms does not reach single-cell level, gene expression is measured at discrete units called *spots*, each aggregating transcripts from multiple cells. Despite this limitation, ST has rapidly become a cornerstone technology in spatial genomics and has motivated the development of numerous computational methods tailored to analyze such data.

ST integrates gene expression and spatial information to decode biological processes, such as cell distribution, gene regulation, and tissue heterogeneity, offering deeper insights into tissue organization and function. Early approaches primarily employed unsupervised algorithms, including PCA, k-means (Likas, Vlassis, and Verbeek 2003) and Louvain (Blondel et al. 2008). These methods treat spots as independent observations and overlook the spatial continuity inherent in tissues. To better capture spatial dependencies, recent methods have leveraged graph-based approaches, particularly *Graph Neural Networks* (GNNs, Kipf and Welling 2017; Veličković et al. 2018; Ju et al. 2024b, 2025), which represent spots as nodes in a graph and incorporate both gene expression and neighborhood relationships (Liu et al. 2023; Li et al. 2022; Dong and Zhang 2022; Zhu et al. 2024). DeepST (Xu et al. 2022) pioneers domain identification with deep learning, enabling effective batch integration. GraphST (Long et al. 2023) builds on this with self-supervised contrastive learning, improving clustering and multi-slice analysis. SpaGCN (Hu et al. 2021) then uses graph convolution to integrate gene expression and spatial data, detecting coherent expression patterns. stMMR (Zhang et al. 2024) enhances this with self-attention for robust multimodal learning, while DUSTED (Zhu et al. 2025) combines gene channel and attention in a graph autoencoder to capture spatial features and expression variability.

*Corresponding authors

Despite the success of existing ST analysis methods, several intrinsic limitations remain. (1) *First*, ST data is inherently noisy due to low capture efficiency and shallow sequencing depth stemming from cost control or technical constraints, as well as biological factors such as variations in cell permeability that cause mRNA drift or diffusion (Du et al. 2024). These factors lead to high sparsity and over-dispersion in ST data, introducing uncertainty in downstream analyses and decisions. Existing methods often denoise data implicitly without explicitly modeling the belief and uncertainty of model decisions, including autoencoder-based (Xu et al. 2022; Long et al. 2023; Zhu et al. 2025), smoothing-based (Holdener and De Vlaminck 2025), and image-enhanced approaches (Wang et al. 2022). These methods typically treat noise as an undesirable factor to be removed, processing all spots equally without indicating which ones are unreliable, and may overfit to low-quality or boundary regions. (2) *Second*, clustering analysis serves as the foundation for many ST downstream tasks, including spatial domain identification, cell type annotation, differentially expressed gene detection, and trajectory inference. However, many deep learning-based ST methods focus primarily on representation learning while neglecting the modeling of class information. A common paradigm involves first learning embeddings and then performing clustering on them using algorithms such as k-means, Leiden, Louvain, or GMM (Long et al. 2023; Zhang et al. 2024; Zhu et al. 2025). This two-step process is often not clustering-friendly, resulting in suboptimal performance in downstream analyses.

To address these issues, we propose the PRototype-based Evidence-aware integration framework for Spatial Transcriptomics data (PREST). We begin by performing multi-scale representation learning followed by fine-grained cross-attention fusion. Then, we introduce learnable class prototypes to capture cluster information and utilize subjective logic (SL, Jøsang 2016) to explicitly quantify both *evidence / belief* and *uncertainty* of model decisions caused by noisy or dropout ST data. To enhance the reliability of these quantifications and prototype learning, we impose constraint to encourage the belief scores to capture the overall semantic information in the latent space. To effectively handle the highly sparse nature of ST data, we adopt the zero-inflated negative binomial (ZINB, Yu et al. 2022) model to reconstruct original expressions. We also leverage spatial coordinates to regularize spot representations, enabling learning discriminative representations enriched with spatial context. Furthermore, we incorporate quantified uncertainties into reconstruction and regularization to assess spot reliability, enabling the model to softly down-weight low-belief regions. It reduces overfitting to low-quality data and enhances overall model trustworthiness. Finally, we conduct various downstream analyses based on the learned spot representations and class prototypes. Extensive experiments on multiple datasets show the superiority of our framework.

Preliminaries & Problem Definition

Notations. Let $\mathbf{P} = \{(u_i, v_i)\}_{i=1}^{N}$ denote the spatial coordinates of N spots and $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^\top \in \mathbb{R}^{N \times G}$ be the gene expression matrix after preprocessing, where G is the number of genes and $\boldsymbol{x}_i \in \mathbb{R}^G$ corresponds to the expression profile at spot i. The objective is to learn class prototypes $\mathbf{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K)^\top \in \mathbb{R}^{K \times D}$ and spot representations $\mathbf{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N)^\top \in \mathbb{R}^{N \times D}$ with $D \ll G$ that jointly encodes both the spatial positions and gene expressions.

Downstream Analyses. We perform several downstream analysis tasks, including: (i) *Domain identification (or spatial clustering)*, which assigns coherent regions by grouping similar representations; (ii) *Differentially expressed gene (DEG) detection*, which identifies marker genes characteristic of each spatial domain; (iii) *Expression imputation*, which predicts missing or unobserved gene expression values to enhance data completeness and downstream analysis. and (iv) *Enrichment analysis*, which interprets the biological significance of identified genes or clusters by linking them to known pathways or ontologies.

Methodology

In this section, we introduce a novel framework, PREST, designed for domain identification and related downstream analyses in ST. The framework is composed of three key modules: (i) *Multi-scale Representation Extraction and Fusion*; (ii) *Prototype-based Belief and Uncertainty Estimation*; and (iii) *Uncertainty-weighted Expression Reconstruction and Spatial Regularization*. An overview of the entire pipeline is depicted in Figure 1.

Multi-scale Representation Extraction and Fusion

ST enables the identification of similar cells by incorporating spatial context. To effectively utilize this spatial information and take advantage of the expressive power of GNNs, we first transform the spatial coordinates $\mathbf{P}$ into an undirected graph. Specifically, we construct the spatial adjacency matrix $\mathbf{A}$ using a fixed-radius nearest neighbor approach, and assign the expression matrix $\mathbf{X}$ as the node attributes.

In GNNs, the graph convolutional network (GCN, Kipf and Welling 2017) updates node representations by aggregating information from their neighbors, thereby effectively capturing the structural patterns of graph data. The layer-wise propagation rule of a GCN is typically defined as:

$$\mathbf{Z}^{\mathrm{s},(l+1)} = \sigma(\tilde{\mathbf{A}}\mathbf{Z}^{\mathrm{s},(l)}\mathbf{W}^{\mathrm{s},(l)}), \tag{1}$$

where $\tilde{\mathbf{A}}$ denotes the normalized adjacency matrix, $\mathbf{W}^{\mathrm{s},(l)}$ is the trainable weight matrix, and $\sigma(\cdot)$ is a non-linear activation function. In contrast, multi-layer perceptrons (MLPs) operate purely on node attributes, making them particularly effective at capturing semantic patterns embedded in attribute space. A standard MLP layer can be form

$$\mathbf{Z}^{\mathrm{e},(l+1)} = \phi\left(\mathbf{Z}^{\mathrm{e},(l)}\mathbf{W}^{\mathrm{e},(l)} + \boldsymbol{b}^{\mathrm{e},(l)}\right), \tag{2}$$

where $\mathbf{W}^{\mathrm{e},(l)}$ and $\boldsymbol{b}^{\mathrm{e},(l)}$ are learnable parameters, and $\phi(\cdot)$ is an activation function such as ReLU. To further enhance representational capacity and stabilize training, each encoder layer can be augmented with self-attention (Vaswani et al. 2017) and skip connection (He et al. 2016), which help mitigate issues like vanishing or exploding gradients.

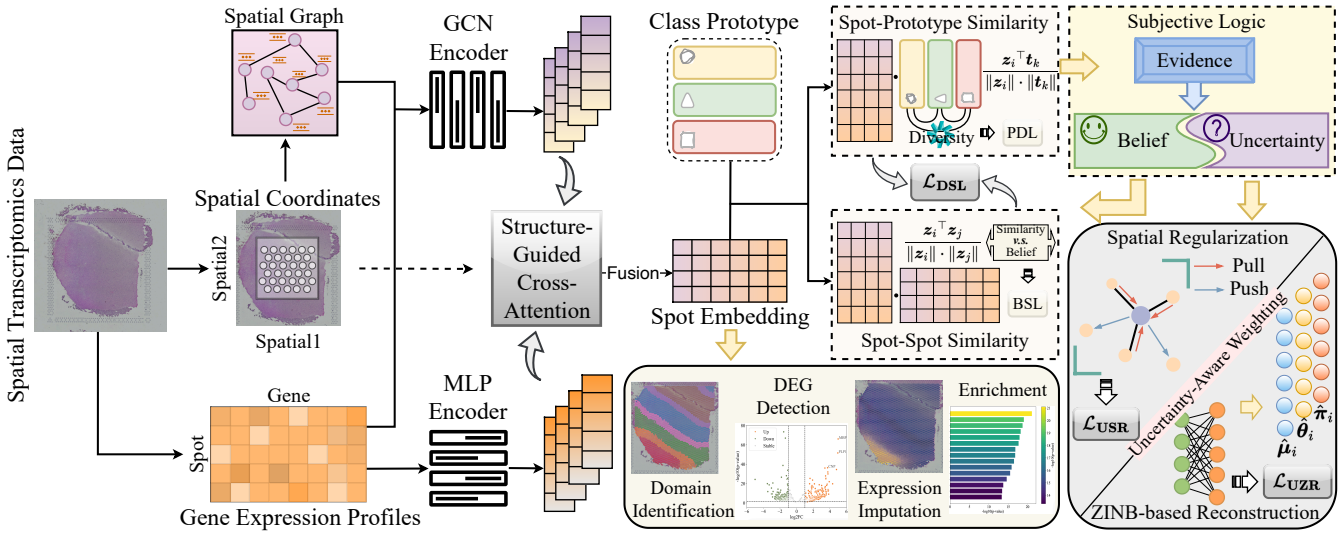


Figure 1: Illustration of the proposed framework PREST.

In ST data, due to technical limitation or biological variation, the measured expression data can be inherently noisy. Meanwhile, due to limitations in imaging resolution, tissue deformation, or inaccuracy in graph construction heuristics, the spatial graph may be also not precise. In such cases, when expression signals are strong but the spatial structure is unreliable, MLPs can help mitigate the influence of noisy connections by focusing on attribute-driven features. Conversely, when expression is sparse or noisy, the spatial topology can provide useful contextual cues that GCN is well-suited to exploit. Even in challenging scenarios where both gene expression and spatial structure are affected by noise, learning from multiple scales remains beneficial, as it enables the model to integrate complementary patterns across views and attenuate view-specific noise. To this end, we integrate GCN and MLP into a unified framework for multi-scale representation learning, facilitating the extraction of more robust and informative embeddings.

Based on the embeddings obtained from GCN and MLP encoders, denoted as $\mathbf{Z}^{\mathrm{s}}$ and $\mathbf{Z}^{\mathrm{e}}$, we introduce a structure-guided cross-attention mechanism (CroAtt) to perform fine-grained representation fusion. The fusion process is:

$$\mathbf{Z} = \mathrm{Norm}\Big(\mathrm{CroAtt}(\mathbf{Z}^{\mathrm{s}}, \mathbf{Z}^{\mathrm{e}}) + \mathrm{CroAtt}(\mathbf{Z}^{\mathrm{e}}, \mathbf{Z}^{\mathrm{s}})\Big), \tag{3}$$

$$\mathrm{CroAtt}(\mathbf{Z}^{\mathrm{s}}, \mathbf{Z}^{\mathrm{e}}) = \mathrm{Softmax}\left(\mathrm{Mask}\left(\frac{\mathbf{Z}^{\mathrm{e}}\mathbf{Z}^{\mathrm{s}\top}}{\sqrt{D}}, \mathbf{A}\right)\right)\mathbf{Z}^{\mathrm{s}}.$$

This mechanism dynamically evaluates the contribution of spatial- and expression-level representations to one another, thereby enabling adaptive and content-aware fusion. The masking operation based on the spatial adjacency matrix helps eliminate redundant or spurious dependencies, ensuring that the attention is focused on meaningful local interactions. By explicitly modeling fine-grained dependencies across the two encoder outputs, rather than performing naive cross-view averaging, it facilitates a more comprehensive integration of spatial and expression information.

Prototype-based Belief and Uncertainty Estimation

The presence of inherent noise in the ST data inevitably leads to predictive uncertainty (Malinin and Gales 2018). Therefore, while ensuring representation richness through multi-scale extraction, it is also crucial to quantify the decision uncertainty of the fused information and leverage it to enhance model training. To this end, we adopt subjective logic (SL, Jøsang 2016; Li et al. 2023) to characterize the *belief* or *evidence* of domain identification decision, as well as the *uncertainty* caused by the lack of sufficient evidence.

Formally, a spot i ($i \in \{1, \ldots, N\}$) in SL is associated with a multinomial opinion consisting of a belief mass distribution $\{b_{ik}\}_{k=1}^{K}$ over K clusters and an uncertainty value u_i, satisfying the constraint $\sum_{k=1}^{K} b_{ik} + u_i = 1$ and $b_{ik}, u_i \geq 0$. Let $e_{ik} (\geq 0)$ represent the extracted evidence from data supporting the association of spot i with cluster k. The belief and uncertainty are then computed as:

$$b_{ik} = \frac{e_{ik}}{\sum_{k=1}^{K}(e_{ik}+1)}, \quad u_i = \frac{K}{\sum_{k=1}^{K}(e_{ik}+1)}. \tag{4}$$

To facilitate the learning of cluster-relevant evidence and support domain identification, we introduce learnable class prototypes $\mathbf{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K)^\top$, initialized via k-means on pre-trained spot embeddings (details provided in the Experiment Section). Based on this and fused representation matrix $\mathbf{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N)^\top$ from Eq. (3), we define the spot-to-prototype similarity matrix $\mathbf{E} = (E_{ik}) \in \mathbb{R}^{N \times K}$ and the spot-to-spot similarity matrix $\mathbf{S} = (S_{ij}) \in \mathbb{R}^{N \times N}$ as:

$$E_{ik} = \frac{\boldsymbol{z}_i{}^\top \boldsymbol{t}_k}{\|\boldsymbol{z}_i\| \cdot \|\boldsymbol{t}_k\|}, \quad S_{ij} = \frac{\boldsymbol{z}_i{}^\top \boldsymbol{z}_j}{\|\boldsymbol{z}_i\| \cdot \|\boldsymbol{z}_j\|}. \tag{5}$$

The similarity score E_{ik}, after applying a positive activation, can be interpreted as the evidence e_{ik}, allowing us to compute belief score b_{ik} and uncertainty u_i according to Eq. (4).

To ensure meaningful quantification, we first impose constraints on the class prototypes by encouraging high diversity among them, which promotes better separability. To this

end, we introduce a prototype diversity learning (PDL) loss:

$$\mathcal{L}_{\text{PDL}} = \frac{1}{2K(K-1)} \sum\nolimits_{1 \le k < k' \le K} \left(\frac{\boldsymbol{t}_k^\top \boldsymbol{t}_{k'}}{\|\boldsymbol{t}_k\| \cdot \|\boldsymbol{t}_{k'}\|} \right)^2.$$

With the prototypes, to ensure better uncertainty quantification, the prototype-based beliefs should effectively encapsulate the global semantic structure of the data in the latent space. To encourage this, we require that the total beliefs assigned to a spot aligns closely with its total similarities to all other spots. Accordingly, we introduce the belief-based semantics learning (BSL) loss:

$$\mathcal{L}_{\text{BSL}} = \frac{1}{N} \sum\nolimits_{i=1}^{N} \left(\frac{1}{K} \sum\nolimits_{k=1}^{K} b_{ik} - \frac{\lambda}{N} \sum\nolimits_{j=1, j \neq i}^{N} S_{ij} \right)^2,$$

where λ is a scaling factor that compensates for potential differences in magnitude. We then formulate a joint diversity-semantics learning loss (DSL) as:

$$\mathcal{L}_{\text{DSL}} = \mathcal{L}_{\text{PDL}} + \mathcal{L}_{\text{BSL}}. \tag{6}$$

Uncertainty-weighted Expression Reconstruction and Spatial Regularization

To effectively handle highly sparse ST data and impute dropout values, we introduce a ZINB-based decoder built upon the fused representations to reconstruct the original gene expression in a self-supervised manner. It is worth noting that the raw expression data may contain noise; thus, different spots should be assigned different self-supervision weights to prevent overfitting on low-quality data. Technically, we leverage the uncertainty estimated in Eq. (4) to reweight the loss, assigning lower weights to spots with higher uncertainties. It leads to the uncertainty-weighted ZINB-based reconstruction (UZR) loss:

$$\mathcal{L}_{\text{UZR}} = -\sum\nolimits_{i=1}^{N} \omega_i \log \text{ZINB}(\boldsymbol{x}_i \,|\, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\pi}}_i), \tag{7}$$
$$\text{with } \omega_i = \frac{1/u_i}{\sum_{j=1}^{N} 1/u_j},$$

where $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\pi}}$ denote the estimations of ZINB parameters $\boldsymbol{\mu}$ (the expected gene expression), $\boldsymbol{\theta}$ (the over-dispersion characteristic of count-based data), and $\boldsymbol{\pi}$ (the dropout probability arising from technical noise or low expression levels). With the dropout indicator function $\delta_0(\cdot)$, the ZINB distribution is formulated as:

$$\text{ZINB}(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\pi}) = \boldsymbol{\pi} \delta_0(\boldsymbol{x}) + (1 - \boldsymbol{\pi}) \,\text{NB}(\boldsymbol{x}), \text{ with}$$
$$\text{NB}(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\theta}) = \Gamma(\boldsymbol{x} + \boldsymbol{\theta}) \left[\boldsymbol{\theta}/(\boldsymbol{\theta} + \boldsymbol{\mu})\right]^{\boldsymbol{\theta}} \left[\boldsymbol{\mu}/(\boldsymbol{\theta} + \boldsymbol{\mu})\right]^{\boldsymbol{x}} / [\boldsymbol{x}! \Gamma(\boldsymbol{\theta})].$$

We utilize three distinct MLPs (g_{enc}) to estimate the parameters of the ZINB distribution from the fused latent representation $\mathbf{z}_i$. Specifically, each MLP is responsible for predicting one of the distribution parameters:

$$\hat{\boldsymbol{\mu}}_i = \text{Exp}\left(g_{\text{dec1}}(\boldsymbol{z}_i)\right), \hat{\boldsymbol{\theta}}_i = \text{Softplus}\left(g_{\text{dec2}}(\boldsymbol{z}_i)\right),$$
$$\hat{\boldsymbol{\pi}}_i = \text{Sigmoid}\left(g_{\text{dec3}}(\boldsymbol{z}_i)\right), i \in \{1, \ldots, N\},$$

where the use of non-linear activations ensures the validity of the parameter ranges required by the ZINB distribution.

To preserve the intrinsic spatial structure in the latent space and also enhance the discriminative power of learned representations, we introduce a spatial regularization loss that encourages neighboring spots to have similar representations while pushing apart those that are not spatially connected. Specifically, with $\mathbf{A} = (A_{ij})$, we define the uncertainty-weighted spatial regularization (USR) loss as:

$$\mathcal{L}_{\text{USR}} = -\sum\nolimits_{i=1}^{N} \omega_i \left[\sum\nolimits_{A_{ij}=1} \log \phi(S_{ij}) + \sum\nolimits_{A_{iq}=0} \log(1 - \phi(S_{iq})) \right], \tag{8}$$

where S_{ij} is defined in Eq. (5) and $\phi(\cdot)$ is a sigmoid function that transforms the similarity scores into the range $[0, 1]$. The weight ω_i, consistent with the definition in Eq. (8), is similarly designed to down-weight the supervisory signals from spurious structures on the latent representations, thereby enabling the model to learn spot representations that better reflect the underlying patterns of the ST data.

Joint Optimization. In summary, the proposed PREST framework combines three loss components: (i) the joint diversity-semantics learning loss $\mathcal{L}_{\text{DSL}}$, (ii) the ZINB-based reconstruction loss $\mathcal{L}_{\text{UZR}}$, and (iii) the spatial regularization loss $\mathcal{L}_{\text{USR}}$. The overall training objective is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{DSL}} + \alpha \mathcal{L}_{\text{UZR}} + \beta \mathcal{L}_{\text{USR}}, \tag{9}$$

where α and β are tunable coefficients that balance the influence of reconstruction and spatial constraints.

After training converges, domain identification is achieved by aligning the optimized spot representations with the class prototypes through belief-based matching. Then the clustering results are used for various downstream analyses to validate the domain identification outcomes. The training procedure of our proposed PREST is summarized in Algorithm 1 in the Appendix.

Experiments

Experimental Setup

To validate the superiority of our proposed PREST, we conduct extensive experiments on three benchmark *datasets*, i.e., DLPFC (Maynard et al. 2021), HBC (Buache et al. 2011), and MAB (Dong 2008). We conduct comparisons on the performance of domain identification with several state-of-the-art *models*, including SCANPY (Wolf, Angerer, and Theis 2018), SpaGCN (Hu et al. 2021), STAGATE (Dong and Zhang 2022), DeepST (Xu et al. 2022), stLearn (Pham et al. 2023), SCGDL (Liu et al. 2023), GraphST (Long et al. 2023), stMMR (Zhang et al. 2024), and DUSTED (Zhu et al. 2025). We adopt *supervised* metrics, i.e., ARI (Vinh, Epps, and Bailey 2009), NMI (Pfitzner, Leibbrandt, and Powers 2009) and Jaccard similarity coefficient (Levandowsky and Winter 1971), and *unsupervised* metrics, i.e., Moran's I score (Moran 1950) to evaluate the performance of domain identification. Several downstream analyses are also performed to validate the effectiveness of PREST. Further experimental configurations and implementation details are provided in the Appendix.

Slice	Metric	SCANPY GB18	SpaGCN NM21	STAGATE NC22	DeepST NAR22	stLearn NC23	SCGDL BIB23	GraphST NC23	stMMR GS24	DUSTED AAAI25	PREST (Ours)
151507	ARI	0.20	0.39	0.54	0.46	0.49	0.49	0.48	0.58	0.46	**0.75**
	NMI	0.21	0.49	0.66	0.64	0.64	0.55	0.64	0.72	0.65	**0.78**
151508	ARI	0.15	0.33	0.49	0.46	0.31	0.34	0.49	0.52	0.45	**0.56**
	NMI	0.21	0.43	0.63	0.61	0.53	0.44	0.54	0.64	0.64	**0.66**
151509	ARI	0.19	0.35	0.53	0.48	0.45	0.32	0.52	0.58	0.45	**0.59**
	NMI	0.27	0.51	0.66	0.62	0.62	0.48	0.64	0.67	0.64	**0.69**
151669	ARI	0.10	0.23	0.35	0.42	0.32	0.24	0.48	0.47	0.47	**0.66**
	NMI	0.16	0.36	0.58	0.53	0.49	0.38	0.59	0.55	0.59	**0.64**
151670	ARI	0.09	0.33	0.32	0.36	0.23	0.26	0.46	0.49	0.29	**0.57**
	NMI	0.16	0.43	0.53	0.54	0.41	0.38	**0.68**	0.56	0.51	0.60
151671	ARI	0.12	0.42	0.51	0.48	0.39	0.31	0.61	0.67	0.50	**0.81**
	NMI	0.24	0.53	0.65	0.63	0.54	0.41	0.72	0.71	0.65	**0.79**
151672	ARI	0.12	0.52	0.54	0.44	0.63	0.34	0.34	0.60	0.57	**0.78**
	NMI	0.23	0.60	0.66	0.59	0.61	0.47	0.46	0.71	0.67	**0.82**
151673	ARI	0.20	0.40	0.45	0.57	0.30	0.34	**0.63**	0.60	0.57	0.59
	NMI	0.29	0.55	0.63	0.71	0.49	0.42	**0.74**	0.68	0.70	0.68
151674	ARI	0.22	0.31	0.48	0.48	0.38	0.27	0.43	0.51	0.46	**0.56**
	NMI	0.31	0.46	0.58	0.60	0.54	0.38	0.61	0.63	0.63	**0.69**
151675	ARI	0.23	0.27	0.36	0.52	0.38	0.30	0.55	0.56	0.56	**0.58**
	NMI	0.32	0.41	0.50	0.65	0.56	0.41	0.62	0.66	0.65	**0.68**
151676	ARI	0.22	0.31	0.49	0.50	0.40	0.29	**0.61**	0.54	0.52	0.56
	NMI	0.31	0.48	0.63	0.60	0.56	0.42	0.66	0.65	0.67	**0.69**

Table 1: Domain identification performance of different methods on the 11 slices of the DLPFC dataset. The **bold** and underlined values indicate the best and the runner-up results, respectively.

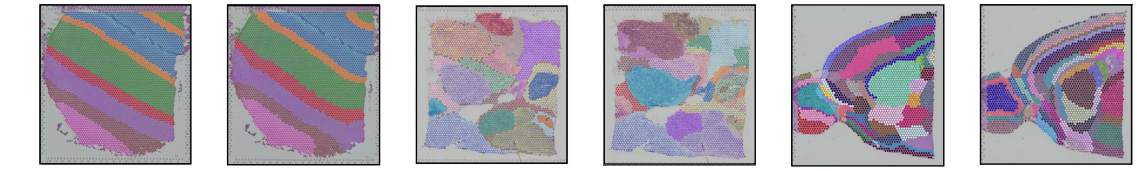


Figure 2: Spatial visualization of identification results for the DLPFC (151507), HBC, and MAB datasets.

Experimental Results

Quantitative Analysis. To evaluate the domain identification performance of our PREST, we conduct quantitative comparisons with several competitive baselines on the DLPFC dataset, as summarized in Table 1. From the results, we observe that spatially-aware methods consistently outperform non-spatial counterparts, highlighting the importance of incorporating spatial context in domain identification. Notably, PREST achieves superior performance on most tissue slices, demonstrating its effectiveness and stability. We attribute this to PREST's capability to effectively integrate spatial information while leveraging uncertainty to dynamically guide and supervise different spots, thereby preventing overfitting to low-quality data. The box plot of the results across all slices of the DLPFC dataset for various methods can be found in Figure 8 of the Appendix. Additional experimental results on HBC and MAB datasets are presented in Table 5 and the comparisons of Moran's I score and Jaccard similarity coefficient across these datasets are shown in Table 6 of the Appendix, which further demonstrate the superiority of our proposed PREST.

Visualization. We visualize the domain identification results of our PREST alongside manually annotated tissues across all three datasets, with slice 151507 from the DLPFC dataset used as an example. As shown in Figure 2, our PREST performs exceptionally well in identifying spatial domains. By comparing the manual annotations with the results detected by PREST, it is clear that our method can accurately delineate spatial domains and effectively capture the boundaries between distinct tissue layers. Furthermore, for a more comprehensive comparison, including domain identification and UMAP projections of other methods, the results are displayed in Figure 7 of the Appendix, which further supports the superiority of our PREST in domain identification.

Ablation Study and Sensitivity Analysis

Effectiveness of Loss Function. We conduct an ablation study to verify the effectiveness of each module we proposed, and the results are presented in Table 2. The term "PREST w/o $\mathcal{L}_{\text{DSL}}$" denotes the model that excludes the $\mathcal{L}_{\text{DSL}}$ loss, while other variants follow similar naming conventions. From the results, we observe that removing any module leads to a decrease in performance. For slice 151507

Model	Metric	DLPFC (151507)	HBC	MAB
PREST w/o $\mathcal{L}_{\text{DSL}}$	ARI	0.62	0.63	0.37
	NMI	0.70	0.67	0.63
PREST w/o $\mathcal{L}_{\text{UZR}}$	ARI	0.62	0.68	0.40
	NMI	0.72	0.69	0.69
PREST w/o $\mathcal{L}_{\text{USR}}$	ARI	0.60	0.66	0.40
	NMI	0.69	0.68	0.66
PREST (Ours)	ARI	**0.75**	**0.72**	**0.43**
	NMI	**0.78**	**0.73**	**0.72**

Table 2: Ablation study across all three datasets.

Dataset	Metric	M_1	M_2	M_3	M_4	Ours
DLPFC (151507)	ARI	0.61	0.63	0.72	0.54	**0.75**
	NMI	0.71	0.73	0.74	0.69	**0.78**
HBC	ARI	0.60	0.63	0.66	0.61	**0.72**
	NMI	0.67	0.70	0.71	0.66	**0.73**
MAB	ARI	0.32	0.30	0.35	0.30	**0.43**
	NMI	0.64	0.63	0.66	0.59	**0.72**

Table 3: Impact of different representation fusion strategies.

of the DLPFC, removing either $\mathcal{L}_{\text{DSL}}$ or $\mathcal{L}_{\text{USR}}$ has a considerable impact, while for HBC and MAB, removing $\mathcal{L}_{\text{DSL}}$ shows a more significant effect. It indicates that uncertainty modeling plays a crucial role in our model.

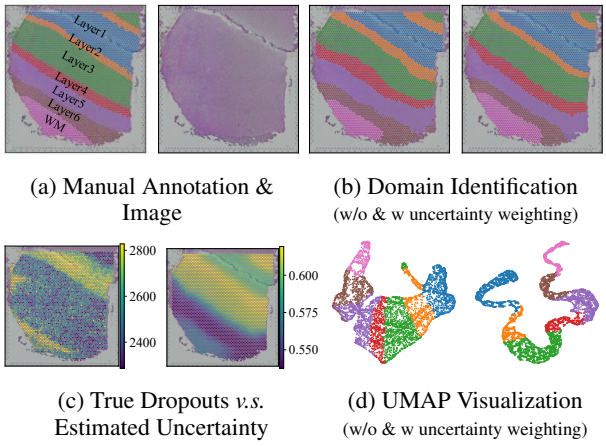
Influence of Representation Fusion Mechanism. To investigate the impact of different representation fusion strategies, we compare four variants: M_1: Additive fusion; M_2: Learnable parameter fusion; M_3: Attention score weighting; M_4: Cross-attention without adjacency information. As shown in Table 3, we observe that M_4 performs worse than the other fusion strategies. This may be because the target spot attends to all other spots indiscriminately, which can easily introduce noise and lead to semantic ambiguity. In contrast, our structure-guided cross-attention fusion effectively leverages spatial topology by attending only to relevant neighboring spots, achieving state-of-the-art performance and underscoring the critical importance of spatial information.

Sensitivity Analysis. We assess hyperparameter sensitivity for α and β in Eq. (9) across all datasets, and full results in Figure 9 (Appendix) consistently confirm model robustness.

Uncertainty and Noise Resistance Analysis

We further investigate the role of estimated uncertainty in our PREST and the model's robustness against noise.

Influence of Uncertainy. To assess the impact of evidence-aware uncertainty, we conduct an analysis on slice 151507 of DLPFC, comparing domain identification results without and with uncertainty weighting in Eqs. (7)-(8). As shown in Figure 3b, the uncertainty-weighted model (right) yields results more consistent with the ground-truth annotations (Figure 3a); for example, White Matter(WM) (pink), Layer 2 (orange), and Layer 4 (red) more closely match the true region shapes. The UMAP plots in Figure 3d further show that our model (right) learns more coherent representations for spatially distant but categorically identical spots (e.g., Layer



(a) Manual Annotation & Image

(b) Domain Identification (w/o & w uncertainty weighting)

(c) True Dropouts *v.s.* Estimated Uncertainty

(d) UMAP Visualization (w/o & w uncertainty weighting)

Figure 3: Visual comparison without (w/o) $v.s.$ with (w) uncertainty weighting on slice 151507 of DLPFC.

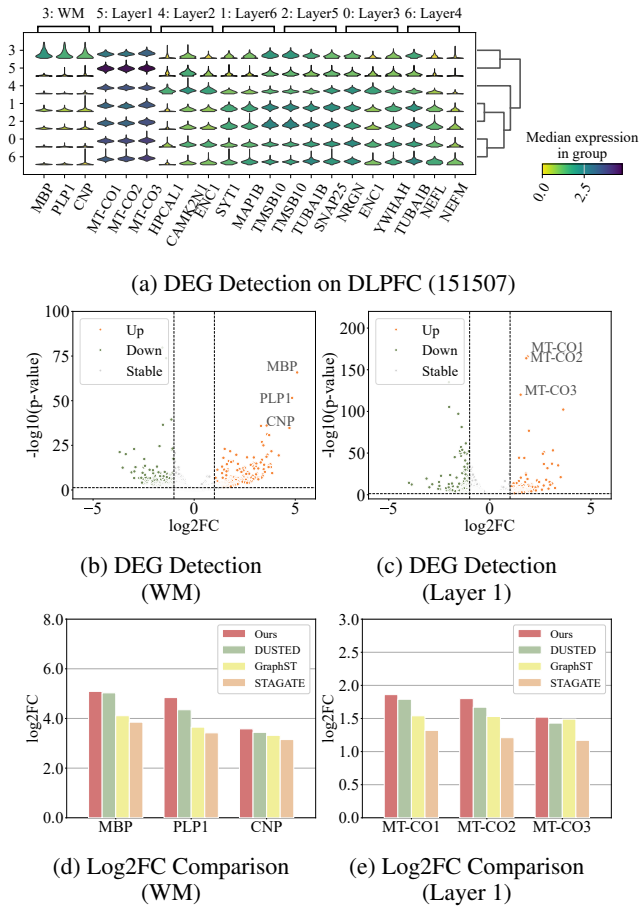
Method	Raw	Gaussian Noise		Dropout	
		$\epsilon = 1$	$\epsilon = 5$	$p = 0.3$	$p = 0.6$
GraphST	0.41	0.21	0.15	0.32	0.15
DUSTED	0.40	0.16	0.13	0.30	0.14
PREST (Ours)	**0.43**	**0.30**	**0.24**	**0.36**	**0.26**

Table 4: Performance comparison of domain identification on MAB via ARI under different expression noise (ϵ: variance of Gaussian noise; p: dropout probability).

2 and Layer 3), while the non-weighted version fails to do so (as seen in the left subplot of Figure 3d, where the green and orange regions are dispersed). It demonstrates that our uncertainty weighting scheme enhances the model's ability to capture the underlying cluster structure.

Uncertainty Attribution. To understand the effectiveness of uncertainty weighting, we analyze the sources of uncertainty. As shown in Figure 3c (right), high-uncertainty spots are mainly located in Layers 1-3 and WM. The left panel in Figure 3c shows dropout counts per spot, with high-dropout regions concentrated in Layer 1 and WM. These observations suggest: (i) high uncertainty in Layer 1 and WM may arise from high dropout rates; while (ii) that in Layers 2-3 may result from discontinuous spatial distributions; the strong influence of spatial information on slice 151507 (see Table 2) contributes higher uncertainty in Layers 2-3. Therefore, for spots in these regions, using raw data for self-supervised learning risks overfitting; down-weighting their supervision promotes more discriminative representations.

Robustness Analysis against Noise. To illustrate model robustness, we artificially add various types of noise to the original dataset and compare our proposed PREST with competitive baselines. We consider Gaussian noise $\mathcal{N}(0, \epsilon)$ (simulate inaccurate measure) with $\epsilon = 1, 5$ and random masking (simulate dropout events) with probability $p = 0.3, 0.6$. Table 4 reports the results on the MAB dataset under these noise conditions. It shows that our PREST consistently outperforms GraphST and DUSTED across different



(a) DEG Detection on DLPFC (151507)

(b) DEG Detection (WM)

(c) DEG Detection (Layer 1)

(d) Log2FC Comparison (WM)

(e) Log2FC Comparison (Layer 1)

Figure 4: DEG detection on slice 151507 of DLPFC.

Gaussian noise levels and dropout rates, highlighting its superior ability to mitigate noise effects. This robustness likely stems from our uncertainty estimation mechanism, which dynamically adjusts supervision strength for each spot, preventing overfitting to low-quality data and effectively compensating for missing data patterns typical in spatial transcriptomics. Additional results on other datasets are provided in Table 7 of the Appendix.

Downstream Analysis

We conduct differentially expressed genes (DEGs) detection, gene expression imputation, and pathway enrichment analysis (in Figure 12 of the Appendix) to highlighting the superior biological interpretability of our PREST.

DEG Analysis. DEGs or marker genes show statistically significant expression differences across biological conditions and serve as key indicators of tissue structure and regulation. Based on the identified domains, we perform DEG detection on all three datasets. Figure 4a presents the top three DEGs for each cluster in slice 151507 of DLPFC. For example, MBP, a gene linked to myelin formation and oligodendrocyte differentiation, is highly expressed in the WM layer, aligning with prior knowledge that oligodendrocytes are enriched in WM (Emery 2010), thus validating our PREST biologically. We further quantify DEGs in WM and Layer 1 using volcano plots (Figures 4b-4c), showing low p-values and high log2FC values, indicating significant expression differences. Additionally, log2FC comparisons are also conducted across methods (Figures 4d-4e), which reveal that our PREST consistently achieves higher log2FC in both clusters than SCANPY, GraphST, and DUSTED, suggesting better inter-cluster separation. The detection results for other datasets are shown in Figure 10 of the Appendix.

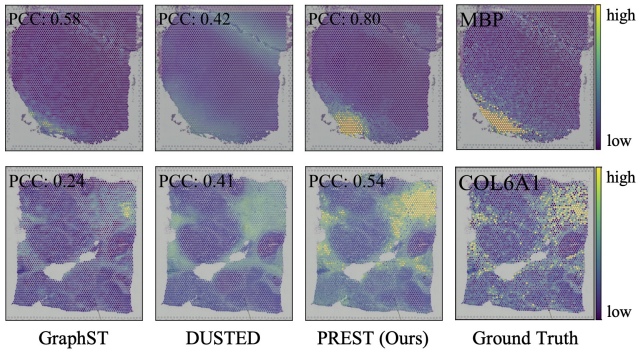


Figure 5: Visual comparison of expression imputation (MAB and COL6A1) for DLPFC and HBC, respectively.

Gene Expression Imputation. Due to the low capture efficiency, ST data is often significantly sparse and noisy. which poses challenges for downstream analysis. The goal of imputation is to recover the biological signal by imputing the dropouts. To validate the biologically meaningful imputation of PREST, we introduce artificial dropouts by randomly masking gene expression with a dropout rate of 0.3, and then perform reconstruction based on the remaining data. Imputation accuracy is quantified via Pearson correlation coefficient (PCC) between imputed and ground-truth expression. As demonstrated in Figure 5, our PREST achieves superior performance in both gene expression restoration and spatial pattern preservation compared to baselines (e.g., +0.22 PCC for MBP gene imputation). The results for other datasets and marker genes can be found in Figure 11 in the Appendix.

Conclusion

In this study, we propose a prototype-based evidence-aware framework PREST for spatial transcriptomics analysis. By integrating multi-scale representation learning, structure-guided cross-attention fusion, and uncertainty-aware modules, PREST facilitates robust uncertainty estimation and clustering-friendly latent representations. Extensive experiments across multiple benchmark datasets demonstrate that our proposed PREST consistently outperforms existing approaches in accuracy of domain identification. Additional experimental analysis illustrates the interpretability of uncertainty estimation and the robustness of our PREST against noise. Various downstream analyses further reveal the effectiveness of our PREST in biological discovery.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 62306014, 12501344 and 12131001, Postdoctoral Fellowship Program (Grade A) of CPSF under Grant BX20250376 and BX20240239, China Postdoctoral Science Foundation under Grant 2024M762201, Sichuan Science and Technology Program under Grant 2025ZNSFSC1506 and 2025ZNSFSC0808, Sichuan University Interdisciplinary Innovation Fund, the Fundamental Research Funds for the Central Universities, LPMC, KLMDASR, and the top-notched student program of Sichuan University.

References

Asp, M.; Bergenstråhle, J.; and Lundeberg, J. 2020. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*, 42(10): 1900221.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Buache, E.; Etique, N.; Alpy, F.; Stoll, I.; Muckensturm, M.; Reina-San-Martin, B.; Chenard, M.; Tomasetto, C.; and Rio, M. 2011. Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. *Oncogene*, 30(29): 3261–3273.

Chen, A.; Liao, S.; Cheng, M.; Ma, K.; Wu, L.; Lai, Y.; Qiu, X.; Yang, J.; Xu, J.; Hao, S.; et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 185(10): 1777–1792.

Codeluppi, S.; Borm, L. E.; Zeisel, A.; La Manno, G.; van Lunteren, J. A.; Svensson, C. I.; and Linnarsson, S. 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods*, 15(11): 932–935.

Dong, H. W. 2008. *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse.* John Wiley & Sons Inc.

Dong, K.; and Zhang, S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1): 1739.

Du, L.; Kang, J.; Hou, Y.; Sun, H.-X.; and Zhang, B. 2024. SpotGF: Denoising spatially resolved transcriptomics data using an optimal transport-based gene filtering algorithm. *Cell Systems*, 15(10): 969–981.

Emery, B. 2010. Regulation of oligodendrocyte differentiation and myelination. *Science*, 330(6005): 779–782.

Gong, C.; Zou, J.; Zhang, M.; Zhang, J.; Xu, S.; Zhu, S.; Yang, M.; Li, D.; Wang, Y.; Shi, J.; et al. 2019. Upregulation of MGP by HOXC8 promotes the proliferation, migration, and EMT processes of triple-negative breast cancer. *Molecular Carcinogenesis*, 58(10): 1863–1875.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Holdener, C.; and De Vlaminck, I. 2025. Smoothie: efficient inference of spatial co-expression networks from denoised spatial transcriptomics data. *bioRxiv*.

Hu, J.; Li, X.; Coleman, K.; Schroeder, A.; Ma, N.; Irwin, D. J.; Lee, E. B.; Shinohara, R. T.; and Li, M. 2021. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods*, 18(11): 1342–1351.

Ji, A. L.; Rubin, A. J.; Thrane, K.; Jiang, S.; Reynolds, D. L.; Meyers, R. M.; Guo, M. G.; George, B. M.; Mollbrink, A.; Bergenstråhle, J.; et al. 2020. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 182(2): 497–514.

Jøsang, A. 2016. *Subjective Logic*, volume 4. Springer.

Ju, W.; Fang, Z.; Gu, Y.; Liu, Z.; Long, Q.; Qiao, Z.; Qin, Y.; Shen, J.; Sun, F.; Xiao, Z.; et al. 2024a. A comprehensive survey on deep graph representation learning. *Neural Networks*, 106207.

Ju, W.; Gu, Y.; Chen, B.; Sun, G.; Qin, Y.; Liu, X.; Luo, X.; and Zhang, M. 2023. Glcc: A general framework for graph-level clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4391–4399.

Ju, W.; Yi, S.; Wang, Y.; Long, Q.; Luo, J.; Xiao, Z.; and Zhang, M. 2024b. A survey of data-efficient graph learning. In *International Joint Conference on Artificial Intelligence*, 8104–8113.

Ju, W.; Yi, S.; Wang, Y.; Xiao, Z.; Mao, Z.; Li, H.; Gu, Y.; Qin, Y.; Yin, N.; Wang, S.; Liu, X.; Yu, P. S.; and Zhang, M. 2025. A Survey of Graph Neural Networks in Real World: Imbalance, Noise, Privacy and OOD Challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–20.

Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Levandowsky, M.; and Winter, D. 1971. Distance between sets. *Nature*, 234(5323): 34–35.

Li, H.; Song, J.; Gao, L.; Zhu, X.; and Shen, H. 2023. Prototype-based aleatoric uncertainty quantification for cross-modal retrieval. *Advances in Neural Information Processing Systems*, 36: 24564–24585.

Li, J.; Chen, S.; Pan, X.; Yuan, Y.; and Shen, H.-B. 2022. Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science*, 2(6): 399–408.

Likas, A.; Vlassis, N.; and Verbeek, J. J. 2003. The global k-means clustering algorithm. *Pattern Recognition*, 36(2): 451–461.

Liu, T.; Fang, Z.-Y.; Li, X.; Zhang, L.-N.; Cao, D.-S.; and Yin, M.-Z. 2023. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Briefings in Bioinformatics*, 24(3): bbad146.

Long, Y.; Ang, K. S.; Li, M.; Chong, K. L. K.; Sethi, R.; Zhong, C.; Xu, H.; Ong, Z.; Sachaphibulkij, K.; Chen, A.; et al. 2023. Spatially informed clustering, integration, and

deconvolution of spatial transcriptomics with GraphST. *Nature Communications*, 14(1): 1155.

Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Lubeck, E.; Coskun, A. F.; Zhiyentayev, T.; Ahmad, M.; and Cai, L. 2014. Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods*, 11(4): 360–361.

Malinin, A.; and Gales, M. 2018. Predictive uncertainty estimation via prior networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 7047–7058.

Maynard, K. R.; Collado-Torres, L.; Weber, L. M.; Uytingco, C.; Barry, B. K.; Williams, S. R.; Catallini, J. L.; Tran, M. N.; Besich, Z.; Tippani, M.; et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, 24(3): 425–436.

Moffitt, J. R.; Bambah-Mukku, D.; Eichhorn, S. W.; Vaughn, E.; Shekhar, K.; Perez, J. D.; Rubinstein, N. D.; Hao, J.; Regev, A.; Dulac, C.; et al. 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, 362(6416): eaau5324.

Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23.

Pfitzner, D.; Leibbrandt, R.; and Powers, D. 2009. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19(3): 361–394.

Pham, D.; Tan, X.; Balderson, B.; Xu, J.; Grice, L. F.; Yoon, S.; Willis, E. F.; Tran, M.; Lam, P. Y.; Raghubar, A.; et al. 2023. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature Communications*, 14(1): 7739.

Rao, A.; Barkley, D.; França, G. S.; and Yanai, I. 2021. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871): 211–220.

Rodriques, S. G.; Stickels, R. R.; Goeva, A.; Martin, C. A.; Murray, E.; Vanderburg, C. R.; Welch, J.; Chen, L. M.; Chen, F.; and Macosko, E. Z. 2019. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434): 1463–1467.

Shah, S.; Lubeck, E.; Zhou, W.; and Cai, L. 2016. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 92(2): 342–357.

Su, J.; Reynier, J.-B.; Fu, X.; Zhong, G.; Jiang, J.; Escalante, R. S.; Wang, Y.; Aparicio, L.; Izar, B.; Knowles, D. A.; et al. 2023. Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. *Genome Biology*, 24(1): 291.

Tang, Z.; Li, Z.; Hou, T.; Zhang, T.; Yang, B.; Su, J.; and Song, Q. 2023. SiGra: single-cell spatial elucidation through an image-augmented graph transformer. *Nature Communications*, 14(1): 5618.

Tu, W.; Zhou, S.; Liu, X.; Guo, X.; Cai, Z.; Zhu, E.; and Cheng, J. 2021. Deep Fusion Clustering Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 9978–9987.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2018. Graph attention networks. In *International Conference on Learning Representations*.

Vinh, N. X.; Epps, J.; and Bailey, J. 2009. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th International Conference on Machine Learning*, 1073–1080.

Wang, Y.; Song, B.; Wang, S.; Chen, M.; Xie, Y.; Xiao, G.; Wang, L.; and Wang, T. 2022. Sprod for de-noising spatially resolved transcriptomics data based on position and image information. *Nature Methods*, 19(8): 950–958.

Williams, C. G.; Lee, H. J.; Asatsuma, T.; Vento-Tormo, R.; and Haque, A. 2022. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, 14(1): 68.

Wolf, F. A.; Angerer, P.; and Theis, F. J. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19: 1–5.

Xu, C.; Jin, X.; Wei, S.; Wang, P.; Luo, M.; Xu, Z.; Yang, W.; Cai, Y.; Xiao, L.; Lin, X.; et al. 2022. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, 50(22): e131–e131.

Yu, Z.; Lu, Y.; Wang, Y.; Tang, F.; Wong, K.-C.; and Li, X. 2022. ZINB-based graph embedding autoencoder for single-cell rna-seq interpretations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4671–4679.

Zhang, D.; Yu, N.; Yuan, Z.; Li, W.; Sun, X.; Zou, Q.; Li, X.; Liu, Z.; Zhang, W.; and Gao, R. 2024. stMMR: accurate and robust spatial domain identification from spatially resolved transcriptomics with multimodal feature representation. *GigaScience*, 13: giae089.

Zhu, J.; Li, Y.; Tang, Z.; and Chang, C. 2025. DUSTED: Dual-Attention Enhanced Spatial Transcriptomics Denoiser. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 1219–1227.

Zhu, Y.; He, X.; Tang, C.; Liu, X.; Liu, Y.; and He, K. 2024. Multi-view Adaptive Fusion Network for Spatially Resolved Transcriptomics Data Clustering. *IEEE Transactions on Knowledge and Data Engineering*.