Pattern Recognition 180 (2026) 114585

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

stHGNN: Deciphering spatial transcriptomics data via dual hypergraph learning enhancement☆

Jinghong Tang [a,1], Lezhi Chen [b,1], Siyu Yi [c,*], Wei Liu [c], Mingyang Li [c], Yifan Wang [d], Ziyue Qiao [e], Bin Guo [b], Xianggen Liu [a], Jiancheng Lv [a], Wei Ju [f]

[a] College of Computer Science, Sichuan University, Chengdu, 610065, China
[b] College of Electrical Engineering, Sichuan University, Chengdu, 610065, China
[c] School of Mathematics, Sichuan University, Chengdu, 610065, China
[d] School of Artificial Intelligence and Data Science, University of International Business and Economics, Beijing, 100029, China
[e] School of Computing and Information Technology, Great Bay University, Dongguan, 523000, China
[f] School of Artificial Intelligence, Sichuan University, Chengdu, 610065, China

ARTICLE INFO

Keywords:
Spatial Transcriptomics
Hypergraph learning
Spatial domain identification
Clustering

ABSTRACT

Spatial Transcriptomics (ST) reveals gene expression patterns within the context of tissue microenvironment by in situ measuring gene expression at native spatial locations, enabling the deciphering of spatial domains. However, existing methods often simplify cellular dependencies to pairwise k-nearest neighbor graphs, overlooking the contextual and higher-order dependencies of cellular networks. To address this issue, we propose stHGNN, a dual-view hypergraph enhanced framework for identifying spatial domains in ST data. stHGNN captures contextual and higher-order cellular dependencies using hypergraphs, adopts hypergraph representation learning from both spatial and gene expression views, and then facilitates cross-view knowledge transfer through a cross-attention mechanism and self-correlation reorganization. Additionally, a multi-view self-supervised clustering strategy is adopted to learn clustering-friendly representations, while hypergraph smoothing is applied to enhance consistency with biological priors. Finally, a zero-inflated negative binomial reconstruction is adopted to handle the inherent over-dispersion and dropouts in ST data. Comprehensive experiments and downstream analysis demonstrate the effectiveness of our stHGNN.

1. Introduction

Spatial Transcriptomics (ST) measures gene expression while preserving the spatial location of cells within tissues, thus providing a multimodal description of tissue spots through expression features and spatial coordinates, facilitating the investigation of previously challenging or impossible biological questions, such as the inherent heterogeneity within classical cell types, dynamic changes in cellular processes, reliable cell–cell communications, and the biological pathways underlying cellular differentiation [1]. Existing ST data are mainly generated by two distinct protocols: barcode-based methods, such as 10x Visium [2] and SLIDE-seq [3]; and image-based methods, such as MERFISH [4], osmFISH [5], and seqFISH [6]. Unlike single-cell resolution sequencing such as scRNA-seq, the resolution of these ST platforms is typically at the spot level (ranging from a few to tens of cells). Despite this limitation, ST has found widespread applications across biological research [7], generating vast data and creating an urgent and growing need for advanced computational methods.

Among the key analytical tasks in ST data, spatial domain identification, which aims to reveal the spatial distribution of different cellular populations or subpopulations, and the architectural features of pathological tissues, has garnered significant attention. It is also considered a foundation for various downstream tasks [8], such as cell type annotation, differential gene expression analysis, and trajectory inference [9]. Early approaches primarily adapted traditional unsupervised algorithms, including PCA [10], with clustering methods, such as K-means [11] and mclust [12]. However, these methods treat cells as independent data points, disregarding the rich biological priors and cellular dependencies shared across the modalities, leading to the great loss of valuable information. Recently, Graph Neural Networks (GNNs) have been widely employed to capture cellular dependencies from the dual modalities of expression and space, which has advanced the

☆ This article is part of a Special issue entitled: 'PR_Biomedical Data' published in Pattern Recognition.

* Corresponding author.
E-mail address: siyuyi@scu.edu.cn (S. Yi).

[1] Equal contribution.

https://doi.org/10.1016/j.patcog.2026.114585
Received 18 April 2026; Received in revised form 9 July 2026; Accepted 29 July 2026
Available online 4 August 2026

development of spatial domain identification algorithms. GraphST [13] combines GNN with contrastive learning, enabling clustering, integration, and deconvolution of ST data, while stHGC [14] builds on a hybrid neighbor graph constructed by integrating different similarity metrics. Spotscape [15] introduces the Similarity Telescope module to capture global relationships, facilitating effective multi-slice integration.

The aforementioned graph-based methods have demonstrated superiority in the spatial domain identification task, however, there are still several intrinsic limitations to be addressed. **(1)** Most existing methods **simplify the complex intercellular dependencies to pairwise k-nearest neighbor graphs.** Some of the methods construct graph structures simply based on spatial coordinates [15], overlooking the fact that spots of the same type or subtype are not always spatially adjacent, while others attempt to alleviate structural information loss by incorporating multi-view information [14,16], which still fail to capture the high-order cellular dependencies. In other words, these existing graph-based strategies mainly encode whether two spots are connected, but they do not explicitly describe whether a group of spots jointly forms a shared cellular microenvironment or functional context. Therefore, accurately modeling such higher-order, multi-participant dependencies is essential for robust spatial domain identification. **(2) Spatial domain identification requires representations that are jointly guided by biological plausibility and clustering objectives.** On the one hand, cells participating in the same functional module tend to exhibit coherent expression profiles due to shared regulatory programs and microenvironments [17], which implies a biological smoothness prior that should be respected in the learned embeddings. On the other hand, for clustering to succeed, the learned representation should actively promote both intra-cluster compactness and inter-cluster separability [18], as opposed to merely preserving the input similarities [16]. The joint constraint could ensure that the learned representation is not only consistent with tissue-level biological organization but also inherently optimized to be friendly to clustering.

To address these issues, we propose Dual-view HyperGraph Neural Network for domain identification in ST data (stHGNN). We begin by constructing hypergraphs corresponding to the information from both spatial and feature views to represent the higher-order cellular dependencies. Subsequently, two hypergraph neural networks are used as encoders to learn low-dimensional representations. We then fuse dual-view embeddings through a cross-attention mechanism and apply a global self-correlation strategy. To learn clustering-friendly representations, we adopt the multi-view self-supervised clustering loss, while the hypergraph smoothing loss is used to ensure biological consistency. Finally, to handle the high sparsity and over-dispersion of ST data, we utilize the zero-inflated negative binomial (ZINB) decoder, enabling self-supervised training. Overall, our method features the following contributions:

- We introduce stHGNN, a hypergraph-enhanced framework that models higher-order cellular dependencies, preserving group-wise interactions among spots, thereby better reflecting the intrinsic organization of tissue.
- We jointly optimize representation learning and clustering, ensuring the representation is not only biologically plausible but also clustering-friendly, with enhanced intra-cluster compactness and inter-cluster separability.
- Extensive experiments and downstream analysis across three benchmark ST datasets spanning diverse species and tissue types show that stHGNN consistently outperforms baseline methods in deciphering spatial domains.

2. Related work

2.1. Spatial domain identification

Spatial domain identification is one of the core tasks in ST data analysis, which facilitates the revelation of cell heterogeneity. Traditional methods, such as K-means [11], do not fully utilize the crucial spatial location information provided by ST data, leading to poor spatial domain identification results. To address this, some methods construct spatial proximity graphs based on coordinates and employ Graph Neural Networks (GNNs) to learn representations. STAGATE [19] prunes the spatial proximity graph using gene expression information and introduces an autoencoder framework for self-supervised model training, while GraphST [13] adopts a contrastive learning strategy to enhance the model's learning capability. To further alleviate the high sparsity and over-dispersion of ST data, DUSTED [20] utilizes a dual-attention enhanced graph autoencoder, while Spotscape [15] captures global relationships, facilitating effective multi-slice integration. Additionally, Recent ST methods further explore evidence-aware integration [21], gene semantic priors [22], and modality conflict modulation [23] for spatial domain identification. scGTN [24] further explores siamese graph transformer modeling for scRNA-seq clustering. While these methods have shown strong performance, they do not consider the complex structure of the cellular network, simplifying the spot relationships into simple pairwise interactions. Our proposed stHGNN utilizes hypergraphs to more explicitly model high-order cellular dependencies, thereby guiding spatial domain identification in ST data.

2.2. Hypergraph representation learning

Hypergraphs have emerged as a powerful generalization of traditional graphs, which are limited to pairwise interactions, by allowing a single hyperedge to connect any number of nodes. By capturing high-order dependencies, hypergraph-based models provide a more expressive framework for analyzing complex, contextual relationships in diverse fields. This capability makes hypergraphs particularly effective for modeling higher-order and group-wise relationships, such as those found in complex systems or networks where multiple entities interact simultaneously. The development of hypergraph convolutional networks, including pioneering works [25] and subsequent extensions [26], has enabled effective representation learning through the aggregation of higher-order features. Hypergraph representation learning has proven successful across not only various tasks, including clustering [27], classification [28], and link prediction [29], but also applications in specific areas, such as cell localization [30], medical image segmentation [31], recommender systems [32] and cancer survival analysis [33]. In this paper, we introduce hypergraphs into domain identification task, designing a dual hypergraph-enhanced representation learning framework.

2.3. Multi-view and cross-modal representation learning

Multi-view and multimodal learning integrates heterogeneous information while reducing modality gaps and noise. For multimodal clustering, DRLMMC [34] reduces noise via task-aware decoupling. ScRL [35], VVI-ReID [36], and HPL [37] respectively exploit shape cues, language prompts, and hierarchical prompts to bridge modality gaps. Video and multi-view studies further explore dual-space structures [38], spatial–temporal cues [39], gait priors [40], and multiview interaction [41]. Together, these studies highlight cross-modal complementarity and structured relation modeling. However, their data objects, noise sources, and structural priors differ from ST data. ST data instead require modeling expression sparsity, spatial continuity, and higher-order cellular dependencies. This makes direct adaptation to ST challenging. Therefore, our stHGNN integrates multi-view learning with biological priors through ST-tailored hypergraphs, fusion modules, and training objectives.

3. Methodology

3.1. Problem definition and data preprocessing

ST provides a gene expression matrix and a spatial coordinate matrix. Due to limitations of sequencing technology and unavoidable

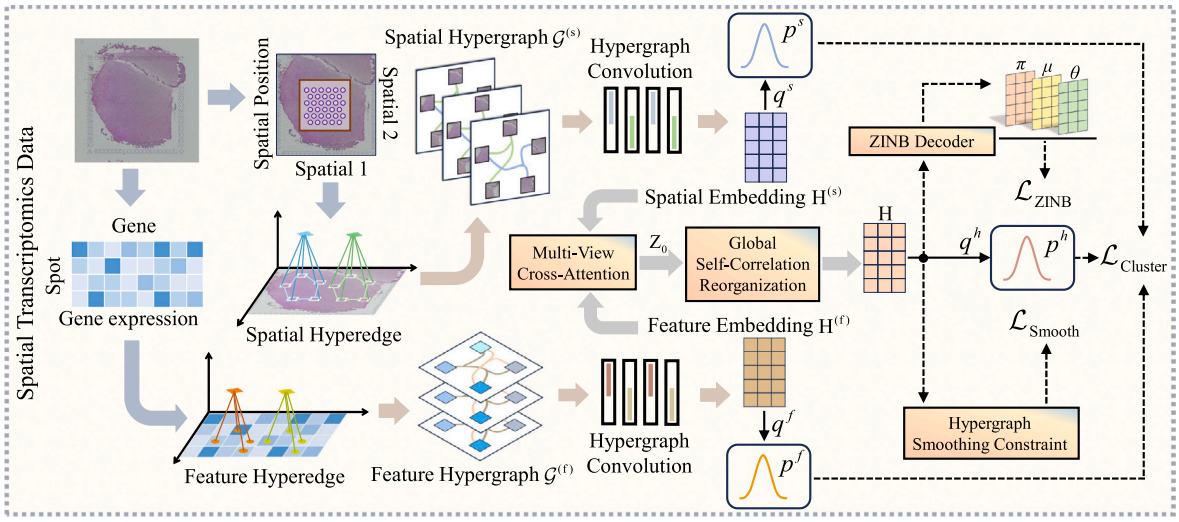

Fig. 1. Overall framework of the proposed stHGNN. Given a ST dataset consisting of gene expression and spatial coordinates, our stHGNN first constructs dual hypergraphs from feature and spatial views to capture higher-order cellular dependencies. Then, it learns view-specific representations via hypergraph neural networks and fuses them through cross-attention and global self-correlation reorganization. Finally, it produces clustering-friendly embeddings that respect both biological smoothness and structural coherence for spatial domain identification.

errors during sample preparation, the gene expression matrix based on UMI counts contains noise. To mitigate technical noise in ST data, we preprocess the raw gene expression matrix $\mathbf{F}_{obs} \in \mathbb{R}^{N\times D_{obs}}$. First, we retain genes that are expressed in at least 100 cells to ensure that remaining genes have sufficient detection rates. Next, we use Seurat_v3 [42] to select the top 3000 highly variable genes (HVGs), which are often associated with key biological features such as cell identity, function, and state. Focusing on HVGs allows us to ignore irrelevant background and directly analyze the core biological structure of the data. Finally, to eliminate differences in sequencing depth between cells, we perform the normalization:

$$\mathbf{X}_{ij} = \frac{\mathbf{F}_{ij}}{\sum_{j=1}^{D}\mathbf{F}_{ij}} \times 10000, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{N\times D}$ is the preprocessed gene expression matrix, and $\mathbf{F}$ is $\mathbf{F}_{obs}$ after cell filtering and HVGs selection. After preprocessing, we obtain a gene expression matrix $\mathbf{X}$ and a spatial coordinate matrix $\mathbf{S}$ from ST data, where each spot's spatial location denotes $\mathbf{s}_i = (x_i, y_i)$. We aim to learn a low-dimensional representation $\mathbf{H} \in \mathbb{R}^{N\times d}$ (with $d \ll D$), and conduct clustering to identify spatial domains. An overview of our proposed framework is described in Fig. 1.

3.2. Dual hypergraph enhanced representing

Existing methods primarily utilize pairwise k-nearest neighbor graphs to capture the expression similarity between spots [13,15], which implicitly introduces an assumption that interactions between spots can be represented as simple pairwise relationships. In ST data, local tissue patterns such as tumor margins, cortical layers, and immune-infiltrated regions are often formed by groups of spatially or transcriptionally related spots, rather than independent spot pairs [43]. To model higher-order dependencies, we introduce hypergraph, which allows edges to connect more than two nodes. Specifically, we derive the incidence matrix of the feature-view hypergraph as:

$$\mathbf{A}^{(\mathrm{f})}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{N}^{(\mathrm{f})}_j, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\mathbf{A}^{(\mathrm{f})} \in \{0,1\}^{N\times N}$ is the incidence matrix. In feature hypergraph construction, we first apply row-wise L2 normalization to the expression matrix $\mathbf{X}$, so the inner product $\mathbf{X}\mathbf{X}^\top$ is equivalent to cosine similarity and reflects expression-pattern correlation. Then, $\mathcal{N}^{(\mathrm{f})}_j = \text{Top-}k_{i\in\{1,\dots,N\}}\left((\mathbf{X}\mathbf{X}^\top)_{ij}\right)$ denotes the set of k most similar nodes to node j based on feature correlation. By contrast, the spatial view focuses on physical proximity between spots in tissue space, which is naturally measured by Euclidean distance.

As a complement, spatial locations of spots are crucial for identifying cellular dependencies. This could be primarily reflected in two aspects: (1) The spatial locations of spots determine the microenvironment in which they reside [17]. Cells within this microenvironment are regulated by both direct contact signals from neighboring cells and soluble diffusion signals within the tissue, which often results in similar phenotypic characteristics. (2) Functionally related cells exist in spatial clusters, and cells that are spatially proximate are more likely to belong to the same cell type or functional subtype. To capture the contextual details of cellular microenvironments, we construct a spatial-view hypergraph by defining a critical distance r, as shown in the following:

$$\mathcal{N}^{(\mathrm{s})}_j = \left\{ i \in \{1,\dots,N\} \,\middle|\, \|\mathbf{s}_i - \mathbf{s}_j\|_2 \le r \right\}, \tag{3}$$

where $\mathcal{N}^{(\mathrm{s})}_j$ denotes the set of spatial neighbors of spot j, and we set $\mathcal{N}^{(\mathrm{s})}_j = \{j\}$ if the set is empty. The incidence matrix $\mathbf{A}^{(\mathrm{s})} \in \{0,1\}^{N\times N}$ of the spatial-view hypergraph is then defined by the following equation:

$$\mathbf{A}^{(\mathrm{s})}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{N}^{(\mathrm{s})}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Graph Convolutional Networks (GCN) [44] have been proven to be highly effective in handling ST data [13,16]. In order to extend GCN to hypergraph structure, we employ two independent hypergraph convolutional networks (HGCN) to encode the feature-view graph $\mathcal{G}^{(\mathrm{f})} = (\mathbf{A}^{(\mathrm{f})}, \mathbf{X})$ and the spatial-view graph $\mathcal{G}^{(\mathrm{s})} = (\mathbf{A}^{(\mathrm{s})}, \mathbf{X})$, respectively. Specifically, the HGCN's forward propagation rule of a single layer can be defined as:

$$\mathbf{H}^{(l+1)} = \sigma\left(\mathbf{D}_v^{-1/2}\mathbf{A}\mathbf{D}_e^{-1}\mathbf{A}^\top\mathbf{D}_v^{-1/2}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right), \tag{5}$$

where $\mathbf{H}^{(l)}$ is the input node feature matrix at the lth layer and $\mathbf{H}^{(0)} = \mathbf{X}$, $\mathbf{H}^{(l+1)}$ denotes the node representations matrix at the $(l+1)$-th layer, σ is a non-linear activation function (e.g., ReLU), $\mathbf{A}$ is the hypergraph incidence matrix, $\mathbf{D}_v$ is the diagonal node degree matrix and $\mathbf{D}_e$ is the diagonal hyperedge degree matrix, $\mathbf{W}^{(l)}$ is the layer-specific trainable weight matrix. Different from pairwise graph convolution,

HGCN propagates information via hyperedges. The multiplication with $\mathbf{A}^\top$ aggregates spot features into hyperedge representations, and the subsequent multiplication with $\mathbf{A}$ sends these hyperedge representations back to spots. In this way, each spot can incorporate information from multiple spots that share the same hyperedge. $\mathbf{D}_v$ and $\mathbf{D}_e$ are used for degree normalization, avoiding biased propagation caused by high-degree spots or large hyperedges. Since the hypergraph structures associated with the feature and spatial views are semantically different, they exhibit distinct adjacency patterns, noise sources, and dependency interpretations. Accordingly, we use independent parameters to learn expression and spatial dependencies separately. The final output representations of two HGNNs are denoted as $\mathbf{Z}^{(\mathrm{f})} = (\mathbf{z}_1^{(\mathrm{f})}, \ldots, \mathbf{z}_N^{(\mathrm{f})})^\top$ and $\mathbf{Z}^{(\mathrm{s})} = (\mathbf{z}_1^{(\mathrm{s})}, \ldots, \mathbf{z}_N^{(\mathrm{s})})^\top$.

3.3. Cross-view knowledge transfer

Cross-Attention Fusion. For spots located at the boundary of a domain, $\mathbf{Z}^{(\mathrm{s})}$ may become ambiguous due to the mixture of heterogeneous information from neighboring cells; if a cell appears outlier in terms of gene expression, $\mathbf{Z}^{(\mathrm{f})}$ may be unreliable. In this case, we fuse the representations from the feature-view and the spatial-view via cross-attention mechanism. Specifically, we use $\mathbf{Z}^{(\mathrm{f})}$ as the query and retrieve relevant information from $\mathbf{Z}^{(\mathrm{s})}$, which can be expressed mathematically as:

$$\begin{aligned}
\mathbf{Q}^{(\mathrm{f})} &= \mathrm{LayerNorm}(\mathbf{Z}^{(\mathrm{f})}),\\
\mathbf{K}^{(\mathrm{s})} &= \mathbf{V}^{(\mathrm{s})} = \mathrm{LayerNorm}(\mathbf{Z}^{(\mathrm{s})}),\\
\mathbf{M}^{(\mathrm{f})} &= \mathrm{MultiHeadAttention}(\mathbf{Q}^{(\mathrm{f})}, \mathbf{K}^{(\mathrm{s})}, \mathbf{V}^{(\mathrm{s})}),\\
\mathbf{Y}^{(\mathrm{f})} &= \mathbf{Z}^{(\mathrm{f})} + \mathbf{M}^{(\mathrm{f})}.
\end{aligned} \tag{6}$$

where $\mathbf{Q}^{(\mathrm{f})}$ denotes the query matrix, $\mathbf{K}^{(\mathrm{s})}$ represents the key matrix and $\mathbf{V}^{(\mathrm{s})}$ represents the corresponding value matrix. The multi-head attention output is denoted as $\mathbf{M}^{(\mathrm{f})}$. LayerNorm($\mathbf{Z}$) denotes layer normalization, which normalizes each row of $\mathbf{Z}$ across its feature dimension and applies learnable scale and shift parameters. $\mathbf{Y}^{(\mathrm{f})}$ is the output of the feature view. Symmetrically, we can obtain $\mathbf{Y}^{(\mathrm{s})}$ using the above formulation. We then average the outputs of the two interactive views to balance complementary information from the spatial and feature views and apply a feed-forward network (FFN) to introduce nonlinearity and enhance representational capacity, formulated as:

$$\mathbf{Z}_0 = \mathrm{LayerNorm}\big(\mathbf{Y} + \mathrm{FFN}(\mathbf{Y})\big), \tag{7}$$

where $\mathbf{Y} = \frac{1}{2}\big(\mathbf{Y}^{(\mathrm{f})} + \mathbf{Y}^{(\mathrm{s})}\big)$, $\mathbf{Z}_0$ is the fused representation. The residual connection $\mathbf{Y} + \mathrm{FFN}(\mathbf{Y})$ preserves the original cross-attended features while allowing the model to learn refined, higher-level abstractions, which is crucial for capturing complex patterns in ST data.

Global Self-Correlation Reorganization. Although $\mathbf{Z}_0$ has already incorporated cross-view information, it still contains redundancies, making further representation enhancement necessary. To capture truly highly similar spots across the global context and enable mutual reinforcement of their features, we introduce a self-correlation module to capture global dependencies, enabling the reorganization and enhancement of features. Specifically, we first compute the self-correlation matrix $\mathbf{C}$ according to $\mathbf{Z}_0$ and utilize it to obtain the soft selection and weighted aggregation of features at all positions, defined as:

$$\mathbf{Z}_1 = \mathbf{C} \cdot \mathbf{Z}_0, \quad \mathbf{C} = \mathrm{softmax}(\mathbf{Z}_0\mathbf{Z}_0^\top). \tag{8}$$

Here, softmax is applied row-wise. Specifically, $\mathbf{C}_{ij}$ represents the global correlation weight from spot i to spot j, and $\sum_{j=1}^{N} \mathbf{C}_{ij} = 1$. Accordingly, $\mathbf{Z}_{1,i} = \sum_{j=1}^{N} \mathbf{C}_{ij}\mathbf{Z}_{0,j}$ means that spot i adaptively aggregates relevant contextual information from all spots. This design aligns with the left-multiplication form $\mathbf{C}\cdot\mathbf{Z}_0$ and assigns each spot an independent global selection distribution. It suppresses weak or noisy correlations while enhancing globally consistent expression patterns, thus reducing redundancy and noise in the fused representation. Finally, we perform an adaptive weighted fusion using learnable parameters α to obtain the globally enhanced representation, defined as:

$$\mathbf{H} = \mathbf{Z}_0 + \alpha \cdot \mathbf{Z}_1. \tag{9}$$

Here, $\mathbf{H}$ is the representation used for downstream analysis.

3.4. Spatial- and expression-view constraints

Hypergraph Smoothing Constraint. Smoothing local features according to cellular dependencies helps identify cellular microenvironments and improves biological consistency. Graph Laplacian regularization is commonly used for this purpose and shows effective performances [45]. We extend this idea to hypergraphs. As this loss encodes the biological prior of spatial organizational continuity, namely that physically adjacent spots in the same local microenvironment tend to have similar tissue states, we apply the smoothing constraint only to the spatial hypergraph. Specifically, we first perform clique expansion on the spatial hypergraph incidence matrix $\mathbf{A}^{(\mathrm{s})}$:

$$\hat{\mathbf{A}}_{ij} = \sum_{k=1}^{N} \mathbf{A}_{ik}^{(\mathrm{s})} \mathbf{A}_{jk}^{(\mathrm{s})}. \tag{10}$$

Then, we construct the hypergraph smoothing loss as:

$$\mathcal{L}_{\mathrm{Smooth}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\mathbf{A}}_{ij}\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\mathbf{A}}_{ij}}. \tag{11}$$

Here, $\mathbf{A}^{(\mathrm{s})}$ is the incidence matrix of the spatial hypergraph, with rows for spots and columns for hyperedges. Therefore, $\mathbf{A}_{ij}^{(\mathrm{s})}$ indicates whether spot i belongs to the spatial hyperedge centered at spot j, not the pairwise adjacency between them. The smoothing loss requires a measure of if two spots share a higher-order spatial context, so $\mathbf{A}^{(\mathrm{s})}$ is not directly applicable. Through clique expansion, $\hat{\mathbf{A}}_{ij}$ counts the spatial hyperedges shared by spots i and j. Thus, $\hat{\mathbf{A}}$ represents the spot–spot higher-order co-occurrence strength induced by the spatial hypergraph and is more appropriate for smoothing.

ZINB-based Reconstruction. To mitigate the zero-inflated, high sparsity and over-dispersion inherent in ST data, we follow DCA [46] to adopt a ZINB decoder to reconstruct the original expression patterns. Specifically, the NB distribution can be expressed as:

$$\mathrm{NB}(\mathbf{f} \mid \boldsymbol{\mu}, \boldsymbol{\theta}) = \Gamma(\mathbf{f} + \boldsymbol{\theta}) \left(\frac{\boldsymbol{\theta}}{(\boldsymbol{\theta} + \boldsymbol{\mu})}\right)^{\boldsymbol{\theta}} \left(\frac{\boldsymbol{\mu}}{(\boldsymbol{\theta} + \boldsymbol{\mu})}\right)^{\mathbf{f}} \frac{1}{\mathbf{f}!\,\Gamma(\boldsymbol{\theta})}. \tag{12}$$

The ZINB distribution is mathematically represented as:

$$\mathrm{ZINB}(\mathbf{f} \mid \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\pi}) = \boldsymbol{\pi}\delta_0(\mathbf{f}) + (1 - \boldsymbol{\pi})\,\mathrm{NB}(\mathbf{f}). \tag{13}$$

In ZINB distribution, $\boldsymbol{\pi}$ is the zero-inflation coefficient, modeling the probability that the observed value comes from inevitable zero component; $\boldsymbol{\mu}$ is regarded as the expected gene expression matrix after implicitly imputing pseudo zeros; $\boldsymbol{\theta}$ is the dispersion parameter, fitting the case where the variance exceeds the mean. We employ three fully connected networks to estimate these parameters based on the representation vector $\mathbf{h}_i$, which can be expressed as:

$$\begin{aligned}
\hat{\boldsymbol{\pi}}_i &= \mathrm{Sigmoid}(\mathrm{MLP}^1(\mathbf{h}_i)),\\
\hat{\boldsymbol{\mu}}_i &= \exp(\mathrm{MLP}^2(\mathbf{h}_i)),\\
\hat{\boldsymbol{\theta}}_i &= \ln\left(1 + \exp\left(\mathrm{MLP}^3(\mathbf{h}_i)\right)\right), i \in [1 : N].
\end{aligned} \tag{14}$$

With the estimators, the reconstruction loss is designed by the negative log-likelihood of the ZINB distribution:

$$\mathcal{L}_{\mathrm{ZINB}} = -\sum_{i=1}^{N} \log \mathrm{ZINB}(\mathbf{x}_i \mid \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\pi}}_i), \tag{15}$$

where $\mathbf{x}_i$ is the gene expression profile of spot i.

3.5. Multi-view self-supervised clustering

Inspired by SDCN [47], we adopt a self-supervised clustering strategy to generate more reliable guidance, precisely adapted to the

clustering task. We aim to mine the clustering structures from three view embeddings simultaneously, encouraging them to evolve collectively, which constrains the model to ensure that both the original views and the fused view are driven towards forming clear and consistent clusters. Concretely, we first compute the soft assignment probability of spot i belonging to pre-calculated clustering center j using Student's t-distribution as kernel:

$$q_{ij}^{(b)} = \frac{(1 + D_{ij}^{(b)}/\nu)^{-\frac{\nu+1}{2}}}{\sum_{j'=1}^{N}(1 + D_{ij'}^{(b)}/\nu)^{-\frac{\nu+1}{2}}}, \tag{16}$$

where $D_{ij}^{(b)} = \|\mathbf{h}_i^{(b)} - \mathbf{c}_j^{(b)}\|^2$, $b \in \{\mathrm{f}, \mathrm{s}, \mathrm{h}\}$ denotes the embedding branch (feature view, spatial view, or fused view), $\mathbf{c}_j^{(b)}$ is the centroid vector of cluster j in the corresponding branch's latent space, and ν is the degrees of freedom parameter for the Student's t-distribution ($\nu = 1$ by default). Subsequently, we sharpen the soft assignment distribution and emphasize high-confidence assignments to construct the target distribution, which can be expressed as:

$$p_{ij}^{(b)} = \frac{(q_{ij}^{(b)})^2 / \sum_{i=1}^{N} q_{ij}^{(b)}}{\sum_{j'=1}^{N}\left((q_{ij'}^{(b)})^2 / \sum_{i=1}^{N} q_{ij'}^{(b)}\right)}, \tag{17}$$

serving as an auxiliary distribution to guide the clustering process. We then adopt multi-view self-supervised clustering loss by adapting KL-divergence as:

$$\mathcal{L}_{\mathrm{Cluster}} = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}^{(\mathrm{h})} \log \frac{p_{ij}^{(\mathrm{h})}}{(\sum_{(b)} q_{ij}^{(b)})/3}, \tag{18}$$

which encourages the model to learn a clustering-friendly representation, thereby achieving tight coupling between representation learning and clustering.

3.6. Joint optimization

Building upon the preceding discussions, we herein propose a comprehensive unified framework, stHGNN, which jointly optimizes three losses. The total loss of stHGNN is formally expressed as:

$$\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{ZINB}} + \beta(t) \cdot \mathcal{L}_{\mathrm{Smooth}} + \gamma \cdot \mathcal{L}_{\mathrm{Cluster}}, \tag{19}$$

where $\beta(t), \gamma$ are used to balance each loss. In the initial few epochs, the model has not yet learned reliable node representations. Applying strong smoothing at this stage would propagate inaccurate structural priors. Therefore, we adopt a linear warm-up strategy for $\mathcal{L}_{\mathrm{Smooth}}$ to stabilize the model during the early stages of training, i.e., $\beta(t) = \beta \cdot \min(1, t/30)$, where t is the current epoch. By contrast, the clustering loss weighted by γ relies on soft assignment and a sharpened target distribution to provide continuous global clustering guidance. Its warm-up may weaken early clustering structure learning. Thus, we use warm-up only for β and keep γ fixed.

3.7. Pseudo-code

The pseudo-code of our proposed stHGNN is shown in Algorithm 1. The source code is available for reproducibility at: https://github.com/LCDYL/stHGNN

4. Experiments

4.1. Experimental setup

Benchmark Datasets. We perform extensive experiments on three benchmark datasets derived from diverse species and organs. The first dataset originates from the LIBD human dorsolateral prefrontal cortex (DLPFC) study [48], which systematically profiles gene expression across four to six cortical layers and white matter (WM), identifying numerous layer-enriched genes. This resource comprises 12 tissue sections, each manually annotated to contain five to seven anatomical regions (151669, 151670, 151671, and 151672 have 5 manually annotated regions, while the remaining 8 slices have 7 regions). The second dataset, generated using the 10x Visium platform, focuses on human breast cancer samples [49] and encompasses diverse tissue regions including pre-malignant lesions (DCIS/LCIS), invasive ductal carcinoma (IDC), tumor margins, and non-cancerous tissues. Across 20 spatially resolved regions, this dataset captures expression profiles of 36,601 genes. The third dataset comprises the mouse anterior brain tissue, containing 2695 spots that profile 32,285 genes, with 52 regions manually annotated according to the Allen Brain Atlas [50].

Baselines. We benchmark stHGNN against the following baselines: SCANPY [51] is a widely used single-cell analysis toolkit that mainly clusters spots according to transcriptional similarity without explicitly incorporating spatial dependence. SpaGCN [16] integrates gene expression, spatial coordinates, and histology information within a graph convolutional framework for spatial domain identification. STAGATE [19] employs an adaptive graph attention auto-encoder to learn spatially aware low-dimensional embeddings from gene expression and neighborhood structure. GraphST [13] leverages graph self-supervised contrastive learning to derive spatially informed representations for downstream clustering and related tasks. stHGC [14] builds a hybrid neighbor graph from multiple similarity metrics and introduces self-supervised graph representation learning with spatial regularization. DUSTED [20] enhances spatial transcriptomics analysis with a dual-attention graph autoencoder that simultaneously considers gene and spatial characteristics during denoising and embedding learning. Spotscape [15] captures global inter-spot relationships via a similarity telescope module for more discriminative representation learning, especially near domain boundaries. For these, we adopt the officially reported results when available. Otherwise, we rerun the official open-source code with recommended parameters. All rerun results use consistent cluster numbers and evaluation metrics, with no additional manual post-processing.

Evaluation Metrics. To evaluate the performance of spatial domain identification, we employ two widely used clustering metrics: Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI):

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

$$\mathrm{NMI}(U, V) = \frac{2 \cdot I(U;V)}{H(U) + H(V)}.$$

ARI ranges from −1 to 1, while NMI ranges from 0 to 1. For both metrics, higher values indicate better spatial domain identification performance.

Implement Details. Our proposed stHGNN is implemented and trained on an NVIDIA RTX 3090 GPU. For both feature-view and spatial-view hypergraph convolutions, the hidden and output dimensions are set to 256 and 128, respectively. And they are 64 and 128 for the ZINB decoder. The entire framework is optimized for 100 epochs using Adam [52]. In self-supervised clustering, K-means is used to initialize the cluster centers, which are subsequently updated based on the embeddings learned in each epoch. Among the key hyperparameters, k determines the expression-similarity neighborhood size and is set to 14 to balance relevant information preservation and noise introduction. The parameter r defines the local spatial receptive field and is set to 200, 550, and 250 on DLPFC, HBC, and MAB, respectively, according to dataset-specific spatial neighborhood densities. Moreover, $\beta = 0.1$ controls the spatial smoothing constraint with warm-up for stable early training, $\gamma = 0.001$ provides a weak auxiliary clustering constraint. Sensitivity analysis demonstrates the stability of these settings.

Algorithm 1 Optimization Algorithm of stHGNN

Input: gene expression matrix $\mathbf{X}$; spatial coordinates $\mathbf{S}$.
Initialize model parameters.
Compute incident matrix $\mathbf{A}^{(\mathrm{f})}, \mathbf{A}^{(\mathrm{s})}$ as Eq. (2) and Eq. (4);
while *not converged* **do**
  *// **Dual Hypergraph Enhanced Representing***
  Learn $\mathbf{H}^{(\mathrm{f})}$ and $\mathbf{H}^{(\mathrm{s})}$ via HGCN as Eq. (5);
  *// **Cross-View Knowledge Transfer***
  Compute fused representation $\mathbf{Z}_0$ as Eq. (7);
  Compute self-correlation representation $\mathbf{H}$ as Eq. (9);
  *// **Spatial- and Expression-view Constraints***
  Compute $\mathcal{L}_{\mathrm{Smooth}}$ as Eq. (11);
  Estimate $\hat{\pi}_i, \hat{\mu}_i, \hat{\theta}_i$ using three different MLPs;
  Compute $\mathcal{L}_{\mathrm{ZINB}}$ as Eq. (15);
  *// **Multi-View Self-Supervised Clustering***
  Compute soft assignments $q^{(b)}, b \in \{f, s, h\}$;
  Construct sharpened target $p^{(b)}$ from $q^{(b)}$;
  Compute $\mathcal{L}_{\mathrm{Cluster}}$ as Eq. (18);
  *// **Joint Optimization***
  Update parameters by minimizing
  $\mathcal{L}_{\mathrm{Total}} = \mathcal{L}_{\mathrm{ZINB}} + \beta(t) \cdot \mathcal{L}_{\mathrm{Smooth}} + \gamma \cdot \mathcal{L}_{\mathrm{Cluster}}$.
  Update latent representation $\mathbf{H}$;
end
Output: The learned representation $\mathbf{H}$, clustering labels.

Table 1
The spatial domain identification performance on DLPFC (151507-151510, 151669-151676), human breast cancer (HBC), and mouse anterior brain (MAB) under ARI and NMI. The best and the second-best results are highlighted with **blue** and green.

Method	151507		151508		151509		151510		151669		151670		151671	
	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
SCANPY [51]	0.20	0.21	0.15	0.21	0.19	0.27	0.14	0.22	0.10	0.16	0.09	0.16	0.12	0.24
SpaGCN [16]	0.39	0.49	0.33	0.43	0.35	0.51	0.37	0.50	0.23	0.36	0.33	0.43	0.42	0.53
STAGATE [19]	0.54	0.66	0.49	0.63	0.53	0.66	0.41	0.61	0.35	0.58	0.32	0.53	0.51	0.65
GraphST [13]	0.48	0.64	0.49	0.54	0.52	0.64	0.50	0.64	0.48	0.59	0.46	**0.68**	0.61	0.72
stHGC [14]	0.56	0.70	0.52	0.65	0.61	0.67	0.46	0.60	0.49	0.60	0.51	0.62	0.63	0.72
DUSTED [20]	0.46	0.65	0.45	0.64	0.45	0.64	0.49	0.62	0.47	0.59	0.29	0.51	0.50	0.65
Spotscape [15]	0.55	0.69	0.50	0.64	0.46	0.63	0.48	0.64	0.39	0.55	0.34	0.48	0.43	0.63
stHGNN (Ours)	**0.57**	**0.71**	**0.65**	**0.72**	**0.66**	**0.69**	**0.59**	**0.71**	**0.55**	**0.64**	**0.61**	0.63	**0.89**	**0.83**

Method	151672		151673		151674		151675		151676		HBC		MAB	
	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
SCANPY [51]	0.12	0.23	0.20	0.29	0.22	0.31	0.23	0.32	0.22	0.31	0.49	0.52	0.23	0.45
SpaGCN [16]	0.52	0.60	0.40	0.55	0.31	0.46	0.27	0.41	0.31	0.48	0.35	0.40	0.43	0.72
STAGATE [19]	0.54	0.66	0.45	0.63	0.48	0.58	0.36	0.50	0.49	0.63	0.52	0.66	0.39	**0.73**
GraphST [13]	0.63	0.61	0.63	**0.74**	0.43	0.61	0.55	0.62	**0.61**	0.66	0.53	0.67	0.41	0.71
stHGC [14]	0.77	**0.75**	0.57	0.69	0.60	0.68	0.55	0.65	0.57	0.67	0.45	0.62	0.37	0.72
DUSTED [20]	0.57	0.67	0.57	0.70	0.46	0.63	0.56	0.65	0.52	0.67	0.57	0.63	0.40	0.65
Spotscape [15]	0.69	0.72	0.43	0.59	0.43	0.57	0.49	0.63	0.43	0.58	0.27	0.41	0.40	0.70
stHGNN (Ours)	**0.79**	**0.75**	**0.62**	0.70	**0.64**	**0.72**	**0.58**	**0.69**	0.58	**0.68**	**0.65**	**0.72**	**0.46**	0.72

4.2. Experiment result

Result on DLPFC Dataset. As shown in Table 1, stHGNN achieves the best or comparable results among the compared methods across the DLPFC, HBC and MAB datasets in terms of both ARI and NMI. Notably, on DLPFC 151671, our stHGNN achieves an ARI of 0.89, showing a clear improvement over the second-best method (stHGC: ARI 0.63). On the biologically heterogeneous HBC dataset, our stHGNN attains average ARI and NMI scores of 0.65 and 0.72, respectively, demonstrating its robustness in distinguishing cancerous from non-cancerous regions. Similarly, on the highly complex MAB dataset with 52 annotated subregions, our stHGNN achieves the highest ARI/NMI (ARI 0.46, NMI 0.72), suggesting its capacity to resolve fine-grained spatial architectures. The observed performance gains of our proposed stHGNN can be attributed to our proposed dual hypergraph-enhanced representation learning module, which enables the model to capture cross-view, high-order cellular dependencies, thereby guiding a more biologically meaningful representation learning.

We evaluate the performance and biological plausibility of stHGNN through visualization and downstream analysis on the DLPFC dataset. Fig. 2(a) compares the spatial domain identification results of GraphST, stHGC, Spotscape, and stHGNN across the 12 DLPFC slices. Across most slices, stHGNN successfully delineates the cortical layers and additionally shows better intra-cluster compactness and inter-cluster separability than the compared baselines. The boundaries identified by stHGNN appear clearer, more continuous, and better aligned with the cortical region's highly ordered laminar structure. On DLPFC 151673, shown in Fig. 2(a), although GraphST achieves a competitive ARI, its layer boundaries are irregular and serrated, inconsistent with the smooth laminar structure in manual annotations. This jagged morphology may obscure accurate biological interpretation. Such jagged boundaries may stem from its contrastive learning strategy, which prioritizes local discriminability over representation smoothness, increasing sensitivity to noise. In contrast, the spatial domain identified by stHGNN show smoother and more continuous boundaries. This improvement may be related to our hypergraph smoothing constraint

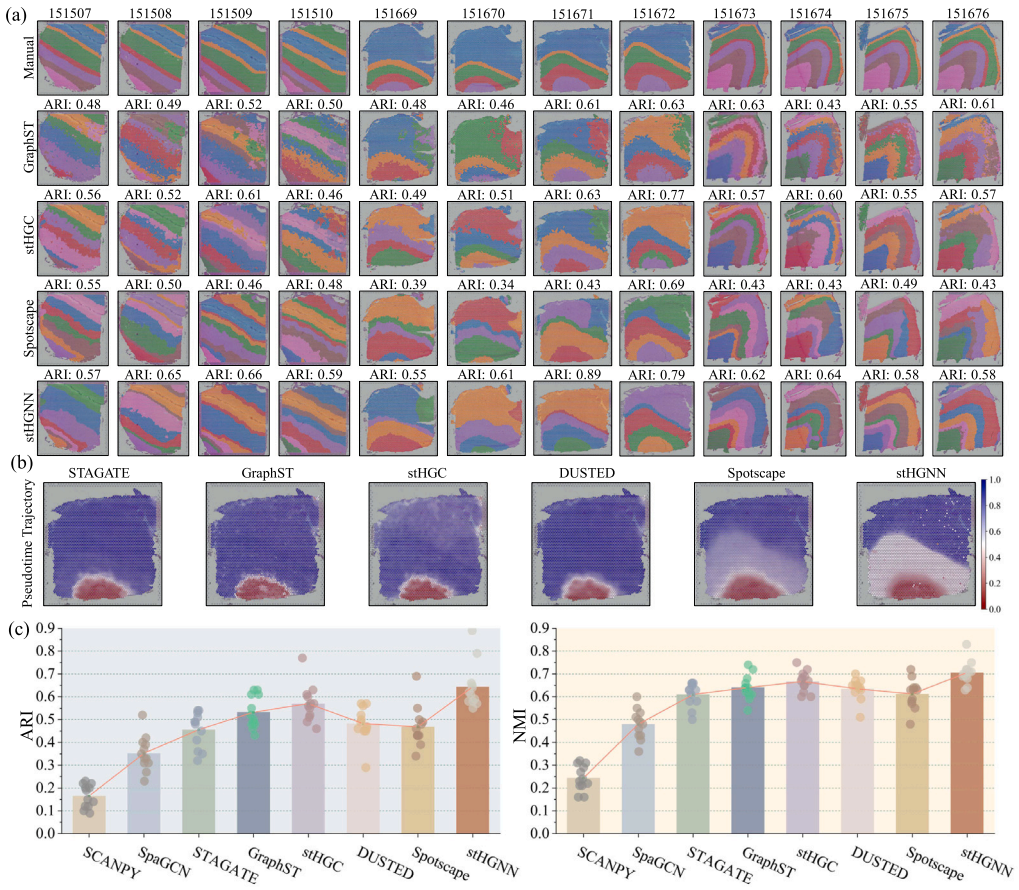

Fig. 2. (a) Manual annotation and visualizations of spatial domains identified by baselines and stHGNN on the 12 slices of DLPFC dataset. (b) Pseudotime analysis of baselines and stHGNN on DLPFC 151671. (c) Bar plot of baselines and stHGNN on DLPFC.

loss module, an approach that extends graph Laplacian regularization to a hypergraph. By promoting embedding similarity among both spatial neighbors and functionally related cells through hyperedges, it suppresses boundary irregularities, yielding more plausible spatial domains.

We further validate the biological dynamics captured by stHGNN's embeddings through pseudotime trajectory inference on DLPFC 151671 using the diffusion pseudotime (DPT) algorithm [53]. Fig. 2(b) shows that stHGNN yields a more continuous and biologically meaningful trajectory than baselines. This trajectory clearly reflects a developmental gradient from WM to Layer 1, consistent with biological knowledge of cortical laminar development. In contrast, the trajectories based on STAGATE, GraphST, stHGC and DUSTED representations exhibit less consistency. This suggests that by integrating dual-view hypergraph information with self-supervised clustering, stHGNN learns embeddings that better capture continuous cellular state transitions, providing a more reliable foundation for understanding spatial heterogeneity evolution.

Moreover, Fig. 2(c) shows grouped mean bar plots of ARI and NMI, demonstrating our advantages. For ARI (left), method performance increases from left to right. stHGNN achieves the highest mean ARI, better than recent methods like DUSTED and Spotscape, and also achieves the highest mean NMI. This quantitative result further strengthens the aforementioned conclusion: the spatial domains identified by stHGNN best align with the manual annotations. Notably, stHGNN exhibits a relatively short error bar, indicating low variability and high stability across slices. In contrast, while some baseline methods show acceptable mean metrics, their larger error bars reveal high sensitivity to the data and low reliability. Thus, these two quantitative evaluation results support the effectiveness of the stHGNN model in ST data clustering analysis. It achieves optimal performance across the reported metrics with low variance, demonstrating robustness and supporting downstream biological analysis.

Result on HBC Dataset. We conduct an evaluation of stHGNN on HBC dataset to evaluate its performance in deciphering the intricate tumor microenvironment and distinguishing between malignant and benign tissue regions. Fig. 3(a)(b) presents a side-by-side comparison of spatial domain identification results generated by six representative methods: SCANPY, STAGATE, GraphST, DUSTED, stHGC, and our stHGNN, with manual annotations. The manual annotation, grounded in histological morphology and established breast cancer marker genes, delineates four key pathological categories: invasive ductal carcinoma (IDC), ductal carcinoma in situ/lobular carcinoma in situ (DCIS/LCIS), tumor edges, and healthy tissue.

stHGNN achieves strong quantitative and visual results. It successfully delineates the spatial extent of diverse pathological states with morphology close to the manual annotation. Specifically, the boundaries of identified IDC, DCIS/LCIS, and healthy regions are spatially coherent and closely aligned with the manual annotation, capturing both large-scale tissue architecture and fine-grained transitions. The UMAP visualization in Fig. 3(b) is consistent with this spatial pattern. revealing compact, well-separated clusters for distinct pathological domains with limited intermixing. In contrast, baseline methods show relatively fragmented or less separated results. including fragmented domains, irregular boundaries, and insufficient discrimination between carcinoma subtypes. Consequently, their UMAP projections display overlapping or collapsed clusters, indicating weaker cluster separation.

Result on MAB Dataset. We also investigate the performance of stHGNN on MAB dataset, a benchmark characterized by high anatomical complexity. This dataset comprises 52 meticulously annotated

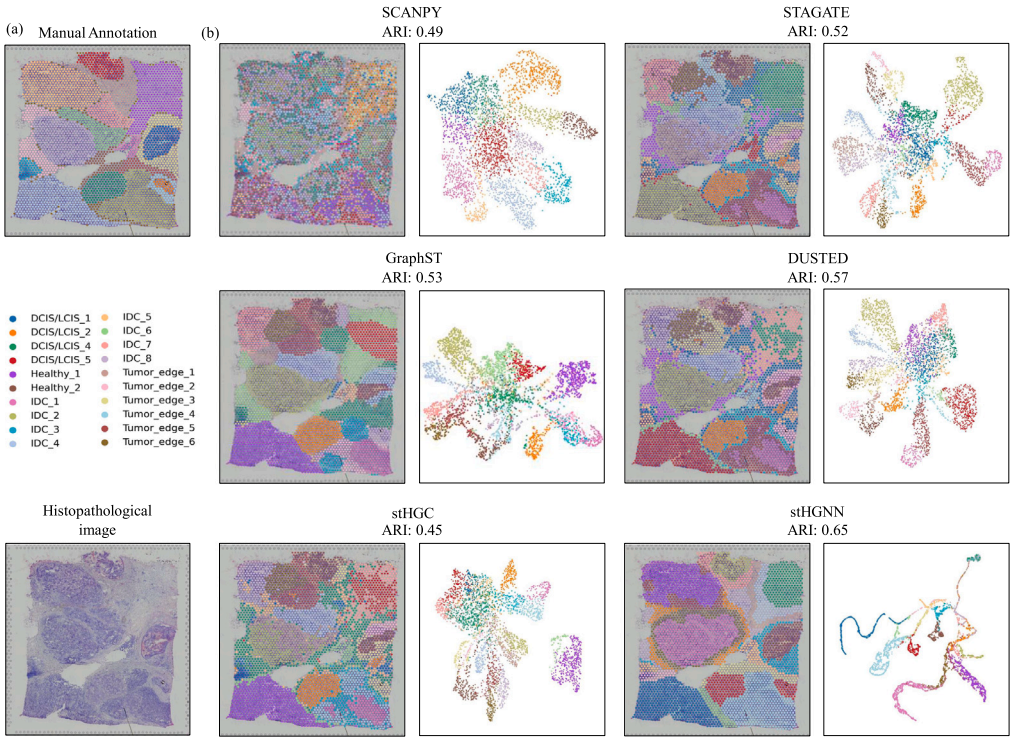

Fig. 3. (a) Manual annotation and histopathological image on HBC dataset. (b) Spatial domains detected by baselines and stHGNN.

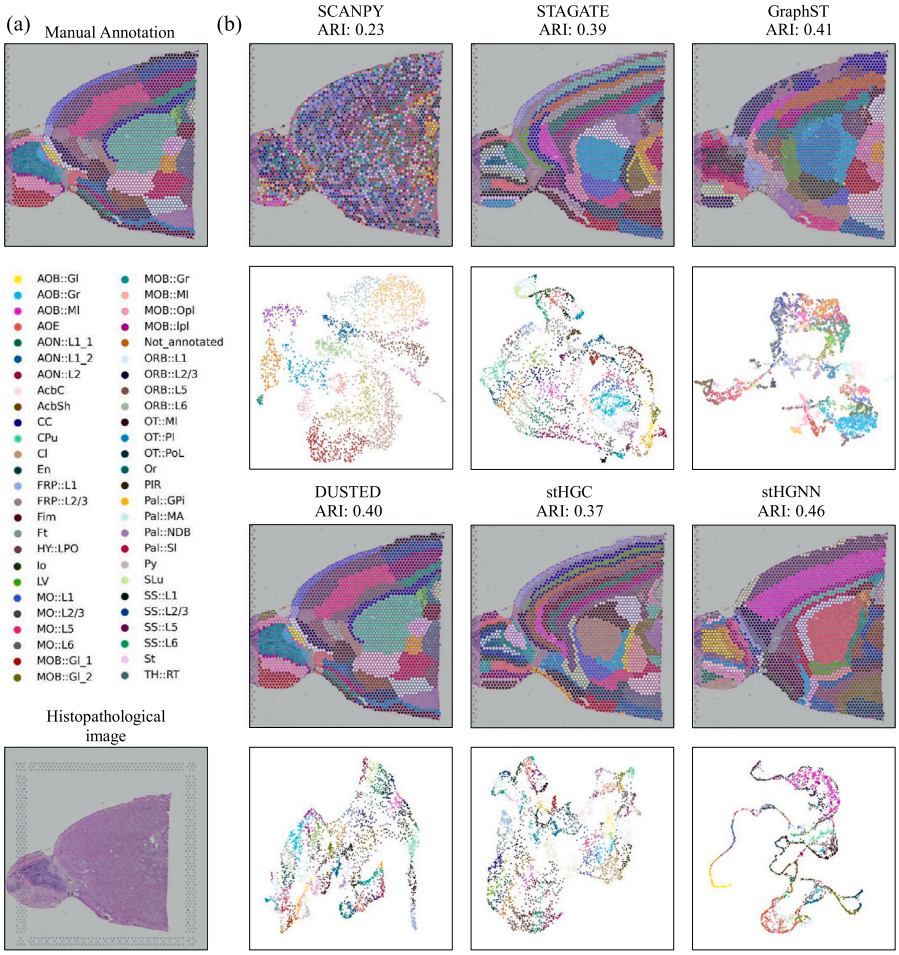

Fig. 4. (a) Manual annotation and histopathological image on MAB dataset. (b) Spatial domains detected by baselines and stHGNN.

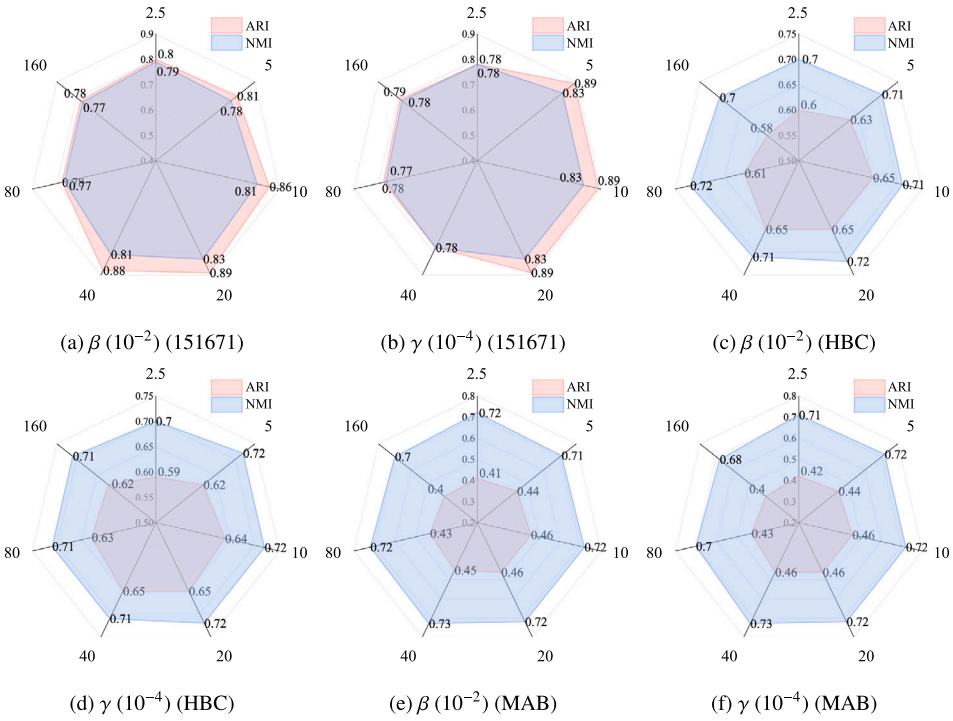

Fig. 5. Sensitivity analysis of balancing hyperparameters of loss functions.

subregions, including diverse nuclei, cortical areas, and fiber tracts, posing a significant and obvious challenging setting for any spatial domain identification method due to the fine-grained, irregular, and often interdigitated nature of these structures. Fig. 4(a)(b) provides a direct visual comparison of the spatial domain identification results produced by SCANPY, STAGATE, GraphST, DUSTED, stHGC, and our proposed stHGNN, all benchmarked against the manual anatomical annotation serving as the ground truth.

The qualitative results show that stHGNN better preserves several anatomical structures on the anatomically complex MAB dataset, which features 52 finely annotated subregions. stHGNN's domains closely match the manual annotations morphologically, successfully delineating the boundaries of small adjacent structures such as hypothalamic nuclei and cortical layers with relatively clear boundary correspondence while maintaining spatial contiguity within regions and clear separation between them. This fidelity is further reflected in the UMAP visualization, where spots from distinct anatomical subregions form compact, well-separated clusters with limited intermixing, indicating that the learned representation captures functional and spatial distinctions to some extent. In contrast, baseline methods show more fragmented or less separated results. yielding maps with fragmentation, boundary errors, or weaker discrimination. The quantitative results confirm these observations, as stHGNN achieves the highest ARI scores, supporting its improved performance in resolving complex tissue architectures through biologically plausible and discriminative representations.

4.3. Sensitivity analysis and ablation study

We perform sensitivity analysis on the two balancing hyperparameters, β and γ, on DLPFC 151671, HBC and MAB datasets. On DLPFC 151671, as shown in Fig. 5(a), when the weight of the smoothing loss β is too small, the constraint becomes weak, leading to lower performance. Conversely, when β is too large, the strong smoothing constraint causes over-smoothing, forcing biologically distinct regions to converge in the embedding space, thereby obscuring genuine heterogeneity and degrading performance. A similar trend is observed for the clustering loss weight γ, as illustrated in Fig. 5(b). A small γ provides insufficient guidance for representation learning, while an excessively large γ causes the model to overfit the current cluster centroids, leading to unstable centroid shifts. On HBC and MAB datasets, as shown in Fig. 5(c)(d)(e)(f), stHGNN maintains relatively stable performance across all tested values. Overall, our proposed stHGNN shows stable performance in general terms, demonstrating its robustness to different hyperparameter choices.

As for the remaining two key parameters, k and r, as shown in Fig. 6. k denotes the number of neighbor nodes for each node in the feature hypergraph, and r represents the spatial neighborhood perception range for constructing hyperedges for each node in the spatial hypergraph. As shown in the results on the DLPFC (151671), HBC, and MAB datasets, stHGNN shows stable performance across the tested values of k. For the parameter r, the only notable performance drop occurs in subfigure (d) for the DLPFC (151671) dataset. This is because when r is set unreasonably small during spatial hypergraph construction, it may fail to capture sufficient neighboring nodes to form meaningful hyperedges. This failure in hypergraph construction consequently thus leads to a loss of crucial spatial information. Otherwise, stHGNN shows limited sensitivity to r within a reasonable range.

As shown in Table 2, the complete stHGNN achieves the best performance across all evaluated settings, indicating complementary contributions of its components. M_1 shows a drop after replacing HGCNs with GCNs, suggesting that pairwise graph convolution is insufficient to capture the group-wise cellular dependencies encoded by hyperedges. M_2, M_9, and M_{10} verify the necessity of dual-view interaction: single-view variants lose either spatial continuity or expression similarity, while mean fusion cannot fully model adaptive compensation between spatial and feature views. The dataset-dependent behavior of M_9 and M_{10} further indicates that different tissues may rely on different view-specific cues. M_{11} overall underperforms the full model, showing that global self-correlation reorganization further refines the fused representation by exploiting global spot-level similarity. Finally, M_3-M_8 demonstrate the complementarity of the three losses. ZINB reconstruction preserves expression distribution, spatial smoothing introduces biological continuity, and clustering loss improves cluster

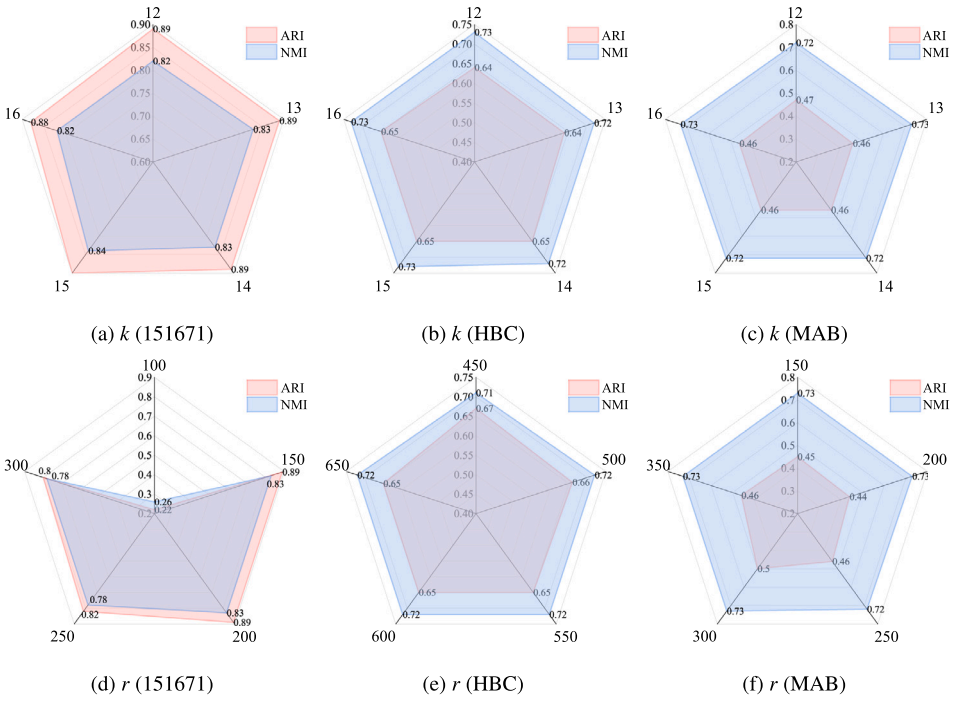

Fig. 6. Sensitivity analysis of parameters for graph construction on DLPFC (151671), HBC and MAB.

Table 2
Ablation study of stHGNN. M_1: GCN instead of HGCN; M_2: mean fusion instead of cross-attention; M_3, M_4, M_5: w/o $\mathcal{L}_{\text{Smooth}}, \mathcal{L}_{\text{ZINB}}, \mathcal{L}_{\text{Cluster}}$; M_6, M_7, M_8: only $\mathcal{L}_{\text{Smooth}}, \mathcal{L}_{\text{ZINB}}, \mathcal{L}_{\text{Cluster}}$; M_9, M_{10}: spatial-, feature-view only; M_{11}: w/o global self-correlation reorganization. The best are marked with **bold**.

Dataset	Metric	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	stHGNN (Ours)
DLPFC (151671)	ARI	0.62	0.75	0.83	0.79	0.83	0.61	0.75	0.60	0.84	0.38	0.85	**0.89**
	NMI	0.71	0.78	0.79	0.74	0.78	0.71	0.70	0.72	0.79	0.25	0.79	**0.83**
HBC	ARI	0.51	0.62	0.61	0.60	0.64	0.59	0.60	0.53	0.52	0.39	0.62	**0.65**
	NMI	0.65	0.64	0.69	0.68	0.70	0.66	0.60	0.66	0.68	0.49	0.69	**0.72**
MAB	ARI	0.38	0.44	0.44	0.43	0.42	0.40	0.39	0.33	0.37	0.45	0.40	**0.46**
	NMI	0.59	0.70	0.71	0.67	0.71	0.69	0.70	0.63	0.62	0.64	0.68	**0.72**

separability. The inferior performance of single-loss variants indicates that no individual objective is sufficient, and the complete model benefits from their joint optimization.

4.4. Downstream analysis

Differential Expression Analysis. Differentially expressed genes (DEGs) are key indicators of tissue structure and regulation. We perform DEG analysis on the spatial domains identified by stHGNN for the DLPFC 151671 and HBC datasets using the Wilcoxon rank-sum test, and subsequently rank the DEGs for each domain according to their adjusted p-values. As shown in Fig. 7(a)(b), the dot plots visualize the expression distribution of these top DEGs across different domains, where each row corresponds to a domain, and each column to a DEG. The dot size reflects the fraction of cells expressing the gene in that domain, while the color intensity encodes the mean expression level. We can find that, for a given domain, top-ranked DEG detected by stHGNN shows both the highest mean expression within that domain (i.e., the reddest dot in its row) and the higher specificity relative to all other domains (i.e., the reddest dot in its column). Specifically, the DEG CXCL14 is enriched in domain 0 (primarily matches Tumor_edge_3) of HBC, as shown in Fig. 7(a). Given that CXCL14 is a chemokine identified as a prognostic marker in breast cancer, known to be involved in recruiting immune cells and can exert tumor-suppressive effects in certain contexts by fostering an anti-tumor microenvironment [54], this finding suggests that domain 0 may represent a tissue niche where tumor progression is potentially restrained by active pro-inflammatory or immune-surveillance responses, which is consistent with the domain identification results. This pattern of high expression within a given domain and strong suppression in all others demonstrates that the domains identified by stHGNN are not only spatially coherent but also molecularly distinct, thereby validating the biological fidelity of stHGNN's spatial domain identification results. To further substantiate the above findings, we quantify the significance of DEGs using log2 fold change (log2FC). For domains where the advantage of our method is less pronounced, specifically IDC_2 and IDC_4 of HBC dataset, we compare the log2FC values of DEGs identified by stHGNN against those from baseline methods, as shown in Fig. 8(a)(b). The results show that our stHGNN performs better than those baselines, indicating stronger and more biologically meaningful differential expression. Similarly, for Layer 2 and Layer 4 of DLPFC dataset, our stHGNN maintains a clear advantage over competing methods, as shown in Fig. 8(c)(d). These consistent log2FC differences further suggest that the spatial domains identified by stHGNN exhibit enhanced molecular discriminability and biological coherence.

Domain-specific Genes Detection. To further validate the spatial domain identification results of our stHGNN from a molecular biology perspective, we examine the expression patterns of domain-specific genes. As shown in Fig. 9, the upper panel displays five representative domains identified by our stHGNN: IDC_8, IDC_2, DCIS/LCIS_4,

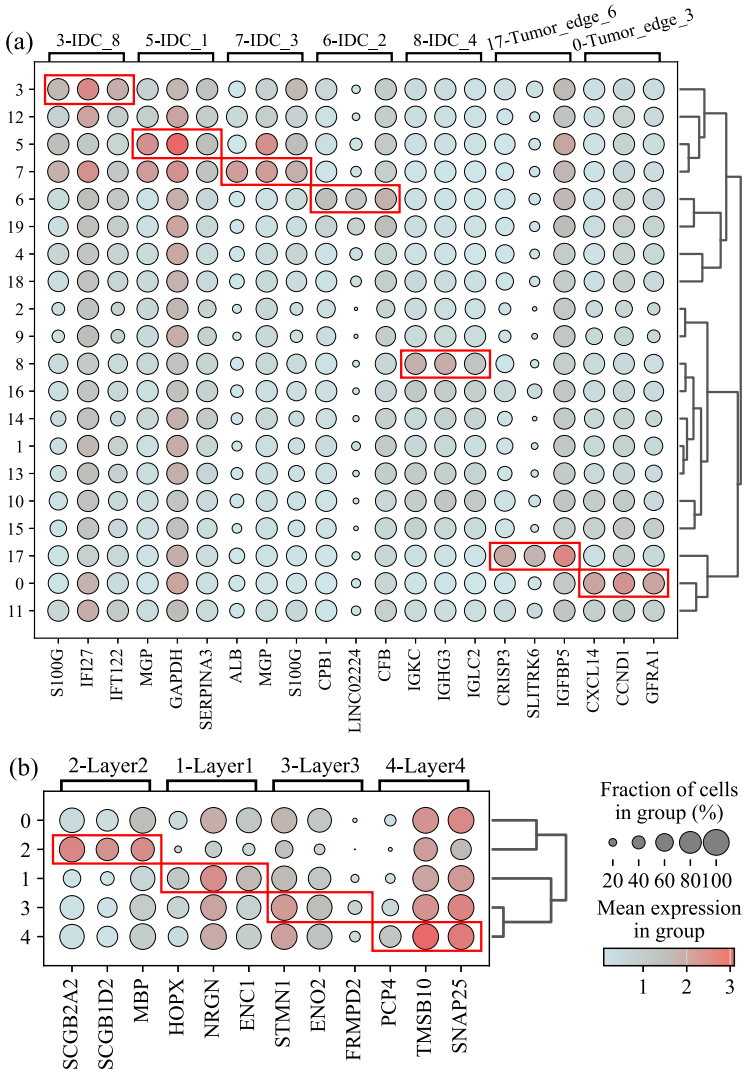

Fig. 7. Dot plots for DEGs detected by stHGNN on (a) HBC and (b) DLPFC 151671.

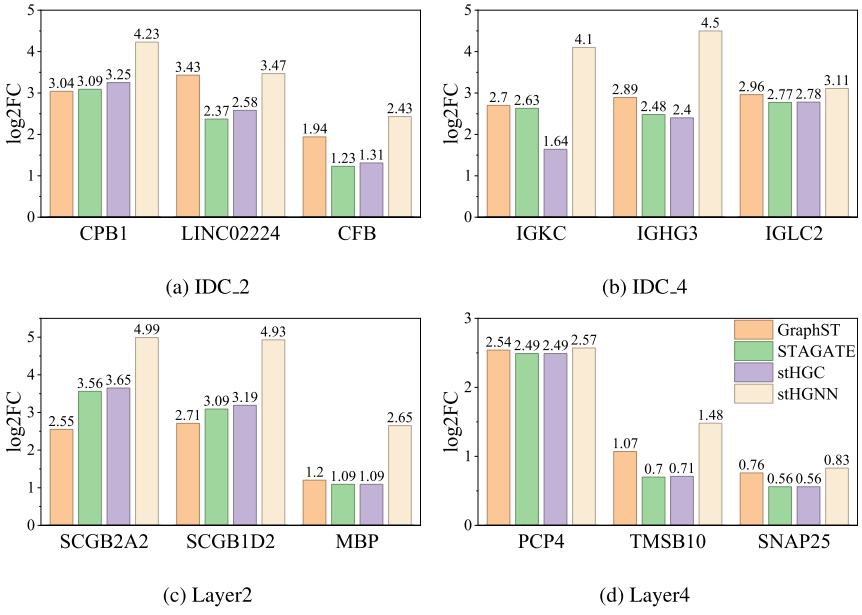

Fig. 8. Comparison of log2FC under stHGNN and competitive baselines STAGATE, GraphST, and stHGC for representative spatial domains.

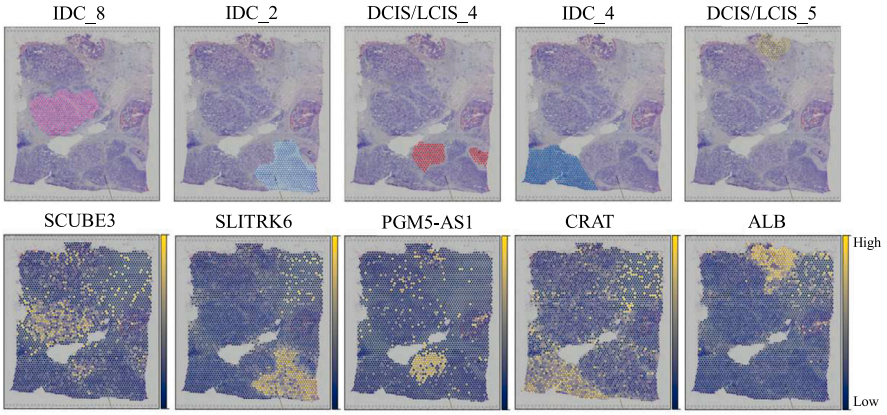

Fig. 9. Visualizations of spatial domains identified by stHGNN vs. expression of domain-specific genes.

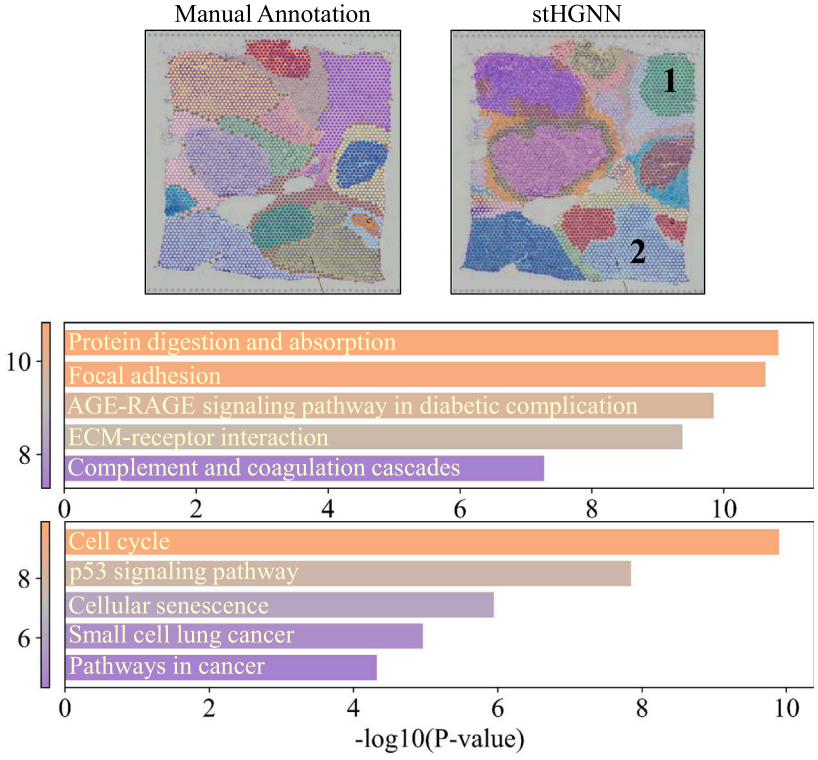

Fig. 10. KEGG Enrichment Visualizations. The upper bar plot corresponds to domain 1 (primarily matches Healthy_1), and the lower to domain 2 (primarily matches IDC_2).

IDC_4, and DCIS/LCIS_5. The lower panel visualizes the spatial expression of five domain-specific genes: SCUBE3, SLITRK6, PGM5-AS1, CRAT, and ALB, each exhibiting a highly localized pattern that aligns precisely with one specific domain. This one-to-one correspondence between spatial domains and gene expression patterns supports that the domains identified by our stHGNN reflect biologically distinct tissue compartments, confirming stHGNN's biological consistency and interpretability.

KEGG Enrichment Analysis. We perform KEGG enrichment on HBC dataset based on DEGs. Fig. 10 shows that in domain 2, the cell cycle and p53 signaling pathways are among the most enriched. The aberrant activation of the cell cycle is a hallmark of unlimited proliferative capacity in tumor cells. As a key tumor suppressor, significant perturbation in the p53 signaling pathway typically indicates a failure in maintaining genomic stability apoptosis. The combined dysregulation of these two pathways constitutes the classical driving force for cellular malignant transformation and progression, suggesting domain 2 primarily matches IDC_2. In contrast, in domain 1, the significant enrichment of the "Protein digestion and absorption" and "Focal adhesion" pathways is consistent with the core physiological functions of normal tissue. The former supports efficient nutrient uptake and metabolic homeostasis, while the latter maintains tissue integrity and orderly cellular arrangement, reflecting the alignment between tissue function and pathway activity, indicating that domain 1 primarily matches Healthy_1. This contrast, and the consistency between enriched pathways and tissue states, support the biological relevance of the domains identified by stHGNN.

5. Conclusion

Strengths. In this work, we propose stHGNN, a dual-view hypergraph enhanced framework for spatial domain identification in ST

data. By modeling higher-order cellular dependencies from spatial and expression views, our stHGNN learns tissue representations that better reflect the contextual and group-wise organization of spots. Through the integration of cross-attention fusion, self-correlation reorganization, hypergraph smoothing, ZINB-based reconstruction, and multi-view self-supervised clustering, the proposed framework jointly promotes structural fidelity, clustering friendliness, and biological consistency.

Limitations. The current framework still depends on manually specified hyperparameters for hypergraph construction and loss balancing, and its performance may be affected when spatial neighborhoods are not appropriately defined. In addition, although stHGNN shows strong results on annotated benchmark datasets, its generalization to broader scenarios, such as more diverse platforms, larger-scale tissues, and weaker supervision settings, still remains to be further explored in future studies.

Benefits. Future work will aim to make the framework more adaptive, scalable, and broadly applicable by exploring adaptive hypergraph construction, richer multimodal integration, and more efficient learning schemes for larger and higher-resolution ST datasets, while extending stHGNN to other spatial omics tasks in broader settings.

CRediT authorship contribution statement

Jinghong Tang: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Lezhi Chen:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Siyu Yi:** Writing – review & editing, Visualization, Project administration, Funding acquisition. **Wei Liu:** Writing – review & editing, Visualization, Conceptualization. **Mingyang Li:** Visualization, Investigation, Data curation. **Yifan Wang:** Writing – review & editing, Visualization, Data curation. **Ziyue Qiao:** Writing – review & editing, Validation, Methodology. **Bin Guo:** Writing – review & editing, Validation, Resources. **Xianggen Liu:** Writing – review & editing, Supervision, Resources, Conceptualization. **Jiancheng Lv:** Writing – review & editing, Supervision, Resources, Project administration. **Wei Ju:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Statement on Generative AI and AI-Assisted technologies in the writing

During the writing process of this work, the author used ChatGPT for language polishing only. After using this tool/service, the author reviewed and edited the content as needed and assumes full responsibility for the content of this publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 12501344 and 62306014, Postdoctoral Fellowship Program (Grade A) of CPSF under Grant BX20240239 and BX20250376, China Postdoctoral Science Foundation under Grant 2024M762201, Sichuan Science and Technology Program under Grant 2025ZNSFSC0808 and 2025ZNSFSC1506.

Data availability

Data will be made available on request.

References

[1] C.G. Williams, H.J. Lee, T. Asatsuma, R. Vento-Tormo, A. Haque, An introduction to spatial transcriptomics for biomedical research, Genome Med. 14 (1) (2022) 68.

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

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

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

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

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

[7] L. Liu, A. Chen, Y. Li, J. Mulder, H. Heyn, X. Xu, Spatiotemporal omics for biology and medicine, Cell 187 (17) (2024) 4488–4519.

[8] Y. Gao, Y.-L. Gao, J. Jing, F. Li, C.-H. Zheng, J.-X. Liu, A review of recent advances in spatially resolved transcriptomics data analysis, Neurocomputing 603 (2024) 128283.

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

[10] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev.: Comput. Stat. 2 (4) (2010) 433–459.

[11] A. Likas, N. Vlassis, J.J. Verbeek, The global k-means clustering algorithm, Pattern Recognit. 36 (2) (2003) 451–461.

[12] C. Fraley, A.E. Raftery, Model-based clustering, discriminant analysis, and density estimation, J. Amer. Statist. Assoc. 97 (458) (2002) 611–631.

[13] Y. Long, K.S. Ang, M. Li, K.L.K. Chong, R. Sethi, C. Zhong, H. Xu, Z. Ong, K. Sachaphibulkij, A. Chen, et al., Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst, Nat. Commun. 14 (1) (2023) 1155.

[14] R. Wang, Q. Dai, X. Duan, Q. Zou, stHGC: a self-supervised graph representation learning for spatial domain recognition with hybrid graph and spatial regularization, Brief. Bioinform. 26 (1) (2024).

[15] Y. Oh, J. Lee, Y. Kim, S. Seo, N. Lee, C. Park, Global context-aware representation learning for spatially resolved transcriptomics, in: Forty-Second International Conference on Machine Learning, 2025, URL https://openreview.net/forum?id=jeJGH6UDOL.

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

[17] R. Browaeys, W. Saelens, Y. Saeys, NicheNet: modeling intercellular communication by linking ligands to target genes, Nature Methods 17 (2) (2020) 159–162.

[18] S. Zhou, H. Xu, Z. Zheng, J. Chen, Z. Li, J. Bu, J. Wu, X. Wang, W. Zhu, M. Ester, A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions, ACM Comput. Surv. 57 (3) (2024) 1–38.

[19] K. Dong, S. Zhang, Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder, Nat. Commun. 13 (1) (2022) 1739.

[20] J. Zhu, Y. Li, Z. Tang, C. Chang, DUSTED: Dual-attention enhanced spatial transcriptomics denoiser, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, (1) 2025, pp. 1219–1227.

[21] W. Zhang, S. Yi, L. Chen, Y. Wang, Z. Qiao, Y. Zhou, W. Ju, Evidence-aware integration and domain identification of spatial transcriptomics data, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, (19) 2026, pp. 16352–16360.

[22] J. Long, Y. Zhu, C. Tang, K. Sun, Y. Liu, X. Yan, When genes speak: A semantic-guided framework for spatially resolved transcriptomics data clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, (1) 2026, pp. 800–808.

[23] Y. Xiao, Z. Li, X. Zheng, C. Tang, K. Sun, X. Liu, Spatially resolved transcriptomics data clustering with modality conflict modulation, IEEE Trans. Emerg. Top. Comput. Intell. (2026).

[24] J. Wu, Y. Wang, S. Yi, C. Yu, Z. Qiao, N. Yin, J. Lv, W. Ju, scGTN: Deep siamese graph transformer network for single-cell rna sequencing clustering, 2026, arXiv preprint arXiv:2606.18672.

[25] Y. Feng, H. You, Z. Zhang, R. Ji, Y. Gao, Hypergraph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (01) 2019, pp. 3558–3565.
[26] Z. Zhang, Y. Feng, S. Ying, Y. Gao, Deep hypergraph structure learning, 2022, arXiv preprint arXiv:2208.12547.
[27] Y. Takai, A. Miyauchi, M. Ikeda, Y. Yoshida, Hypergraph clustering based on pagerank, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 1970–1978.
[28] X. Sun, H. Yin, B. Liu, H. Chen, J. Cao, Y. Shao, N.Q. Viet Hung, Heterogeneous hypergraph embedding for graph classification, in: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021, pp. 725–733.
[29] N. Yadati, V. Nitin, M. Nimishakavi, P. Yadav, A. Louis, P. Talukdar, Nhp: Neural hypergraph link prediction, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1705–1714.
[30] B. Li, Y. Zhang, C. Zhang, X. Piao, Y. Hu, B. Yin, Multi-scale hypergraph-based feature alignment network for cell localization, Pattern Recognit. 149 (2024) 110260.
[31] W. Jing, J. Wang, D. Di, D. Li, Y. Song, L. Fan, Multi-modal hypergraph contrastive learning for medical image segmentation, Pattern Recognit. 165 (2025) 111544.
[32] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, X. Zhang, Self-supervised hypergraph convolutional networks for session-based recommendation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (5) 2021, pp. 4503–4511.
[33] M. Qu, G. Yang, D. Di, T. Su, Y. Gao, Y. Song, L. Fan, Multimodal cancer survival analysis via hypergraph learning with cross-modality rebalance, 2025, arXiv preprint arXiv:2505.11997.
[34] Z. Jin, X. Zheng, C. Tang, C. Li, Y. Liu, X. Liu, Task-aware information decoupling for multimodal clustering, IEEE Trans. Knowl. Data Eng. (2026).
[35] S. Li, J. Leng, J. Gan, M. Mo, X. Gao, Shape-centered representation learning for visible–infrared person re-identification, Pattern Recognit. 167 (2025) 111756.
[36] S. Li, J. Leng, C. Kuang, M. Tan, X. Gao, Video-level language-driven video-based visible-infrared person re-identification, IEEE Trans. Inf. Forensics Secur. (2025).
[37] L. Zhou, S. Li, N. Dong, Y. Tai, Y. Zhang, H. Li, Hierarchical prompt learning for image-and text-based person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, (16) 2026, pp. 13728–13736.
[38] J. Leng, C. Kuang, S. Li, J. Gan, H. Chen, X. Gao, Dual-space video person re-identification, Int. J. Comput. Vis. 133 (6) (2025) 3667–3688.
[39] S. Tao, S. Li, J. Ye, N. Dong, F. Li, H. Li, Spatial-temporal high-frequency learning for video-based visible-infrared person re-identification, IEEE Trans. Circuits Syst. Video Technol. (2026).
[40] Y. Yang, S. Li, J. Ye, N. Dong, F. Li, H. Li, DINOv2 driven gait representation learning for video-based visible-infrared person re-identification, in: Proceedings of the 33rd ACM International Conference on Multimedia, 2025, pp. 8283–8292.
[41] S. Li, F. Li, J. Li, H. Li, B. Zhang, D. Tao, X. Gao, Logical relation inference and multiview information interaction for domain adaptation person re-identification, IEEE Trans. Neural Networks Learn. Syst. 35 (10) (2023) 14770–14782.
[42] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data, Cell 177 (7) (2019) 1888–1902.
[43] S. Klamt, U.-U. Haus, F. Theis, Hypergraphs and cellular networks, PLoS Comput. Biol. 5 (5) (2009) e1000385.
[44] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2017,
[45] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Adv. Neural Inf. Process. Syst. 14 (2001).
[46] G. Eraslan, L.M. Simon, M. Mircea, N.S. Mueller, F.J. Theis, Single-cell RNA-seq denoising using a deep count autoencoder, Nat. Commun. 10 (1) (2019) 390.
[47] D. Bo, X. Wang, C. Shi, M. Zhu, E. Lu, P. Cui, Structural deep clustering network, in: Proceedings of the Web Conference 2020, 2020, pp. 1400–1410.
[48] K.R. Maynard, L. Collado-Torres, L.M. Weber, C. Uytingco, B.K. Barry, S.R. Williams, J.L. Catallini, M.N. Tran, Z. Besich, M. Tippani, et al., Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex, Nature Neurosci. 24 (3) (2021) 425–436.
[49] E. Buache, N. Etique, F. Alpy, I. Stoll, M. Muckensturm, B. Reina-San-Martin, M. Chenard, C. Tomasetto, M. Rio, Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice, Oncogene 30 (29) (2011) 3261–3273.
[50] H.W. Dong, The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse, John Wiley & Sons Inc, 2008.
[51] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (2018) 1–5.
[52] D.P. Kingma, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[53] L. Haghverdi, M. Büttner, F.A. Wolf, F. Buettner, F.J. Theis, Diffusion pseudotime robustly reconstructs lineage branching, Nature Methods (2016).
[54] N.S. Giacobbi, S. Mullapudi, H. Nabors, D. Pyeon, The chemokine CXCL14 as a potential immunotherapeutic agent for cancer therapy, Viruses 16 (2) (2024) 302.