PLOS Computational Biology

METHODS

Histology-informed spatial domain identification through multi-view graph convolutional networks

Huihui Zhang[1,2], **Jiaxing Chang**[1,3], **Zirong Li**[1], **Yue Sun**[1], **Pinli Hu**[1], **Haoxiu Wang**[1], **Hang Yang**[1,3], **Yonglin Ren**[4], **Xingtan Zhang**[1], **Zehua Chen**[3]*, **Kok Wai Wong**[2]*, **Haojing Shao**[1]*

1 Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, **2** School of Information Technology, College of Science, Technology, Engineering and Mathematics, Murdoch University, Murdoch, Australia, **3** College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Taiyuan, China, **4** College of Environmental and Life Sciences, Murdoch University, Murdoch, Australia

* chenzehua@tyut.edu.cn (ZC); K.Wong@murdoch.edu.au (KKW); shaohaojing@caas.cn (HS)

OPEN ACCESS

Citation: Zhang H, Chang J, Li Z, Sun Y, Hu P, Wang H, et al. (2026) Histology-informed spatial domain identification through multi-view graph convolutional networks. PLoS Comput Biol 22(6): e1014281. https://doi.org/10.1371/journal.pcbi.1014281

Editor: Hatice Ulku Osmanbeyoglu, University of Pittsburgh, UNITED STATES OF AMERICA

Received: October 25, 2025

Accepted: April 28, 2026

Published: June 1, 2026

Data availability statement: Code Availability The source code has been publicly released under the MIT open-source license, which complies with the requirements of the Open-Source Initiative. The source code of STESH is deposited as a Python package on Zenodo (https://doi.org/10.5281/zenodo.14954828),

Abstract

Identifying spatial domains is crucial in spatial transcriptomics, yet effectively integrating gene expression, spatial location, and histology remains challenging. We present STESH, a Spatial Transcriptomics clustering method that combines Expression, Spatial information and Histology. STESH extracts histological features using a convolutional neural network and generates expression, histology, spatial, and collaborative convolution modules for a multi-view graph convolutional network with a decoder and attention mechanism. We evaluated STESH on multiple tissue types and technology platforms. STESH consistently outperformed ten state-of-the-art methods, achieving superior clustering accuracy with the highest scores in adjusted Rand index, normalized mutual information, and Fowlkes-Mallows index.

Author summary

Identifying spatial domains in spatial transcriptomics is key to understanding tissue structure and gene expression patterns, yet existing approaches struggle to fully and effectively combine gene expression data, spatial location information, and histological images—three critical pieces of spatial transcriptomics data. To address this gap, we developed a new method that integrates all three types of information to accurately detect spatial domains. We used deep learning to extract detailed features from histological images, built separate analytical frameworks for each data type, and fused these frameworks with an attention mechanism to capture the intrinsic links between different data modalities. We tested this method on multiple tissue types and technical platforms, comparing it against ten leading approaches, and found it consistently achieved higher clustering accuracy and could precisely reconstruct complex biological tissue

and it is also available on GitHub (https://github.com/haojingshao/STESH). Data Availability The datasets used in this paper can be downloaded from the following websites: 1. The LIBD human dorsolateral prefrontal cortex (DLPFC) dataset https://research.libd.org/spatialLIBD/, 2. 10x Visium spatial transcriptomics dataset of human breast cancer https://www.10xgenomics.com/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard, 3. The processed Stereo-seq dataset from mouse olfactory bulb tissue https://github.com/JinmiaoChenLab/SEDR_analyses/blob/master/data/Stero-seq.tar.gz.

Funding: This study was supported by the National Natural Science Foundation of China (Grant number: 32200517 to H.S.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

structures, even uncovering fine details of tumor heterogeneity. This method enhances downstream spatial transcriptomics analyses and provides a more reliable tool for researchers, helping to deepen our understanding of the spatial organization of tissues and the relationship between gene expression and histological morphology.

Introduction

Identifying the spatial domain is the first important step in spatial transcriptomics (ST). Currently, methods for identifying spatial domains can be roughly divided into three categories by the number of dimensions: one-dimension (expression) clustering, two-dimensions (expression and spatial) clustering, and three-dimensions (expression, spatial and histological) clustering. Expression only clustering methods, such as k-means and Louvain [1], focus solely on gene expression data, and ignore spatial and histological contexts, which leads to discontinuous domain clustering. To address this, expression and spatial clustering methods such as Giotto's Hidden Markov Random Field model, SEDR's [2] deep autoencoder network, and STAGATE's [3] attention mechanism incorporate spatial information to improve domain identification. Other methods, like CoSTA [4], BASS [5], CCST [6], BayesSpace [7], GraphST [8], SCGDL [9], and AE-GCN [10], use various strategies, including convolutional neural networks, Bayesian approaches, and graph neural networks to enhance clustering accuracy by integrating spatial and gene expression data. These methods jointly utilize gene expression and spatial localization but do not exploit the histological information that high-resolution histological images can provide.

Expression, spatial and histological methods extract features from histological images to assess histological similarity. Combined with the spatial neighborhood structure, this similarity is then used to refine gene expression. SpaGCN [11] represents a pioneering graph convolutional network (GCN) algorithm that integrates histology, spatial coordinates, and gene expression data. This approach constructs an undirected weighted graph by combining histology images, spatial coordinates, and gene expression profiles, which is subsequently incorporated into the GCN framework. stLearn [12] combines histological features from histological images with gene expression from neighboring spots to achieve domain clustering. TIST [13] converts hematoxylin and eosin (H&E) images to grayscale and extracts image features based on a Hidden Markov Random Field (HMRF) model. This image processing strategy results in the loss of color information, hindering the consideration of highly correlated color pixels. DeepST [14] uses a pre-trained deep neural network model to extract feature vectors from histological image tiles, then integrates the extracted features with gene expression and spatial location data to characterize the correlation of spatially adjacent spots and create a spatially enhanced gene expression matrix. These methods establish a connection between image features and gene expression data but do not fully involve image information in subsequent training, which may lead to inaccuracies.

Integrating histological images, gene expression data, and spatial location information offers a promising approach for spatial clustering of spatial transcriptomics (ST) data. However, existing methods struggle to incorporate spatial information and align with high-resolution histological images. In this study, we developed a novel framework to effectively utilize histological images for spatial domain identification. Specifically, we employed a state-of-the-art convolutional neural network (CNN) to extract and learn detailed texture features from histological images, constructing corresponding graphs. To ensure the independence of histological patterns from spatial and expression information, we constructed three separate GCNs for spatial, histological, and expression data, respectively. Furthermore, we added a fourth integrative GCN that collaboratively learns from all three data types to capture the intrinsic relationships among these modalities.

Results

Overview of STESH

The main workflow of STESH is a multi-view graph convolutional network in deep learning. The first step is constructing three adjacent graphs for histological information from images, spatial information, and gene expression (Fig 1A). The histological graph is calculated by pre-trained convolutional neural networks methods such as ResNet50 (Fig 1B) [15]. The next step is reducing the dimension by autoencoder and decoder (Fig 1C). The last step is explaining the low-dimensional data by spatial clustering and trajectory inference (Fig 1D).

For benchmarking STESH, we utilized three public datasets that were widely used for comparison. The human dorsolateral prefrontal cortex (DLPFC) dataset contains 12 sections (3460–4789 spots), which were manually annotated in seven cortical layers [16]. The human breast cancer dataset (4898 spots) is also manually annotated in 20 layers. These two datasets were generated by 10X platform. To demonstrate that STESH can be applied to high-resolution data, we used a mouse olfactory bulb dataset (19109 spots) generated by Stereo-seq [17].

To quantitatively evaluate the contribution of histological image integration in STESH, we developed an image-free version of our framework. Comparative analysis revealed that the image-integrated GCN architecture significantly enhances clustering accuracy, demonstrating the substantial value of histological information in spatial transcriptomics analysis (S1 Fig). STESH uses a fixed random seed (seed = 100) throughout its implementation and our experiments (S2 Fig).

In order to comprehensively and systematically evaluate the performance of the spatial domain recognition method proposed in this article, ten representative spatial domain recognition methods were selected as comparison benchmarks. Leiden is a non-spatial method widely used and implemented in Seurat [18] and Scanpy [19]. GraphST [8], Spatial-MGCN [20], BayesSpace [21], SpaceFlow [22], SEDR [2], and STAGATE [3] are spatial methods that do not require histological images, while stLearn [12], SpaGCN [11], and DeepST [14] are spatial methods that use histological images. The spatial domain recognition performance of each method is measured using three common clustering metrics: the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Fowlkes-Mallows Index (FMI).

Benchmarking STESH on DLPFC dataset

Our analysis revealed that all spatial methods outperformed the non-spatial method Leiden, demonstrating the significant contribution of spatial information to clustering accuracy (S3–S12 Figs and S1–S5 Tables). As illustrated in Fig 2A–2C, STESH achieved superior performance with average ARI (0.61), NMI (0.69), and FMI (0.70), surpassing ten state-of-the-art methods. For detailed evaluation, we selected slice #151672, which exhibits a well-defined five-layer cortical structure (Fig 2D and 2S). STESH accurately delineated all five cortical layers, achieving an exceptional ARI of 0.81—the highest reported value for this dataset among recently published methods, including DeepST, SEDR, GraphST, SpaGCN, and Spatial-MGCN. While most methods maintained clear cluster boundaries, Leiden and stLearn exhibited less distinct separations. Notably, both Spatial-MGCN and STESH correctly identified layer 3 as a single cluster, whereas other methods erroneously split it into two clusters.

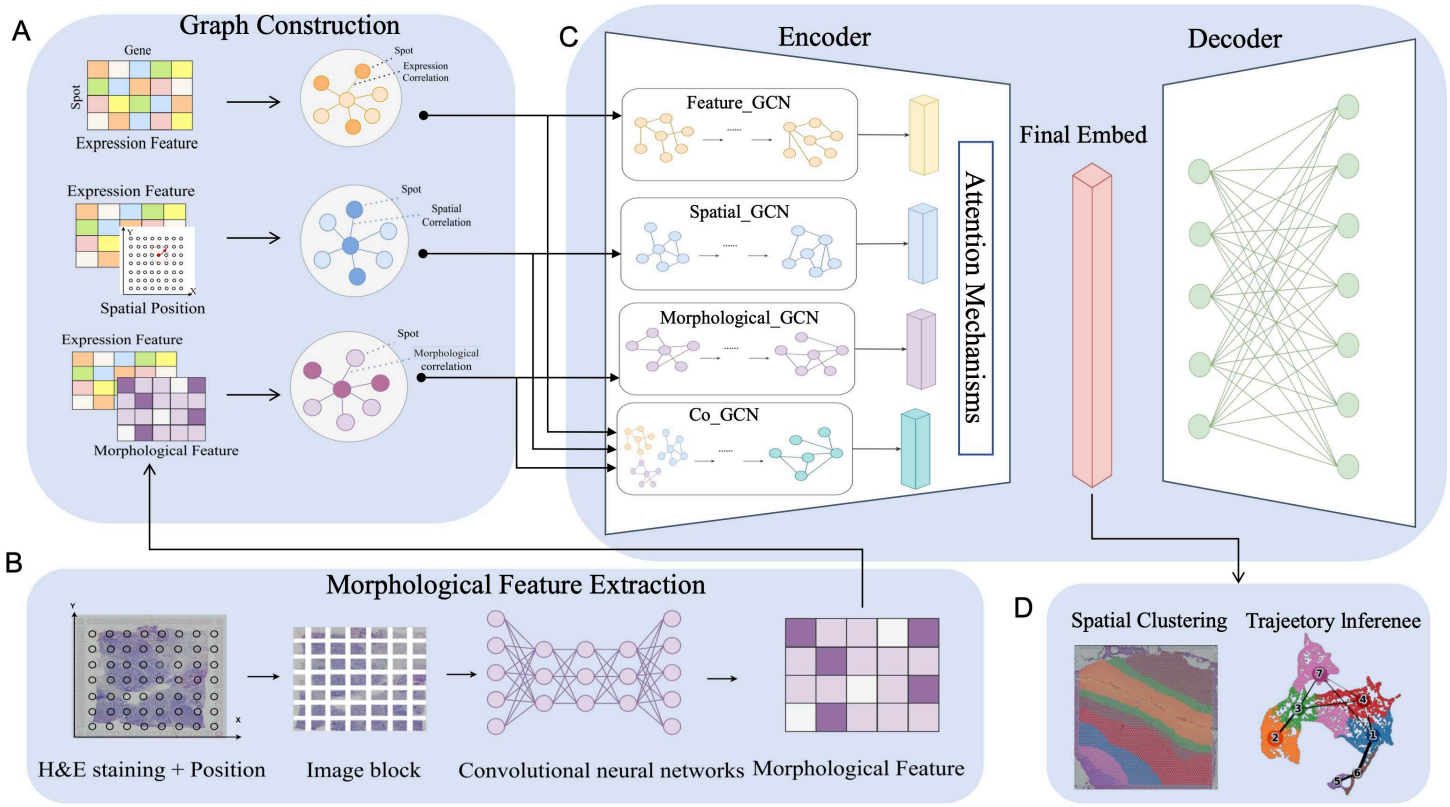

Fig 1. Workflow of STESH algorithm. A, Graph Construction. Based on the measured gene expression data, spatial location information and tissue section images in the spatial transcriptome, the spatial adjacency matrix, the expression adjacency matrix and the morphological adjacency matrix were constructed through different similarity measures. Combined with the gene expression feature, the expression graph, the spatial graph and the morphology graph were constructed. **B**, Morphological Feature Extraction. STESH utilizes H&E staining to extract histological features. Initially, the center of the image patch is determined based on the actual coordinates of the spot within the section. Subsequently, the tissue section image is cropped into square image patches of an appropriate size. Then, a pre-trained convolutional neural network (CNN) model is employed as the feature extractor, taking the segmented image patches as input to extract morphological information from the tissue section images. **C**, Multi-view GCN encoder. The GCN encoder uses feature GCN, spatial GCN, Morphological GCN,and Co-CCN, to learn the latent embeddings of spots on the feature graph, spatial graph and morphology graph, and adaptively fuses these embeddings with the learned weights using an attention mechanism. **D**, Application. Applying learned low-dimensional embeddings to spatial clustering and trajectory inference.

https://doi.org/10.1371/journal.pcbi.1014281.g001

The UMAP and PAGA trajectory plots further confirmed the robustness of STESH, GraphST, and Spatial-MGCN, all of which captured the sequential development of cortical layers with well-organized topological structures (Fig 2E).

We further identified differentially expressed genes (DEGs) and performed pathway enrichment analysis on the clustering results of STESH, spatial-MGCN, and GraphST. STESH exhibited the smallest discrepancy in the number of enriched genes relative to the ground truth(S13a Fig), together with the highest correlation ($R^2 = 0.949$, S13b Fig). Layer 6 (L6) is characterized by axons from diverse neuronal subtypes (e.g., L6cc, L6inv) traversing all cortical layers and projecting directly to the white matter, making it the layer with the densest axonal output to the white matter. Notably, STESH revealed that genes associated with axon and postsynapse pathways were abundantly expressed in L6(S13c Fig), which is consistent with known neurobiological functions.

Slice #151510, characterized by a complex cortical structure comprising seven layers across nine regions, was selected for analysis. In this structure, layers 2 and 3 were divided into two regions flanking layer 1 (Fig 3A). STESH, GraphST, Spatial-MGCN, SEDR, BayesSpace, and SpaceFlow demonstrated well-defined cluster boundaries, whereas

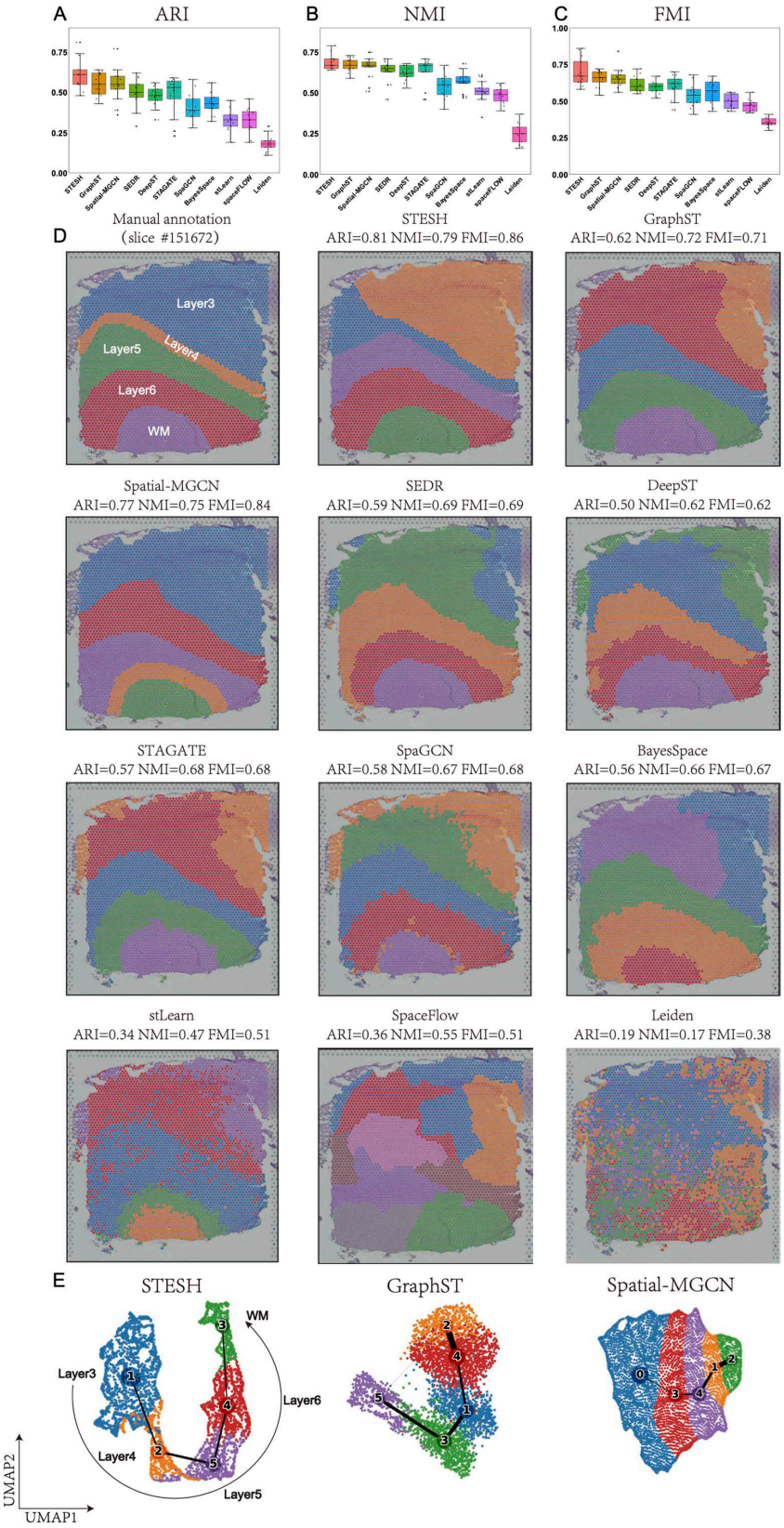

Fig 2. STESH improves spatial domain recognition in brain tissue. (A–C), Boxplot of the performance of STESH and other algorithms for all 12 DLPFCs. The y-axes are adjusted rand index (ARI), normalized mutual information (NMI), and Fowlkes-Mallows index (FMI) for A, B, and C,

respectively. **D**, DLPFC layers of slide 151672 were annotated by Maynard [16] et al., and identification of spatial domains by STESH, GraphST, Spatial-MGCN, SEDR, DeepST, STAGATE, SpaGCN, BayesSpace, stLearn, spaceFLOW, and Leiden. D, Trajectory inference and visualization of dimension reduction by Uniform Manifold Approximation and Projection (UMAP) for STESH, GraphST, and Spatial-MGCN, respectively.

https://doi.org/10.1371/journal.pcbi.1014281.g002

SpaGCN, stLearn, STAGATE, and DeepST exhibited discontinuous boundary delineation. Notably, GraphST, SEDR, BayesSpace, and SpaceFlow erroneously split layer 1 into two distinct clusters. Among all methods, only STESH and Spatial-MGCN successfully reconstructed the cortical structure with accuracy comparable to manual annotation. The UMAP and PAGA trajectory plots (Fig 3B) further revealed that both STESH and Spatial-MGCN exhibited a linear topological structure, precisely capturing the sequential development of cortical layers from Layer 1 to Layer 6 and the white matter (WM). In contrast, GraphST displayed an incorrect topological configuration, showing anomalous connections between clusters (1, 2, 6) and clusters (3, 7, 4), indicating potential misclassification of cortical layer identities. These results collectively demonstrate STESH's superior capability in spatial domain identification and its ability to accurately reconstruct complex biological structures.

Benchmarking STESH on a breast cancer dataset

Tumor heterogeneity detection represents another critical application of spatial transcriptomics clustering. We evaluated STESH and ten competing methods on a 10× Genomics Visium dataset of breast cancer, which was manually annotated into 20 distinct regions: four ductal carcinoma in situ/lobular carcinoma in situ (DCIS|LCIS) regions, eight invasive ductal carcinoma (IDC) regions, six tumor edge regions, and two healthy tissue regions. To ensure a fair comparison, all methods were constrained to identify 20 clusters, matching the manual annotation. STESH outperformed all other methods, achieving the highest clustering accuracy with an ARI of 0.61, NMI of 0.69, and FMI of 0.64 (Fig 4A). Notably, STESH, along with Leiden, stLearn, and STAGATE, correctly identified the IDC_4 tumor region as a single cluster without erroneous splitting, consistent with the manual annotation.

Cancer tissues exhibit significantly greater heterogeneity compared to normal cortical tissues. Accurately characterizing this complexity based solely on morphological features is challenging. However, spatial transcriptomics clustering methods that integrate gene expression data can uncover tumor heterogeneity missed by manual annotation. For instance, STESH, DeepST, and SEDR identified IDC_3 as comprising an outer "ring" and an inner core. Similarly, IDC_2 was subdivided by STESH, stLearn, DeepST, STAGATE, and SEDR, while DCIS/LCIS1 was split by STESH, STAGATE, and SEDR. To explore the biological significance of these classifications, we performed deconvolution analysis using Scanpy, aligning spatial data with scRNA-seq reference datasets [23]. We found that the percentage of tumor-associated macrophages (TAMs) of the outer ring is significantly higher than the inner core. TAM plays multi-function roles in cancer initiation and promotion, immune regulation, metastasis, and angiogenesis, such as providing essential support on tumor progression and metastasis [24]. Furthermore, the outer ring is also heterogeneous. STESH split the outer ring into clusters 16 and 17 (Fig 4B). Cluster 16 is directly adjacent to healthy tissue and had a higher TAM percentage, while cluster 17 is adjacent to another tumor tissue and had a similar TAM percentage to cluster 11.

Benchmarking STESH on Stereo-seq dataset

New high-resolution spatial transcriptomics methods such as Stereo-Seq and Seq-Scope [25] provided new opportunities for spatial transcriptomics tools. STESH on a Stereo-Seq mouse olfactory bulb dataset (19109 spots) was evaluated against two fine methods (STAGATE and SEDR) in previous benchmarking and one popular method (Leiden, Fig 4B). The histological image of the mouse olfactory bulb is briefly divided into seven layers, being the olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), granule cell

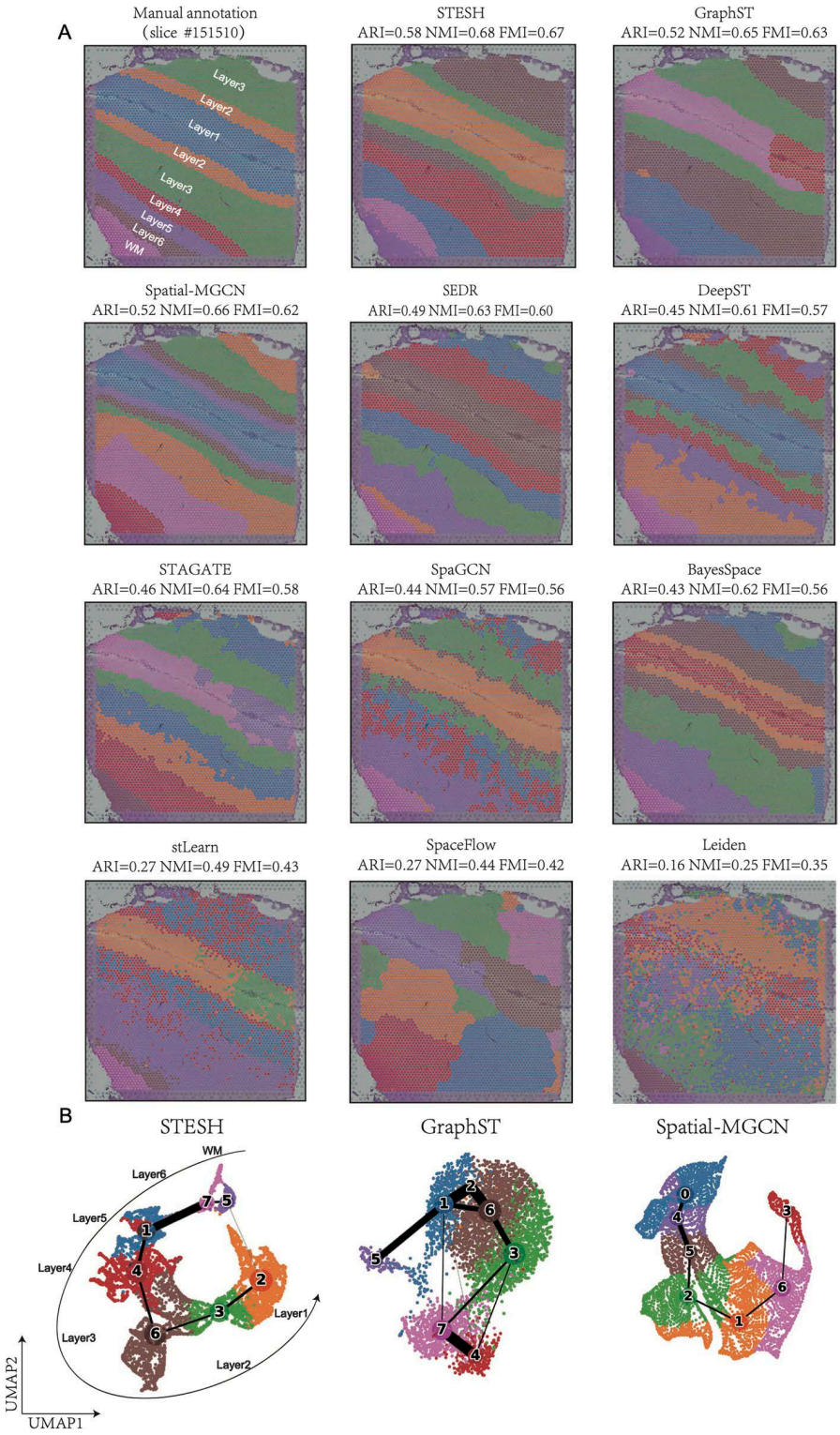

Fig 3. STESH improves spatial domain recognition in brain tissue slide 151510. A, The DLPFC slide was annotated by Maynard et al., and identification of spatial domains by STESH, GraphST, Spatial-MGCN, SEDR, DeepST, STAGATE, SpaGCN, BayesSpace, stLearn, spaceFLOW, and Leiden. **B**, Trajectory inference and UMAP visualization for STESH, GraphST, and Spatial-MGCN, respectively.

https://doi.org/10.1371/journal.pcbi.1014281.g003

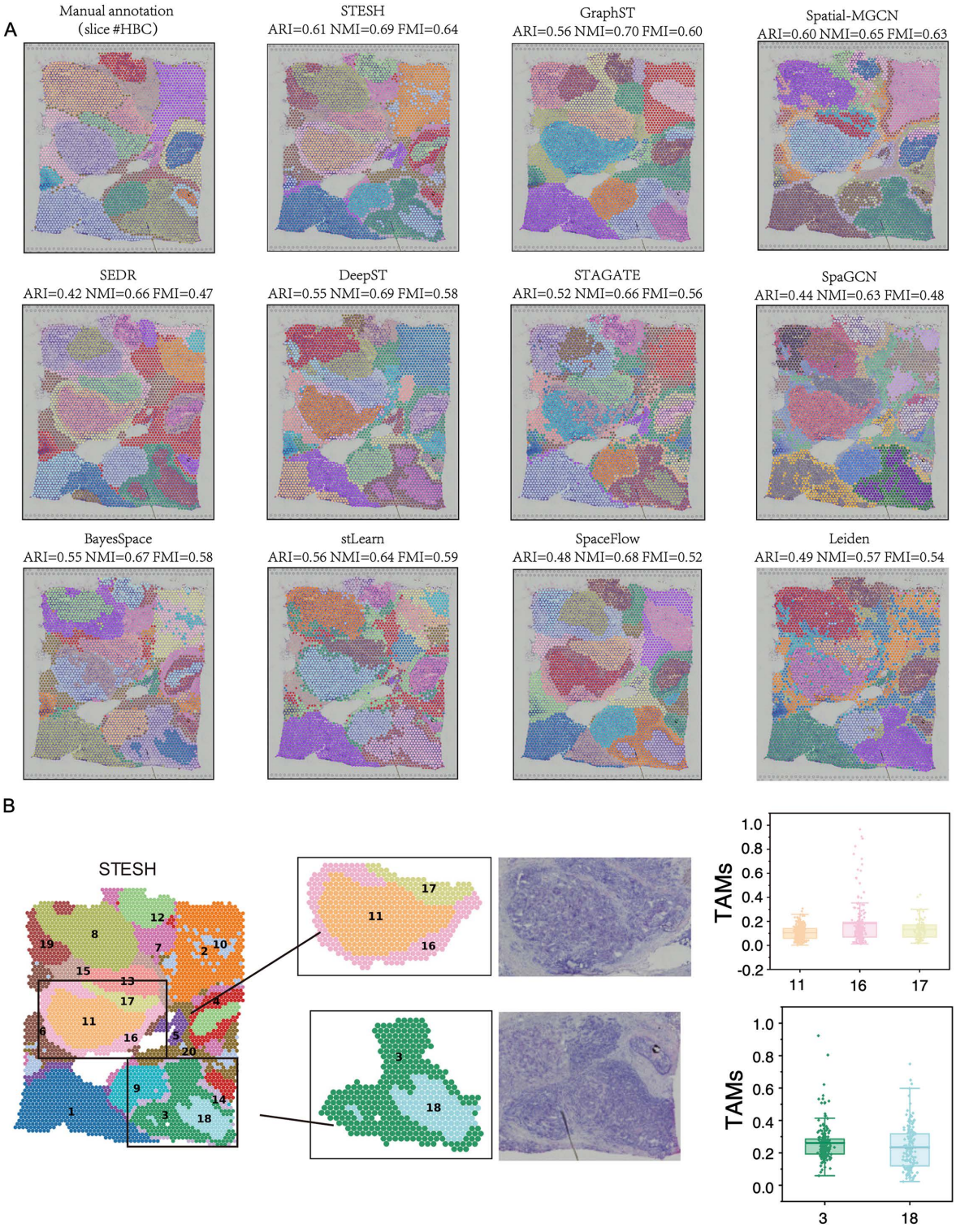

Fig 4. STESH improves spatial domain recognition in human breast cancer tissue. A, Identification of spatial domains by manual annotation, STESH, GraphST, Spatial-MGCN, SEDR, DeepST, STAGATE, SpaGCN, BayesSpace, stLearn, spaceFLOW, and Leiden. **B**, Visualization of cluster 11, 16, 17, 3 and 18 and their percentage of tumor-associated macrophages..

https://doi.org/10.1371/journal.pcbi.1014281.g004

layer (GCL), and rostral migratory stream (RMS, Fig 4A). Since there is no manual annotation for each spot, quantitative evaluation of the method performance using ARI, NMI or FMI was not possible. Visually, all four methods reconstructed the biological structure of the mouse olfactory bulb layers (Fig 4C). STESH clustered this dataset into seven circular regions, which were similar to the marker genes for each layer (Fig 4D). In the UMAP and PAGA trajectory plots, SEDR, STAGATE, and Leiden exhibited network-like topological structures, indicating erroneous connections between clusters (Fig 5E). In contrast, STESH demonstrated an almost linear topological structure, precisely capturing the sequential development from the ONL to the RMS. These results highlight STESH's superior clustering accuracy.

STESH operating parameter evaluation

To verify the effectiveness, stability and resource efficiency of the STESH algorithm, a comprehensive evaluation of its operating parameters was conducted, and the detailed experimental results are presented as follows. Firstly, ablation experiments were carried out to assess the contribution of each core component of the STESH algorithm (Fig 6a and 6c). Statistical analysis shows that the p-values of all ablation experimental groups are less than 0.05, indicating that there is a statistically significant difference between the complete model and the ablated models. After removing each core component, the performance of the STESH algorithm decreases by 10% to 80%, which fully demonstrates the necessity and irreplaceability of each component in ensuring the algorithm's performance.

Secondly, random seed experiments were performed to test the stability of the STESH algorithm (Fig 6b). We repeated the experiment 10 times with different random seeds, and the performance difference between each run was less than 7%. Further statistical analysis confirmed that there was no statistically significant difference in the average performance of the 10 experiments, proving that the STESH algorithm has good stability and is not sensitive to the setting of random seeds.

Thirdly, parameter sensitivity experiments were conducted to optimize the key parameters of the STESH algorithm (Fig 6d–6h). Loss weights α, β, and γ balance the contributions of different loss terms. Neighborhood size k defines the number of neighbors in the local structure modeling. A small k fails to capture sufficient contextual information, while an overlarge k introduces noisy irrelevant points. Spatial radius r controls the spatial range of the local region. We tested multiple groups of different parameter values and compared their mean ARI. The experimental results show that the selected values of α, β, γ, radius, and k all achieve the highest average performance among all tested parameter combinations, which verifies the rationality and optimality of the parameter setting in this study. Finally, we statistically analyzed the resource consumption and running efficiency of the STESH algorithm during operation, including memory usage, GPU memory occupancy, and running time(Fig 6i).

Conclusion and discussion

In this work, we introduce STESH, a novel method for processing spatial transcriptomics data to generate low-dimensional latent embeddings. STESH integrates spatial, RNA expression, and histological data using a multi-view graph convolutional network. Specifically, it constructs spatial adjacency matrices, feature adjacency matrices, and histological adjacency matrices, each based on different similarity metrics derived from tissue slice images, gene expression data, and spatial location information. By combining the gene expression matrix with the adjacency matrix, STESH builds spatial, feature, and histological graphs. The method employs a multi-view graph convolutional model to learn specific embeddings for each of these graphs. An attention mechanism is designed to adaptively learn the importance of each embedding, which is used to generate the final low-dimensional embedding. Clustering is performed based on learned embeddings to achieve spatial domain recognition, integrating multi-modal information in spatial transcriptomics. STESH enhances downstream analyses such as spatial clustering, UMAP visualization, and trajectory inference, offering new platforms and tools for spatial transcriptomics research. It addresses the limitations of existing methods in effectively utilizing spatial information, matching high-resolution histological images, and improving the accuracy of spatial domain

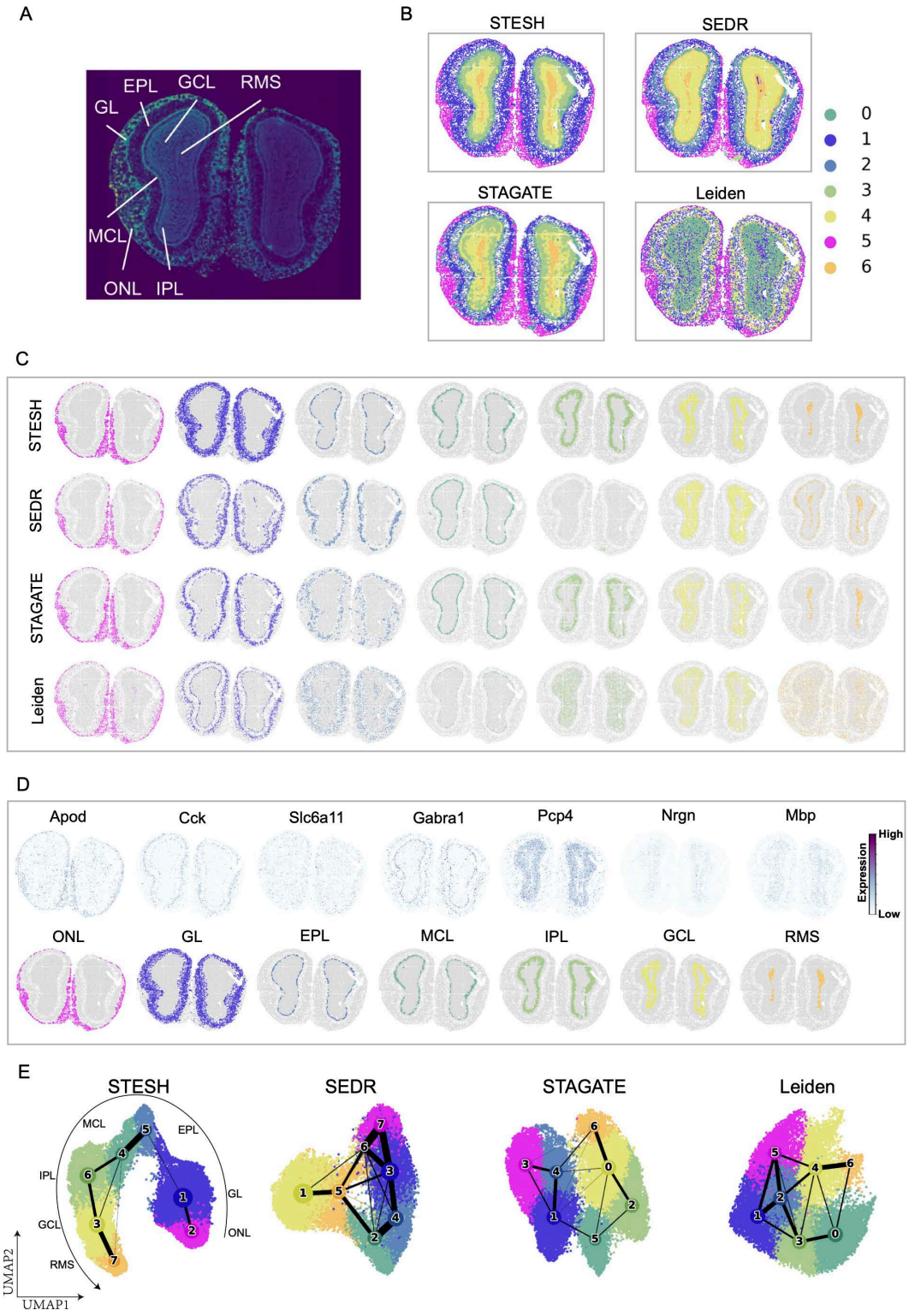

Fig 5. Application of STESH on Stereo-seq dataset. A, Laminar organization of a DAPI-stained mouse olfactory bulb. **B**, Clustering results from STESH, SEDR, STAGATE, and Leiden on the olfactory bulb Stereo-seq data. **C**, Visualization of individual identified clusters. **D**, Marker genes and predicted layers from STESH. **E**, Trajectory inference and UMAP visualization.

https://doi.org/10.1371/journal.pcbi.1014281.g005

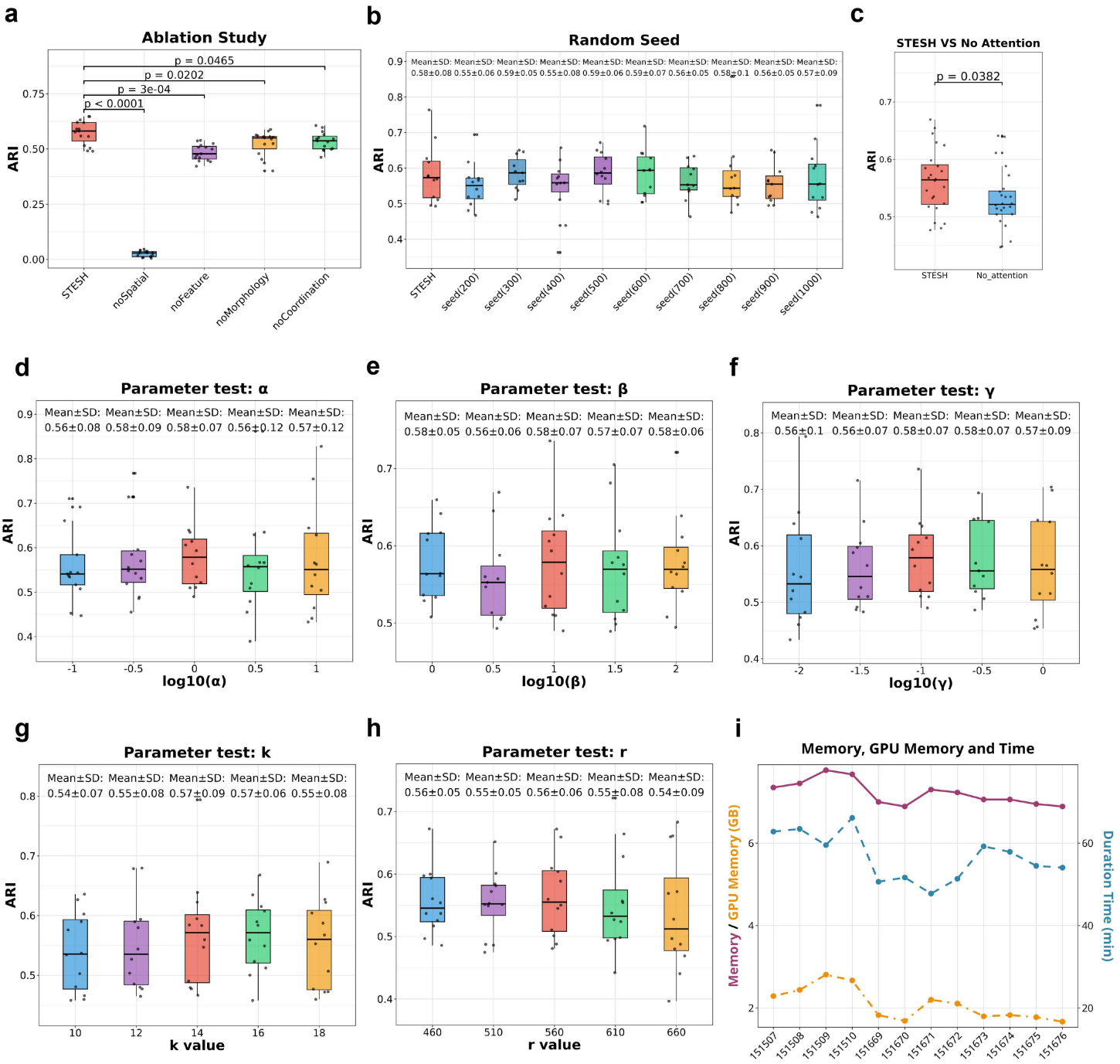

Fig 6. STESH parameter evaluation. (a) Ablation study: Performance comparison of model variants with different component ablations. **(b)** Random seed robustness: Performance stability across different random initializations. **(c)** Attention validation: Performance with and without the attention module. **(d–h)** Parameter sensitivity: Performance changes for parameters α, β, γ, k, and z at different values. **(i)** Resource consumption: Memory, GPU usage, and runtime across experimental configurations.

https://doi.org/10.1371/journal.pcbi.1014281.g006

recognition. It provides a more effective approach for the precise analysis of spatial domains with similar gene expression patterns and in situ histological features.

Accurate identification of spatial domains is fundamental for describing genomic heterogeneity and cell interactions, as well as for various downstream tasks in spatial transcriptomics analysis. Existing methods have limitations in effectively utilizing spatial information and match high-resolution histological images. The strength of STESH lies in its deep integration of tissue image information. Instead of simply overlaying tissue image information on gene expression data and spatial location information, morphological dependencies are constructed between spots using tissue images. This relationship is continuously and deeply considered at every stage of model training, achieving a greater level of data fusion.

Method

Morphological feature extraction

Image segmentation of tissue slice images is performed by determining the center of image blocks based on the actual coordinates of the spots in the slices. The tissue slice images are cropped into 50 × 50 pixel square image blocks and normalized to 224 × 224 pixels. For spots near the image edge, black background (0 value) is automatically filled to ensure uniform patch size. The original image is normalized to the range of [0, 255]. For CNN input, normalization is performed using the mean and standard deviation of the ImageNet dataset. All data augmentation operations use a fixed random seed, including 7 types of random augmentation methods such as random autocontrast, Gaussian blur, and random affine transformation. A pre-trained CNN model (ResNet50 by default) is used to extract 2048-dimensional original features, which are then reduced to 50 dimensions by PCA.

Construction of multi-dimensional adjacency matrices

By calculating different similarity metrics, the spatial adjacency matrix A_S, expression adjacency matrix A_f, and morphological adjacency matrix A_m are constructed. These matrices, along with the gene expression matrix, are used for subsequent graph construction.

The spatial adjacency matrix is used to represent the spatial relationships between spots in the tissue. By calculating the physical distance between spots, it is possible to determine which cells are neighbors. This helps the subsequent model learn the gene expression patterns of spatially adjacent cells. The calculation is as follows:

First, the Euclidean distance *ds* between each spot and all other spots is computed based on the spatial location of the spots to measure spatial similarity. To define their adjacency relationships, a pre-defined radius r is used. The spatial adjacency matrix S between two spots is calculated by combining the Euclidean distance with the predefined radius r.. The specific construction method is as follows: for a given spot, if the distance between the centers of two spots is smaller than the calculated radius r, they are considered adjacent. The corresponding position in the adjacency matrix is set to 1; otherwise, it is set to 0. The calculation formula is as follows:

$$As_{ij} = \begin{cases} 1 & if\ Ds_{ij} < r \\ 0 & if\ Ds_{ij} \geq r \end{cases}$$

The expression adjacency matrix reflects the connectivity between spots at the gene expression level and is an important component of the expression graph. The calculation method is as follows: the cosine distance d_f is used to measure gene expression similarity, revealing the underlying structure of gene expression. For a given spot i and spot j, assuming their gene expressions are x_i and x_j, the cosine distance is calculated using the formula:

$$df_{ij} = \frac{x_i \cdot x_j}{|x_i|\,|x_j|}$$

To better define gene expression similarity, a k-nearest neighbor graph is constructed for the gene expression matrix, referred to as the expression graph. This graph is represented by the expression adjacency matrix A_f for N spots, where the adjacency matrix is calculated based on the cosine distance d_f.. The top k most similar gene expression spots are defined as neighbors. The specific calculation method is as follows: for a given spot i, if spot j is a neighbor of spot i, then the corresponding position in the adjacency matrix is set to 1; otherwise, it is set to 0.

$$Af_{ij} = \begin{cases} 1 & if\ j \in Nf_i \\ 0 & if\ j \notin Nf_i \end{cases}$$

Morphological adjacency matrix

The morphological adjacency matrix A_m reflects the relationship between cells or tissue regions based on their morphological features. First, the tissue images are segmented based on the coordinates of each spot, and features are extracted from the images using a pre-trained convolutional neural network (CNN), which are then used as the morphological feature vectors for each spot. Since the feature vectors extracted by the pre-trained convolutional neural network are high-dimensional, Principal Component Analysis (PCA) is used to select the top 50 features as the latent morphological feature representation M for each spot. Finally, for a given spot i and spot j, the Pearson correlation *dm* between the morphological latent features is calculated using the formula:

$$dm_{ij} = \frac{cov(M_i, M_j)}{\sigma_{M_i}\sigma_{M_j}} = \frac{E\left(M_iM_j\right) - E\left(M_i\right)E(M_j)}{\sqrt{E\left(M_i^2\right) - E^2(M_i)}\sqrt{E\left(M_j^2\right) - E^2(M_j)}}$$

The top k most similar image blocks for each spot are identified as neighbors. for a given spot i, if spot j is a neighbor of spot i, then the corresponding position in the adjacency matrix is set to 1; otherwise, it is set to 0. The construction formula is the same as for the expression adjacency matrix.

$$Am_{ij} = \begin{cases} 1 & if\ j \in Nm_i \\ 0 & if\ j \notin Nm_i \end{cases}$$

Graph construction

Graph construction mainly includes the creation of the expression graph, spatial graph, and morphological graph.

1) The expression graph is based on the cosine distance d_f between each spot to measure gene expression similarity. These similarity metrics reflect the degree of similarity in gene expression between different spots, which is the key basis for constructing the expression graph. Based on the calculated gene expression similarity, the top k nearest neighbors for each spot are selected, and an adjacency matrix A_f representing gene expression similarity is constructed. Finally, the gene expression data is used as the node attribute feature matrix X and combined with the adjacency matrix, the complete expression graph $G_f\left(A_f, X\right)$ is constructed.
2) The spatial graph is primarily based on spatial location information. By calculating the Euclidean distance between each spot in the tissue slice, spatial similarity is measured, and an adjacency matrix that reflects spatial similarity is constructed. Based on the computed Euclidean distance d_s and a pre-defined radius r, the adjacency matrix A_s is constructed. The gene expression feature matrix X is used as the node attribute feature matrix. The spatial graph $G_s\left(A_s, X\right)$ is constructed using the spatial similarity adjacency matrix A_s and the node attribute feature matrix X..

3) The morphological graph is constructed based on the morphological information extracted from the tissue slices. The Pearson correlation d_m between the image blocks corresponding to each spot is calculated to measure morphological similarity. Using the computed Pearson correlation $d_{m'}$, the top k nearest neighbors for each spot are selected, and an adjacency matrix A_m representing morphological similarity is constructed. The gene expression feature matrix X is used as the node feature matrix. The morphological graph $G_m\left(A_m, X\right)$ is constructed using the morphological similarity adjacency matrix A_m and the node feature matrix X..

Multi-view graph convolutional autoencoder

The Multi-View Graph Convolutional Autoencoder (MCGCN) consists of spatial convolution modules, feature convolution modules, morphological convolution modules, and collaborative convolution modules. The method for generating low-dimensional embeddings is as follows:

1. The spatial convolution module performs convolution operations on the spatial graph and applies the following hierarchical propagation rule to generate low-dimensional embeddings E_s: $E_s^{(l+1)} = MVGCN_s(A_s, X) = ReLU\left(\widetilde{A}_s E_s^{(l)} W_s^{(l)}\right)$.

Where: $W_s^{(l)}$ is the weight parameter of the l-th layer in the spatial convolution module;
$E_s^{(l)}$ is the low-dimensional embedding generated by the l-th layer in the spatial convolution module; ReLU represents the ReLU activation function. $\widetilde{A}_s$ is the symmetric normalized adjacency matrix in the spatial graph, and its calculation method is shown below: $\widetilde{A}_s = \widetilde{D}_s^{-\frac{1}{2}} A_s \widetilde{D}_s^{-\frac{1}{2}}$, where $\widetilde{D}_s$ represents the degree matrix of A_s.

2. The feature convolution module performs convolution operations on the expression graph and applies the following hierarchical propagation rule to generate low-dimensional embeddings E_f: $E_f^{(l+1)} = MVGCN_f(A_f, X) = ReLU\left(\widetilde{A}_f E_f^{(l)} W_f^{(l)}\right)$. Where: $W_f^{(l)}$ is the weight parameter of the l-th layer in the feature convolution module; $E_f^{(l)}$ is the low-dimensional embedding generated by the l-th layer in the feature convolution module; $\widetilde{A}_f$ is the symmetric normalized adjacency matrix in the expression graph, and its calculation method is the same as for the spatial graph.

3. The morphological convolution module performs convolution operations on the morphological graph and applies the following hierarchical propagation rule to generate low-dimensional embeddings E_m:

$E_m^{(l+1)} = MVGCN_m(A_m, X) = ReLU\left(\widetilde{A}_m E_m^{(l)} W_m^{(l)}\right)$. Where: $W_m^{(l)}$ is the weight parameter of the l-th layer in the morphological convolution module; $E_m^{(l)}$ is the low-dimensional embedding generated by the l-th layer in the morphological convolution module; $\widetilde{A}_m$ is the symmetric normalized adjacency matrix in the morphological graph, and its calculation method is the same as for the spatial graph.

4. Since gene expression has a certain correlation with spatial distribution and histological information, a collaborative convolution module is introduced for collaborative convolution of three graphs to extract the collaborative embedding E_{cm}, E_{cs}, and E_{cf} for histological, spatial, and expression graph, respectively.

$$E_{cs}^{(l+1)} = MVGCN_c(A_s, X) = ReLU\left(\widetilde{A}_s E_{cs}^{(l)} W_s^{(l)}\right)$$

$$E_{cf}^{(l+1)} = MVGCN_c(A_f, X) = ReLU\left(\widetilde{A}_f E_{cf}^{(l)} W_f^{(l)}\right)$$

$$E_{cm}^{(l+1)} = MVGCN_c(A_m, X) = ReLU\left(\widetilde{A}_m E_{cm}^{(l)} W_m^{(l)}\right)$$

Through the above calculated $E_{cm}^{(l)}$, $E_{cs}^{(l)}$ and $E_{cf}^{(l)}$, the collaborative embedding $E_c^{(l)}$ is defined as: $E_c^{(l)} = \frac{E_{cm}^{(l)}+E_{cs}^{(l)}+E_{cf}^{(l)}}{3}$.

In order to learn a more consistent representation, the consistency is measured by comparing the difference in the covariance matrices between E_{cm}, E_{cs}, and E_{cf}. Define the consistency constraint loss L_{con} as:

$$L_{con} = \| \widetilde{E}_{cm} - \widetilde{E}_{cs} \|_2^2 + \| \widetilde{E}_{cm} - \widetilde{E}_{cf} \|_2^2 + \| \widetilde{E}_{cs} - \widetilde{E}_{cf} \|_2^2$$

Where: $\widetilde{E}_{cs}$ is the embedding extracted from the spatial graph,

$\widetilde{E}_{cf}$is the embedding extracted from the expression graph,

$\widetilde{E}_{cm}$is the embedding extracted from the morphological graph.

Attention mechanism

We introduce an adaptive multi-view attention mechanism to fuse the four view-specific embeddings E_m, E_s, E_f, and E_c, (each with dimension d_{in}) generated by the multi-view graph convolutional autoencoder. The attention module learns view-wise importance weights in a data-driven manner, with the detailed computation as follows:

First, we compute the raw attention scores for each view embedding via a two-layer projection network with Tanh activation:

$$\widehat{\omega}_i = W_2 \cdot \tanh\left(W_1 \cdot E_i + b_1\right), \; i \in \{s, f, m, c\}$$

where $W_1 \in R^{d_{hidden} \times d_{in}}$ and b_1 denote the weight matrix and bias of the first linear layer (hidden dimension d_{hidden} =16), and $W_2 \in R^{1 \times d_{hidden}}$ is the weight matrix of the second bias-free linear layer.

Next, we normalize the raw scores using the Softmax function over the view dimension to ensure the weights sum to 1 (non-negativity and unit sum):

$$\omega_i = \frac{\exp(\widehat{\omega}_i)}{\sum_{i \in \{s,f,m,c\}} \exp(\widehat{\omega}_i)}$$

Finally, the fused embedding is obtained by weighted summation of the view-specific embeddings (followed by a linear projection as our final step):

$$E_{fusion} = \sum_{i \in \{s,f,m,c\}} \omega_i \cdot E_i$$

$$E_{final} = W_{linear} \cdot E_{fusion} + b_{linear}$$

In practice, the input to the attention module is a tensor $Z = [E_s, E_f, E_m, E_c] \in R^{B \times 4 \times d_{in}}$ (where B is batch size), and the above operations are vectorized for efficient batch computation. The learned weights $\{\omega_s, \omega_f, \omega_m, \omega_c\}$ reflect the relative importance of each view, and the module outputs both the final embedding E_{final} and the attention weights $\{\omega_i\}$ for interpretability analysis.

Negative binomial decoder

The negative binomial decoder used the negative binomial distribution to model the characteristics of the data. When reconstructing the gene expression matrix, it accounts for the discrete and variable nature of gene expression data, enabling it to capture the complex global patterns present in ST data. Its composition is as follows.

Firstly, an intermediate layer containing a linear layer and a batch normalization layer is defined, which is used to map the low-dimensional latent embedding E_{final} (the encoder output) to a higher-dimensional space for extracting higher-level features. The ReLU activation function is used to introduce nonlinearity. Secondly, two linear layers are introduced to map the output of the middle layer to the original dimensions, respectively. Dispersion θ and mean μ of the distribution are obtained. Both the dispersion. and the mean are processed by different activation functions to ensure that their values are within range. This design helps the model to better fit the real data. Specifically, for a given gene expression matrix X, assuming that it conforms to the negative binomial distribution, the probability distribution of gene expression is defined as follows:

$$f_{NB}\left(X|\mu,\theta\right) = \frac{\Gamma(\mathrm{X}+\theta)}{\Gamma(\mathrm{X}+1)\Gamma(\theta)}\left(\frac{\theta}{\theta+\mu}\right)^{\theta}\left(\frac{\mu}{\mu+\theta}\right)^{\mathrm{X}}$$

Where μ and θ are calculated by the decoder, representing the mean and dispersion, respectively. Γ represents the gamma function. In order to minimize the difference between the predicted value and the true value, the negative log-likelihood estimation is used to reconstruct the loss of the original gene $L_{\mathrm{NB\_rec}}$:

$$L_{\mathrm{NB\_rec}} = -\log\left(f_{NB}\left(X|\mu,\theta\right)\right)$$

Loss function

To better capture the structural information of the graph, regularization constraints are based on the spatial adjacency matrix, spatial negative sampling matrix, histological adjacency matrix, and histological negative sampling matrix. Taking into account the influence of multiple graph structures, the model can learn the structural information of the graph more comprehensively by comparing the cosine similarity between nodes.

The regularization constraint loss $\mathrm{L}_{\mathrm{reg_m}}$ of the histological graph is defined as follows:

$$\mathrm{L}_{\mathrm{reg_m}} = -\sum_{i=1}^{N_{\mathrm{spots}}}\left(\sum_{j\in \mathrm{Nm}_i}\log\left(\sigma\left(mat_{ij}\right)\right) + \sum_{k\notin \mathrm{Nm}_i}\log\left(1-\sigma\left(mat_{ik}\right)\right)\right)$$

Where: Nm_i is the histological neighbor set of the spot i and mat_{ij} is based on the cosine similarity matrix of the learned latent representation E_{final}. The loss function contains two parts the histological adjacency matrix positive sample loss is to encourage the embedding vectors of histological adjacent nodes to be closer in the embedding space to learn the local structure in the histological graph, histological adjacency matrix negative sample loss is to encourage the embedding vectors of histological non-adjacent nodes to be further apart in the embedding space to avoid noise and overfitting in the learned graph.

Similarly, the regularization constraint loss $\mathrm{L}_{\mathrm{reg_s}}$ of the spatial graph is defined as follows:

$$\mathrm{L}_{\mathrm{reg_s}} = -\sum_{i=1}^{N_{\mathrm{spots}}}\left(\sum_{j\in \mathrm{Ns}_i}\log\left(\sigma\left(mat_{ij}\right)\right) + \sum_{k\notin \mathrm{Ns}_i}\log\left(1-\sigma\left(mat_{ik}\right)\right)\right)$$

Taking into account the regularization constraint loss $\mathrm{L}_{\mathrm{reg_m}}$ of the histological graph and the regularization constraint loss $\mathrm{L}_{\mathrm{reg_s}}$ of the spatial graph, the overall regularization constraint loss $\mathrm{L}_{\mathrm{reg}}$ is defined as follows:

$$\mathrm{L}_{\mathrm{reg}} = \frac{\mathrm{L}_{\mathrm{reg_m}} + \mathrm{L}_{\mathrm{reg_s}}}{2}$$

In summary, the STESH model consists of the multi-view graph convolution encoder and the negative binomial decoder. The total loss function L is composed of the reconstruction loss L_{NB_rec} of the original gene, the consistency constraint loss L_{con}, and the regularization constraint loss L_{reg}.

$$L = \alpha L_{NB\_rec} + \beta L_{con} + \gamma L_{reg}$$

Downstream analysis

The learned latent representation can be applied to downstream analysis tasks. To cluster the spatial data, STESH uses mclust() in R. To visualize it in two dimension, STESH uses sc.pl.umap() function in Scanpy. To infer the trajectory, STESH uses sc.tl.paga() function in Scanpy.

Data preprocessing

For all datasets, to reduce technical noise in ST data, we first remove spots outside the main tissue regions. Then, we use the SCANPY package to filter out genes with low expression or low variance from the raw gene expression data, selecting the top 3000 highly variable genes. Finally, we normalize the data using a scaling factor. The normalization function is defined as follows: $X_{ij} = \frac{count_{ij}}{s_i} \times 10^4$ and $s_i = \text{median}\left(\sum_{j=1}^{M} count_{ij}\right)$ $(1 \leq i \leq N, 1 \leq j \leq M)$, where X_{ij} denotes the original expression value of the j-th gene in the i-th spot.

Parameter settings

The entire model is implemented using PyTorch 2.6.0. In the experiments of this model, two graph convolution layers are used for the three views, where the output dimension of the first convolution layer is set to 128, and the output dimension of the second convolution layer is set to 64. To optimize the model, the Adam optimizer is employed with a learning rate of lr = 0.001. The weight parameter values are set to α = 1, β = 10a nd γ = 0.1. For all baseline methods, we use the default parameters from the original paper.All experiments are conducted on a server with an NVIDIA GeForce RTX 3090 GPU running Ubuntu 20.04.5.

Supporting information

S1 Fig. Identification of spatial domains for DLPFC by STESH with no histological images.
(DOCX)

S2 Fig. Replication of spatial domain identification in DLPFC using STESH with a consistent random seed (100).
(DOCX)

S3 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151507.
(DOCX)

S4 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151508.
(DOCX)

S5 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151509.
(DOCX)

S6 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151669.
(DOCX)

S7 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151670.
(DOCX)

S8 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151671.
(DOCX)

S9 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151673.
(DOCX)

S10 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151674.
(DOCX)

S11 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151675.
(DOCX)

S12 Fig. Annotation of DLPFC layers and identification of spatial domains by multiple methods in slide 151676.
(DOCX)

S13 Fig. Analysis of gene numbers in brain cortex pathways for STESH, Spatial-MGCN, and GraphST in slice 151672.
(DOCX)

S1 Table. ARI value for DLPFC and HBC dataset.
(XLSX)

S2 Table. NMI value for DLPFC and HBC dataset.
(XLSX)

S3 Table. FMI value for DLPFC and HBC dataset.
(XLSX)

S4 Table. Attention weights of 151510 across 200 epochs.
(XLSX)

S5 Table. Attention weights of 151672 across 200 epochs.
(XLSX)

Acknowledgments

We thank Jing Liu and Jiuju Luo for their assistance with visualization.

Author contributions

Conceptualization: Haojing Shao.

Formal analysis: Jiaxing Chang.

Funding acquisition: Haojing Shao.

Methodology: Jiaxing Chang.

Software: Huihui Zhang, Jiaxing Chang, Hang Yang.

Supervision: Xingtan Zhang, Zehua Chen, Kok Wai Wong.

Validation: Zirong Li.

Visualization: Huihui Zhang, Zirong Li, Yue Sun, Pinli Hu, Haoxiu Wang.

Writing – original draft: Huihui Zhang.

Writing – review & editing: Yonglin Ren, Kok Wai Wong.

References

1. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008:10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
2. Xu H, Fu H, Long Y, Ang KS, Sethi R, Chong K, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med. 2024;16(1):12. https://doi.org/10.1186/s13073-024-01283-x
3. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat Commun. 2022;13(1):1739. https://doi.org/10.1038/s41467-022-29439-6 PMID: 35365632
4. Yang XU, McCord RP. CoSTA: unsupervised convolutional neural network learning for spatial transcriptomics analysis. BMC Bioinformatics. 2021;22(1):397. https://doi.org/10.1186/s12859-021-04314-1
5. Li Z, Zhou X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. Genome Biol. 2022;23(1):168. https://doi.org/10.1186/s13059-022-02734-7 PMID: 35927760
6. Li J, Chen S, Pan X, Yuan Y, Shen H-B. Cell clustering for spatial transcriptomics data with graph neural networks. Nat Comput Sci. 2022;2(6):399–408. https://doi.org/10.1038/s43588-022-00266-5 PMID: 38177586
7. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol. 2021;39(11):1375–84. https://doi.org/10.1038/s41587-021-00935-2
8. Long Y, Ang KS, Li M, Chong KL, Sethi R, Zhong C, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun. 2023;14(1):1155. https://doi.org/10.1038/s41467-023-36796-3
9. Liu T, Fang ZY, Li X, Zhang LN, Cao DS, Yin MZ. Graph Deep learning enabled spatial domains identification for spatial transcriptomics. Brief Bioinform. 2023;24(3):bbad146. https://doi.org/10.1093/bib/bbad146
10. Li X, Huang W, Xu X, Zhang HY, Shi Q. Deciphering tissue heterogeneity from spatially resolved transcriptomics by the autoencoder-assisted graph convolutional neural network. Front Genet. 2023;14:1202409. https://doi.org/10.3389/fgene.2023.1202409
11. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021;18(11):1342–51. https://doi.org/10.1038/s41592-021-01255-8 PMID: 34711970
12. Pham D, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. biorxiv. 2020. https://doi.org/10.1101/2020.05.31.125658
13. Shan Y, Zhang Q, Guo W, Wu Y, Miao Y, Xin H, et al. TIST: transcriptome and histopathological image integrative analysis for spatial transcriptomics. Genomics Proteomics Bioinformatics. 2022;20(5):974–88. https://doi.org/10.1016/j.gpb.2022.11.012 PMID: 36549467
14. Xu C, Jin X, Wei S, Wang P, Luo M, Xu Z, et al. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Res. 2022;50(22):e131. https://doi.org/10.1093/nar/gkac901 PMID: 36250636
15. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. 2021;40(11):e107333. https://doi.org/10.15252/embj.2020107333 PMID: 33950524
16. Maynard KR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci. 2021;24(3):425–36. https://doi.org/10.1038/s41593-020-00787-0
17. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell. 2022;185(10):1777-1792.e21. https://doi.org/10.1016/j.cell.2022.04.003 PMID: 35512705
18. Hao Y, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304. https://doi.org/10.1038/s41587-023-01767-y
19. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. https://doi.org/10.1186/s13059-017-1382-0 PMID: 29409532
20. Wang B, Luo J, Liu Y, Shi W, Xiong Z, Shen C, et al. Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. Brief Bioinform. 2023;24(5):bbad262. https://doi.org/10.1093/bib/bbad262 PMID: 37466210
21. Zhao E d w a r d, et al. BayesSpace enables the robust characterization of spatial gene expression architecture in tissue sections at increased resolution. bioRxiv. 2020.
22. Ren H, Walker BL, Cang Z, Nie Q. Identifying multicellular spatiotemporal organization of cells with SpaceFlow. Nat Commun. 2022;13(1):4076. https://doi.org/10.1038/s41467-022-31739-w PMID: 35835774
23. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. 2023;1(1):1–10.
24. Lin Y, Xu J, Lan H. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. J Hematol Oncol. 2019;12(1):76. https://doi.org/10.1186/s13045-019-0760-3 PMID: 31300030
25. Cho C-S, Xi J, Si Y, Park SR, Hsu JE, Kim M, et al. Microscopic examination of spatial transcriptome using seq-scope. Cell. 2021;184(13):3559–3572.e22. https://doi.org/10.1016/j.cell.2021.05.010