

RESEARCH ARTICLE

# SpaMWGDA: Identifying spatial domains of spatial transcriptomes using multi-view weighted fusion graph convolutional network and data augmentation

Lin Yuan  <sup>1,2,3</sup>, Boyuan Meng <sup>1,2,3</sup>, Qingxiang Wang <sup>1,2,3</sup>, Chunyu Hu <sup>1,2,3</sup>, Cuihong Wang  <sup>4\*</sup>, De-Shuang Huang  <sup>5,6\*</sup>

**1** Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, **2** Shandong Engineering Research Center of Big Data Applied Technology, Faculty of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, **3** Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, China, **4** The Second Qilu Hospital of Shandong University, Jinan, China, **5** Ningbo Institute of Digital Twin, Ningbo Key Laboratory of Multi-Omics & Multimodal Biomedical Data Mining and Computing, Eastern Institute of Technology, Ningbo, China, **6** Collaborative Innovation Center (2011) for Medical and Health Big Data of Guizhou Province, School of Biology and Engineering (School of Modern Industry for Health and Medicine), Guizhou Medical University, Guiyang, China

\* [xqawc1120@163.com](mailto:xqawc1120@163.com) (CW) [dshuang@eitech.edu.cn](mailto:dshuang@eitech.edu.cn) (DSH)



## OPEN ACCESS

**Citation:** Yuan L, Meng B, Wang Q, Hu C, Wang C, Huang D-S (2025) SpaMWGDA: Identifying spatial domains of spatial transcriptomes using multi-view weighted fusion graph convolutional network and data augmentation. PLoS Comput Biol 21(11): e1013667. <https://doi.org/10.1371/journal.pcbi.1013667>

**Editor:** Guang-Zhong Wang, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, CHINA

**Received:** July 4, 2025

**Accepted:** October 27, 2025

**Published:** November 12, 2025

**Copyright:** © 2025 Yuan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** The code and data of SpaMWGDA is available at <https://github.com/nathanyl/SpaMWGDA>.

**Funding:** LY is supported by the National Natural Science Foundation of China (Nos.

## Abstract

The rapid development of spatial transcriptomics (ST) has made it possible to effectively integrate gene expression and spatial information of cells and accurately identify spatial domains. A large number of deep learning (DL)-based methods have been proposed to perform spatial domain identification and achieved impressive results. However, these methods have some limitations. First, these methods rely on a fixed similarity metric and cannot fully utilize neighborhood information. Second, they cannot efficiently and adaptively integrate key information when fusing and reconstructing gene expression using purely additive methods. Finally, these methods ignore key nonlinear features and introduce noise during clustering. To address these limitations, we propose a novel DL model SpaMWGDA based on multi-view weighted fused graph convolutional network (GCN) and data augmentation. By modeling spatial information using different similarity metrics, the model is able to successfully capture comprehensive neighborhood information of the spot features. By combining data augmentation and contrastive learning, SpaMWGDA is able to learn key gene expressions. SpaMWGDA uses a multi-view GCN encoder to model the similarities between spatial information and gene features, and uses a view-level attention mechanism for weighted fusion to adaptively learn the dependencies between them and learn the key features of each view. Experimental results not only demonstrate that SpaMWGDA outperforms competing methods in spatial domain identification and

62472239, 62002189), the Shandong Provincial Natural Science Foundation (No. ZR2024MF011), the Ability Improvement Project of Science and Technology SMES in Shandong Province (2023TSGC0279), the Youth Innovation Team of Colleges and Universities in Shandong Province (2023KJ329). CHW is supported by the Cultivation Fund of the Second Hospital of Shandong University (No. 2023JX16), supported by the Shandong Medical Association Qilu medical special project (YKH2022K02112), and supported by the Shandong Province Key Research and Development Program—International Scientific and Technological Cooperation Project (2024KJHZ029). DSH is supported by grants from the National Natural Science Foundation of China, Nos. 62333018, 62372255, 62073231, partly supported by the Joint Project of National Natural Science Foundation of China and Russian Science Foundation (W2412087), and supported by the Natural Science Foundation of Guizhou Province, No. ZK2024ZD035, and supported by the Natural Science Foundation of Ningbo City under Grant No.2023J199, and supported by Key Research and Development (Digital Twin) Program of Ningbo City under Grant Nos.2023Z219, 2023Z226. 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.

trajectory inference but also show the ability of SpaMWGDA to analyse tissue structure and function. The source code for SpaMWGDA is available at <https://github.com/nathanyl/SpaMWGDA>.

## Author summary

DL-based methods have some limitations. First, these methods rely on a fixed similarity metric and cannot fully utilize neighborhood information. Second, they cannot efficiently and adaptively integrate key information when fusing and reconstructing gene expression using purely additive methods. Finally, these methods ignore key nonlinear features and introduce noise during clustering. We propose a novel DL model SpaMWGDA based on multi-view weighted fused graph convolutional network (GCN) and data augmentation. SpaMWGDA uses a multi-view GCN encoder to model the similarities between spatial information and gene features, and uses a view-level attention mechanism for weighted fusion to adaptively learn the dependencies between them and learn the key features of each view. Experimental results not only demonstrate that SpaMWGDA outperforms competing methods in spatial domain identification and trajectory inference but also show the ability of SpaMWGDA to analyse tissue structure and function.

## Introduction

Spatial transcriptomics (ST) technologies [1–5] can pinpoint gene expression while maintaining the structural integrity of tissues, helping us to understand molecular communication and tissue structure. Spatial domains are areas with similar spatial gene expression distribution.

Precise identification of the spatial domain is a crucial step in ST data analysis. It helps to elucidate the complex relationships between gene expression, its spatial characteristics, and tissue functions, and also aids in understanding the distribution and interactions of cells within the tissue structure [6–8]. However, accurately identifying spatial domains remains challenging due to the frequent presence of loss and noise in the generated data. Many computational methods have been developed for spatial domain identification. The shallow learning algorithms (e.g., K-means, Louvain [9] and Leiden [10]) are often used in Scanpy [11] or Seurat [12] packages to form integrated analysis workflows. The spatial domains identified by these methods are usually discontinuous because they ignore the adjacent relationships between spatial domains [13]. Giotto [14] identified spatial domains by comparing the intrinsic gene expression patterns of neighboring cells using hidden Markov random field (HMRF). The method is computationally complex and difficult to deal with large-scale datasets. In addition, HMRF may have limitations in capturing nonlinear relationships between cells. BayesSpace [15], based on a fully Bayesian statistical approach, encouraged

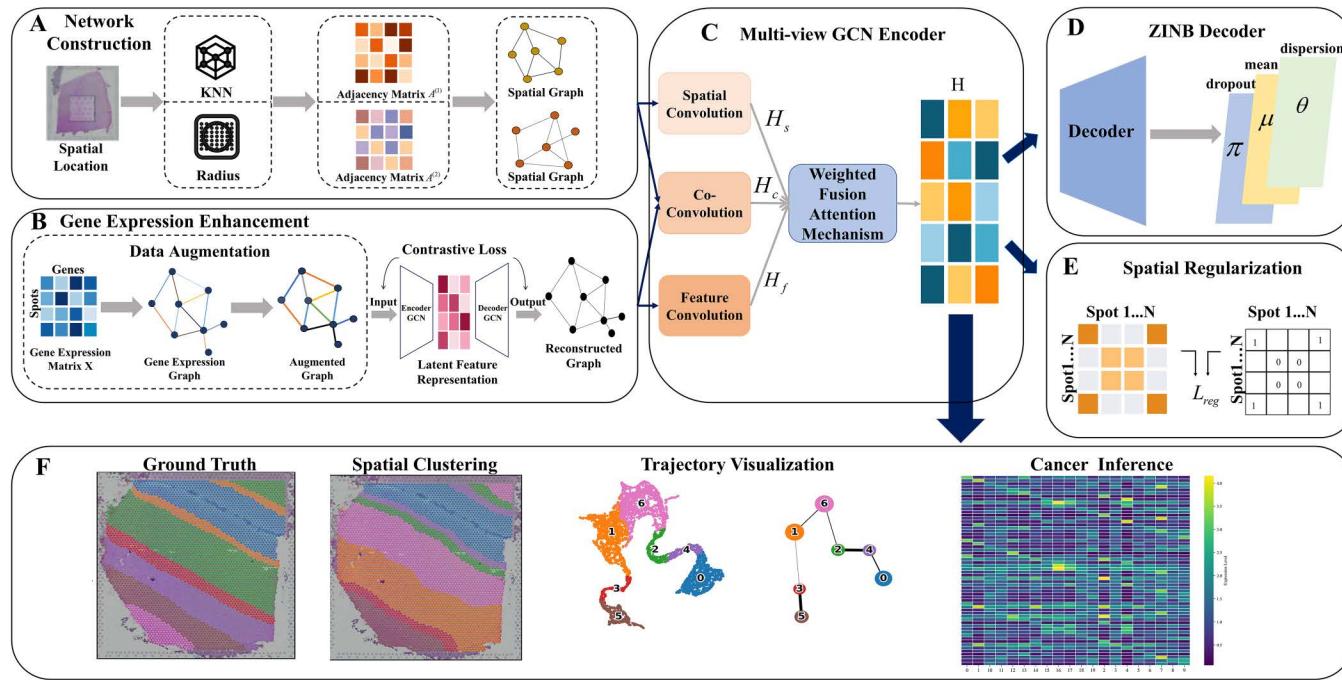
neighboring cells to belong to the same group via predefined spatial priors. However, BayesSpace fails to effectively utilize spatial coordinates, and the choice of spatial priors may affect the clustering results.

Recently, deep learning (DL)-based spatial clustering methods have emerged as powerful tools for spatial domain identification [16–24]. For example, stlearn [25] efficiently performed spatial domain identification by integrating gene expression, spatial location, and tissue morphology. spaGCN [26] employed an undirected weighted graph to represent spatial data dependencies and incorporated spatial location, histological images, and gene expression into the graph's construction to identify spatial domains. DeepST [27] used a denoising autoencoder and a graph neural network (GNN) autoencoder to infer latent embeddings in enhanced ST data. However, noise and inaccurate spot relationships in histological images may lead to erroneous spatial domain identification results. To better utilize high-resolution ST data, researchers considered the spatial dependencies of gene expression by modeling the similarity between neighboring spots. For example, SEDR [28] used a deep autoencoder and a variational graph autoencoder to integrate transcriptomics data with relevant spatial information. STAGATE [29] employed a graph attention autoencoder to integrate spatial information and gene expression, learning latent representations to distinguish spatial domains. GraphST [30] combined GNNs with self-supervised contrastive learning to identify spatial domains by learning spot representations. Although these methods can identify spots as distinct regions, they cannot adaptively capture the interrelationships between gene expression and spatial information. Recently, multi-view graph convolutional network (GCN)-based methods have been proposed to learn the relationship between spatial location and gene expression more effectively [31]. For example, Spatial-MGCN [32] utilized a multi-view GCN encoder to identify spatial domains. STMGCN [33] introduced an unsupervised learning framework that learns view-specific spot representations and integrates them with an attention mechanism to generate the final representation of the spots. But multi-view convolutional networks encounter several challenges, including noise and inconsistent feature distributions across views, which can hinder effective fusion and compromise stability. Additionally, certain views may contribute minimally, and inappropriate fusion strategies may introduce redundancy, further affecting model performance.

Although these methods have achieved remarkable results, there are still some limitations. First, these methods rely on a fixed similarity metric and cannot fully utilize neighborhood information. Second, they cannot efficiently and adaptively integrate key information when fusing and reconstructing gene expression using purely additive methods. Finally, these methods ignore key nonlinear features and introduce noise during clustering [34,35].

In this article, we are interested in integrating multi-view weighted fusion graph convolutional network (GCN) and data augmentation to guide deep learning architecture to accurately identify spatial domains. Multi-view Graph Convolutional Networks enhance the analytical accuracy and generalization capabilities of spatial transcriptomics data by integrating data from multiple perspectives to augment the model's expressive power. The weighted fusion attention mechanism dynamically adjusts feature weights to identify critical regions or genes, thereby improving prediction precision. Data augmentation enhances model robustness by generating new samples, particularly in scenarios with sparse or imbalanced data, further boosting generalization performance.

Here, we name the proposed spatial domain identification method as SpaMWGDA. The workflow of SpaMWGDA is presented in [Fig 1](#). First, SpaMWGDA employs KNN and Radius to construct a neighborhood graph [36,37], which can help the model successfully capture the comprehensive neighborhood information of point features. Second, SpaMWGDA learns key feature representations from original and augmented gene expression data by constructing graph-embedded contrastive encoder [38,39]. Third, SpaMWGDA integrates a multi-view GCN encoder and a ZINB (Zero-Inflated Negative Binomial) decoder [40,41] to reconstruct the gene expression matrix. In multi-view GCN encoder, the weighted fusion attention mechanism dynamically adjusts the weight of each view, enabling the model to better adapt to the contribution of different views, thus improving model performance. The view-level attention mechanism can adaptively and effectively integrate information from multiple views, thereby improving the performance and robustness of the model. Finally, SpaMWGDA uses a spatial regularization constraint to train the model to cluster neighboring points in space and effectively separates spatially non-neighboring points. We compared the performance of SpaMWGDA with seven state-of-the-art



**Fig 1. Schematic overview of SpaMWGDA.** (A) Spatial neighborhood network construction module. (B) Gene expression enhancement module. (C) Multi-view weighted fusion GCN encoder. (D) ZINB decoder. (E) Spatial regularization constraint. (F) Downstream analysis.

<https://doi.org/10.1371/journal.pcbi.1013667.g001>

methods on five datasets (three from 10 x Visium, one from Stereo-seq platform and one from Xenium). Experimental results not only indicate that SpaMWGDA outperforms seven state-of-the-art methods in spatial domain identification and trajectory inference but also demonstrate the ability of SpaMWGDA to analyse tissue structure and function.

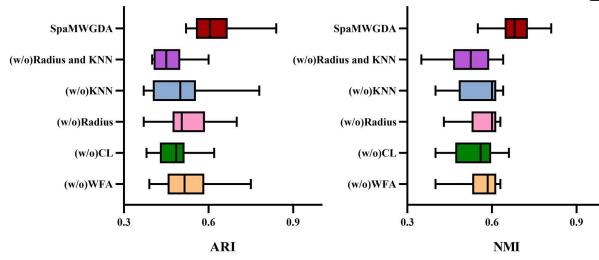
## Results

### Ablation experiments

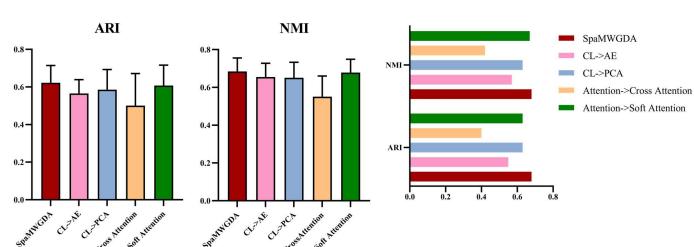
To explore the contribution of key modules to the performance of SpaMWGDA, we constructed five variants of SpaMWGDA and performed ablation experiments on DLPFC datasets. The five variant models are: (i) (w/o)Radius; (ii)(w/o)KNN; (iii) (w/o)CL; (iv) (w/o)WFA and (v) (w/o)KNN and Radius. (w/o)Radius represents that SpaMWGDA only uses KNN (without Radius) to identify neighboring points. (w/o)KNN represents that SpaMWGDA only uses Radius (without KNN) to identify neighboring points. (w/o)CL represents that SpaMWGDA only use original gene expression to obtain feature representation without using contrastive learning. (w/o)WFA represents that SpaMWGDA uses equal weights instead of dynamic weights in the attention layer. (w/o)KNN and Radius represents that SpaMWGDA uses a distance-based similarity matrix to model spatial information.

As shown in Fig 2A, compared to SpaMWGDA, the ARI and NMI of (w/o)Radius, (w/o)KNN, (w/o)CL, (w/o)WFA and (w/o)KNN and Radius decreased by 16.1% and 16.2%, 17.7% and 17.6%, 21% and 20.6%, 16.1% and 19.1%, and 37.1% and 17.6%, respectively. The ARI and NMI of each module were listed in Table 1. The experimental results highlight the importance of constructing spatial graphs using different similarity metrics, contrastive learning using augmented gene features, and using weighted fusion attention mechanisms. These modules can help the model better identify the interaction between spatial information and genetic features, improve the accuracy of spatial domain identification, and ultimately improve model performance.

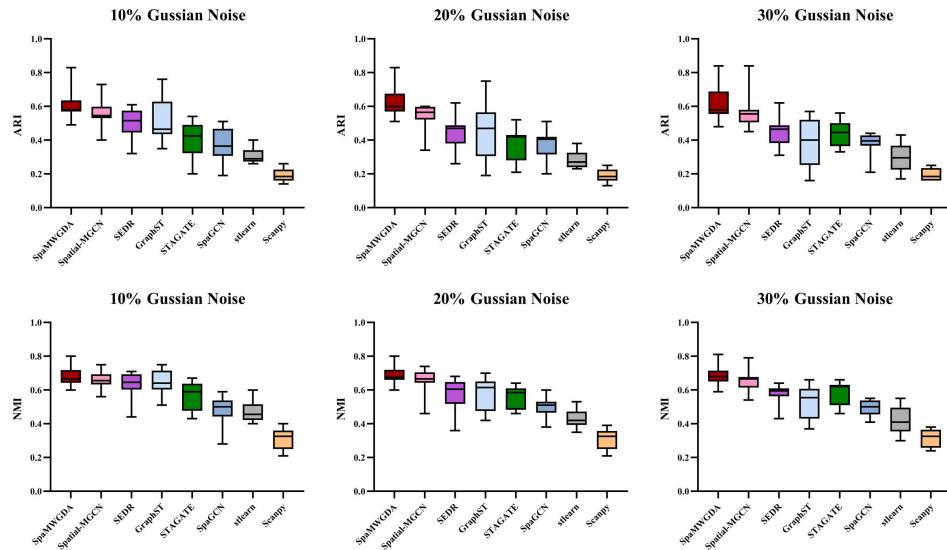
A



B



C



**Fig 2. (A) Ablation experiment results. (B)** The performance of contrastive learning and weighted fusion attention on SpaMWGDA. **(C)** The results of SpaMWGDA and seven competing methods on the noisy DLPFC dataset.

<https://doi.org/10.1371/journal.pcbi.1013667.g002>

### The performance of contrastive learning and weighted fusion attention on model

In order to further evaluate the impact of contrastive learning and weighted fusion attention on the model, we constructed four variants of SpaMWGDA: (i) CL->PCA; (ii) CL->AE [42], (iii) Attention-> Cross Attention; and (iv) Attention-> Soft Attention [43]. Principal component analysis (PCA) and autoencoder (AE) are widely used feature learning methods. CL->PCA indicates using PCA instead of contrastive learning. and CL->AE replaces the contrastive learning with autoencoder. The weighted fusion attention mechanism is replaced by the cross-attention mechanism and the soft attention mechanism respectively to evaluate the impact of the weighted fusion attention mechanism on SpaMWGDA.

We compared the performance of these four variants with SpaMWGDA on the DLPFC and human breast cancer datasets. As shown in the Fig 2B, SpaMWGDA outperforms these variants in terms of ARI and NMI. Contrastive learning enables SpaMWGDA to learn critical and discriminative nonlinear features. The adaptive property of the weighted fusion attention mechanism is particularly beneficial in integrating information from different views, helping to extract features that are critical for spatial clustering, thereby improving the performance of the model.

PCA is a linear dimensionality reduction method that cannot capture nonlinear relationships. Autoencoder is effective in learning latent representations, but may focus too much on global features during reconstruction and ignore subtle differences between cell types that are critical for spatial clustering. Cross attention mechanism uses multiple attentions, which increases computational complexity and may affect model performance. Soft attention mechanism assigns weights to input features to focus on important information. However, it may over-rely on global features and ignore local details.

**Table 1.** The ARI and NMI of four variant models and SpaMWGDA.

	(w/o)WFA	(w/o)CL	(w/o)Radius	(w/o)KNN	(w/o)KNN and Radius	SpaMWGDA
ARI	0.60	0.61	0.54	0.57	0.5	0.73
	0.54	0.45	0.37	0.38	0.4	0.66
	0.52	0.50	0.50	0.50	0.45	0.67
	0.51	0.49	0.47	0.48	0.42	0.53
	0.75	0.62	0.70	0.78	0.6	0.55
	0.60	0.52	0.60	0.65	0.5	0.52
	0.47	0.47	0.62	0.47	0.45	0.62
	0.52	0.50	0.51	0.51	0.5	0.84
	0.45	0.38	0.50	0.50	0.5	0.60
	0.39	0.39	0.37	0.37	0.4	0.57
	0.48	0.48	0.55	0.50	0.45	0.61
	0.40	0.42	0.48	0.38	0.4	0.57
AVG.	0.52	0.49	0.52	0.51	0.39	0.62
NMI	0.63	0.60	0.60	0.60	0.6	0.78
	0.59	0.55	0.43	0.49	0.49	0.72
	0.62	0.64	0.62	0.62	0.62	0.69
	0.40	0.45	0.63	0.64	0.64	0.66
	0.63	0.58	0.55	0.60	0.6	0.55
	0.54	0.40	0.52	0.48	0.48	0.60
	0.53	0.59	0.57	0.60	0.6	0.73
	0.60	0.61	0.63	0.63	0.63	0.81
	0.58	0.53	0.60	0.60	0.6	0.68
	0.40	0.43	0.49	0.40	0.4	0.64
	0.59	0.57	0.60	0.55	0.55	0.68
	0.53	0.54	0.60	0.45	0.45	0.67
AVG.	0.55	0.54	0.57	0.56	0.56	0.68

<https://doi.org/10.1371/journal.pcbi.1013667.t001>

### The impact of noise on model robustness and the scalability of SpaMWGDA

To assess the impact of noise on the performance of SpaMWGDA, we constructed three kinds of random gaussian noise datasets based on the DLPFC dataset: (i) Gaussian Noise (10%); (ii) Gaussian Noise (20%); (iii) Gaussian Noise (30%). Gaussian Noise (10%), Gaussian Noise (20%), and Gaussian Noise (30%) respectively represent adding 10%, 20%, and 30% Gaussian noise to the original gene expression data, respectively. We compared the performance of SpaMWGDA with state-of-the- art methods (Scanpy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN) on three noisy datasets. As shown in [Fig 2C](#), although the performance of SpaMWGDA decreases due to the increase of noise, it still outperforms the competing methods. The results of SpaMWGDA and seven competing methods on the Gaussian Noise 10% DLPFC dataset were listed in [Table 2](#), and the results of the remaining noise dataset were listed in [S1 Table](#). These results showed that SpaMWGDA is robust to noise and its performance remains excellent even as noise increases.

We calculated the running time of SpaMWGDA under different data scales and spot counts. The results in [Table 3](#) show the scalability of SpaMWGDA on large-scale datasets.

### Performance comparison of methods for identifying spatial domain

To evaluate the performance of SpaMWGDA in spatial domain identification, we conducted a comparison with seven state-of-the-art methods: Scanpy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN, using

**Table 2. The ARI and NMI of SpaMWGDA and seven competing methods on Gaussian Noise 10% DLPFC dataset.**

ARI	SpaMWGDA	Spatial-MGCN	STAGATE	SEDR	GraphST	SpaGCN	stLearn	Scanpy
151507	0.74	0.66	0.43	0.49	0.45	0.47	0.40	0.23
151508	0.65	0.54	0.47	0.49	0.52	0.35	0.34	0.26
151509	0.59	0.53	0.36	0.61	0.43	0.39	0.40	0.25
151510	0.50	0.56	0.42	0.46	0.45	0.33	0.34	0.21
151669	0.57	0.44	0.24	0.44	0.35	0.23	0.28	0.14
151670	0.49	0.40	0.20	0.32	0.42	0.19	0.28	0.16
151671	0.58	0.61	0.49	0.56	0.65	0.46	0.29	0.18
151672	0.83	0.73	0.53	0.54	0.76	0.51	0.27	0.14
151673	0.59	0.56	0.54	0.58	0.65	0.47	0.26	0.21
151674	0.57	0.54	0.36	0.59	0.48	0.33	0.29	0.19
151675	0.58	0.55	0.31	0.56	0.56	0.30	0.30	0.16
151676	0.57	0.53	0.49	0.38	0.45	0.38	0.26	0.16
<b>AVG.</b>	0.61	0.55	0.40	0.50	0.51	0.37	0.31	0.19
NMI	SpaMWGDA	Spatial-MGCN	STAGATE	SEDR	GraphST	SpaGCN	stLearn	Scanpy
151507	0.77	0.75	0.59	0.63	0.66	0.54	0.52	0.36
151508	0.72	0.67	0.63	0.61	0.64	0.47	0.53	0.36
151509	0.64	0.67	0.59	0.70	0.61	0.53	0.60	0.40
151510	0.65	0.65	0.59	0.62	0.64	0.49	0.50	0.33
151669	0.62	0.57	0.47	0.56	0.51	0.40	0.42	0.21
151670	0.60	0.56	0.43	0.44	0.56	0.28	0.41	0.24
151671	0.71	0.74	0.64	0.67	0.75	0.53	0.46	0.29
151672	0.80	0.70	0.66	0.66	0.74	0.58	0.40	0.24
151673	0.68	0.64	0.67	0.71	0.73	0.59	0.45	0.38
151674	0.65	0.66	0.50	0.71	0.63	0.44	0.43	0.32
151675	0.66	0.65	0.45	0.67	0.67	0.45	0.46	0.33
151676	0.67	0.63	0.60	0.60	0.60	0.51	0.44	0.28
<b>AVG.</b>	0.68	0.66	0.57	0.63	0.65	0.48	0.47	0.31

<https://doi.org/10.1371/journal.pcbi.1013667.t002>

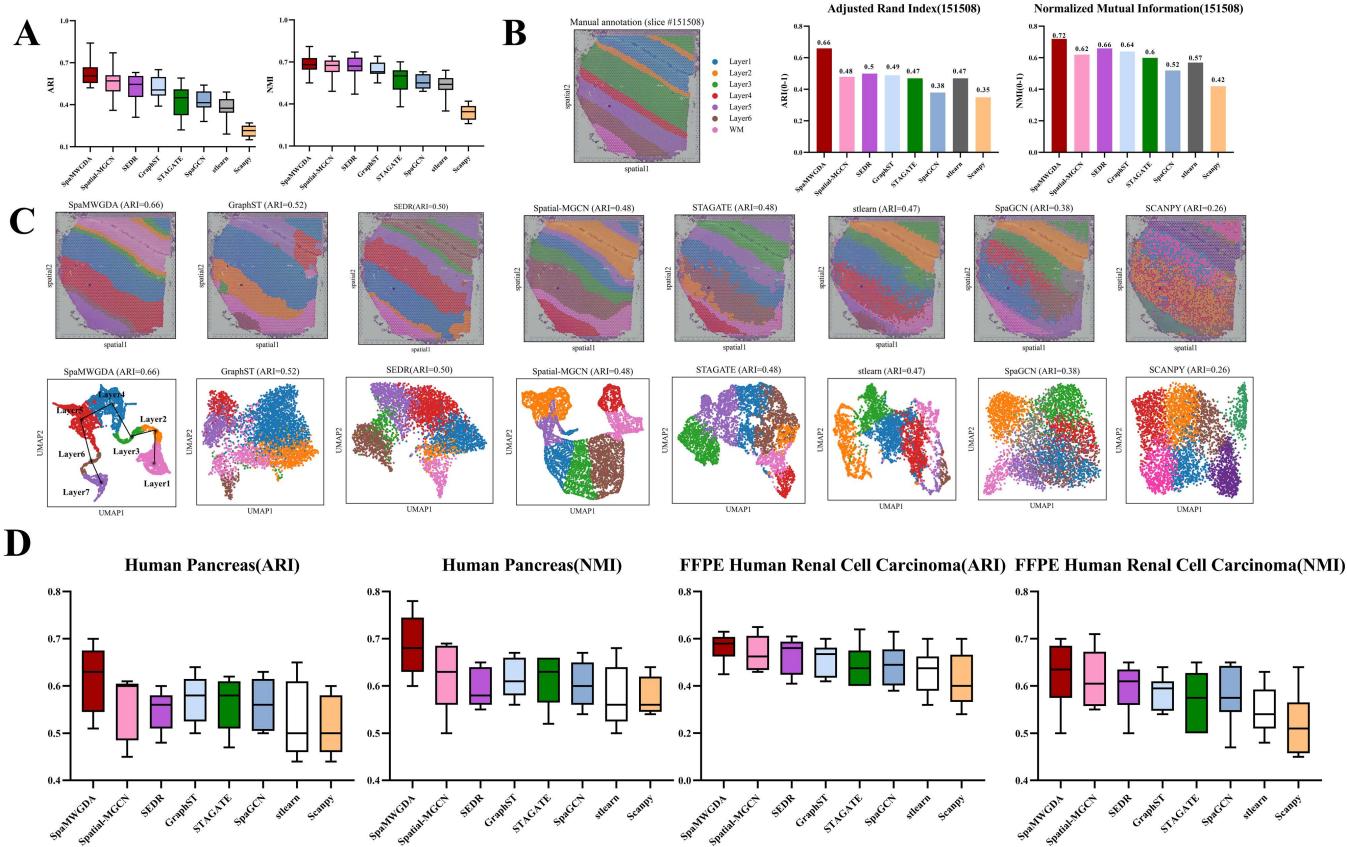
**Table 3. Running time of SpaMWGDA under different data scales and spot counts.**

Data Scale	10%	20%	30%	40%	50%
Running Time(minute)	6mins	12mins	18mins	24mins	30mins
Spot Scale	10%	20%	30%	40%	50%
Running Time(minute)	2mins	5mins	8mins	10mins	12mins

<https://doi.org/10.1371/journal.pcbi.1013667.t003>

the 10x Visium DLPFC dataset, which contains 12 slices [44]. These seven methods include two shallow learning methods (Scanpy, stlearn) and five state-of-the-art DL-based methods (SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN).

As shown in Fig 3A, SpaMWGDA achieves the best clustering across these slices, with the highest mean ARI of 0.62 and the highest mean NMI of 0.68, surpassing Spatial-MGCN, SEDR, STAGATE and GraphST by 0.06 and 0.02, 0.1 and 0.01, 0.13 and 0.05, and 0.1 and 0.04, respectively. The mean ARI and mean NMI of other methods (Scanpy, stlearn and SpaGCN) are below 0.45 and 0.6, respectively. Detailed results for all 12 slices in the DLPFC dataset were shown in S1 Fig. It is important to note that while Scanpy exhibits minimal variation in ARI between slices, its median NMI is only 0.35. In contrast, the performance of Spatial-MGCN and STAGATE shows instability, with significant ARI variation across slices.



**Fig 3. (A)** The performance comparison of spatial domain identification of SpaMWGDA and seven state-of-the-art methods (Scampy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN) on DLPFC dataset. **(B)** Comparison of ARI and NMI of SpaMWGDA with seven state-of-the-art methods (Scampy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN) on DLPFC slice 151508. **(C)** The performance comparison of spatial domain identification of SpaMWGDA and seven state-of-the-art methods (Scampy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN) on DLPFC slice 151508. **(D)** The performance comparison of SpaMWGDA and seven state-of-the-art methods on Human Pancreas dataset and FFPE Human Renal Cell Carcinoma dataset.

<https://doi.org/10.1371/journal.pcbi.1013667.g003>

The mean values of three histology-informed spatial clustering methods (stlearn and SpaGCN) are all below 0.45, which indicates that histological images may introduce noise that adversely affects clustering performance.

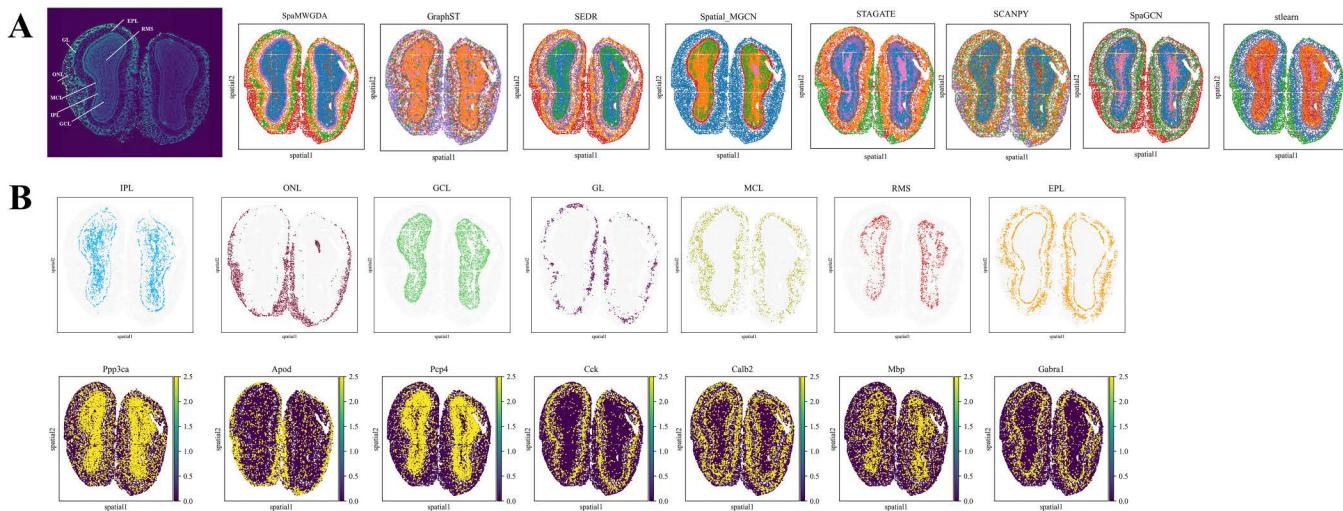
To further validate the effectiveness of SpaMWGDA, we show the identification results of DLPFC slice 151508. As shown in Fig 3B, compared to competing methods, SpaMWGDA achieves a better fit with the ground truth, with ARI and NMI of 0.66 and 0.72 respectively, while the ARI and NMI of competing methods are all below 0.6 and 0.7 respectively. As shown in Fig 3C, Scampy shows layer mixing, SEDR shows unclear cluster boundaries, Spatial-MGCN only achieves an ARI of 0.48, GraphST and STAGATE confuse spots from layer 3 and WM (White Matter), and both stlearn and SpaGCN fail to distinguish most DLPFC layers. In UMAP visualization analysis, the clustering divisions of Scampy and SpaGCN are not clear. Although STAGATE, SEDR, GraphST, and stlearn achieve good clustering results, the hierarchical order is unclear. The embeddings of both SpaMWGDA and Spatial-MGCN clearly show the cortical development trajectories, and the UMAP map of SpaMWGDA not only reveals the different clustering divisions in each domain, but also highlights the clear sequential relationships between layers. These results indicate the excellent clustering performance of SpaMWGDA and highlight the advantage of SpaMWGDA in identifying spatial domains. In addition, we also compared the performance of SpaMWGDA and seven competing methods on the Human Pancreas dataset from 10x Visium HD and FFPE Human Renal Cell Carcinoma dataset from 10x Xenium. The results show that SpaMWGDA is better than other methods.

## SpaMWGDA reveals the laminar structure of the olfactory bulb from Stereo-seq data and infers its biological functions

Stereo-seq technology provides highly detailed spatial and molecular data for spatial transcriptomics studies of the mouse olfactory bulb (MOB) [45], enabling the evaluation of models' ability to explore biological functions. For comparative studies, we selected Scanpy, GraphST, SEDR, STAGATE, Spatial-MGCN, stlearn, and SpaGCN as benchmark methods to assess the spatial domain identification capability of SpaMWGDA.

As shown in [Fig 4A](#), Scanpy does not clearly delineate the partitions, and GraphST and Spatial-MGCN confuse different regions. SEDR, stlearn, STAGATE, SpaGCN, and SpaMWGDA all accurately capture the hierarchical structure of the MOB. SEDR divides the region into seven layers. However, it fails to link the identified regions to any specific areas in the histological images and fails to detect the narrow glomerular layer. STAGATE, stlearn and SpaGCN segment the MOB into seven regions, most of which are accurately identified, but there is still considerable overlap between clusters. SpaMWGDA exploits a multi-view weighted fusion GCN encoder to deeply explore the intrinsic relationship between gene expression and spatial information, thereby obtaining more informative and discriminative latent representations.

We then performed a detailed analysis of the differentially expressed genes (DEGs) between the coronal tissue layers of the mouse olfactory bulb identified by SpaMWGDA. As shown in [Fig 4B](#), the results revealed that the identified clusters were strongly correlated with the expression of well-known marker genes such as Apod and Cck [46]. The olfactory nerve layer (ONL) corresponds to the expression domain of the Apod gene, which is involved in encoding components of high-density lipoproteins, suggesting that the ONL may be linked to lipoprotein metabolism. The Cck gene encodes cholecystokinin (CCK), which modulates neuronal excitability and synaptic transmission and plays a key role in the processing and feedback regulation of olfactory signals [47,48]. CCK has an important influence on the glomerular layer (GL), which is responsible for the initial integration of olfactory information. CCK enhances the transmission and perception of olfactory signals by regulating the activity of neurons in this region. These findings not only demonstrate the ability of SpaMWGDA to handle tissue structures at different resolutions, but also indicate that SpaMWGDA can infer potential tissue functions from identified spatial domains, offering valuable insights for further exploration of unknown tissue structures and functions.



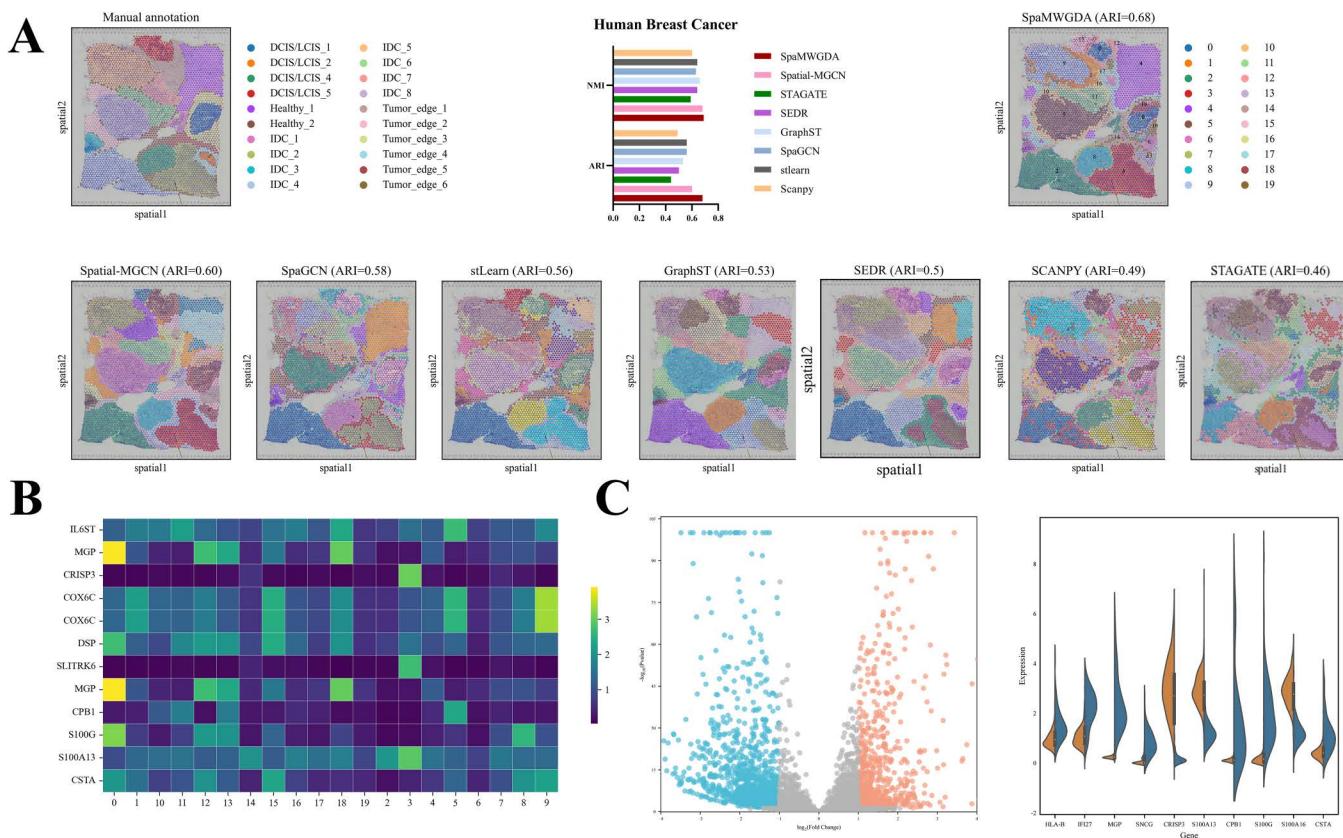
**Fig 4. (A) Comparison of the results of identifying the laminar structure of the olfactory bulb using SpaMWGDA and seven state-of-the-art methods (Scanpy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN). (B) Analysis results of differentially expressed genes (DEGs) between the coronal tissue layers of the mouse olfactory bulb identified by SpaMWGDA.**

<https://doi.org/10.1371/journal.pcbi.1013667.g004>

## SpaMWGDA delivers detailed insights into tumor heterogeneity in human breast cancer

In this section, we analysed the human breast cancer dataset from 10x Visium platform, which contains 20 domains and 4 main tissue types: DCIS/LCIS, healthy tissues, IDC, and hypomalignant tumor margins. As shown in [Fig 5A](#), SpaMWGDA achieves the highest ARI and NMI among all methods. Compared to competing methods, SpaMWGDA can consistently identify domains that align with manual annotations and accurately detect domains such as Healthy\_1 and IDC\_5, with smoother boundaries for each domain. In contrast, the boundaries of regions identified by Scanpy are highly irregular and contain a large amount of noise. Although Spatial-MGCN, SEDR, SpaGCN, stlearn, GraphST, and STAGATE identify more domains than Scanpy, these domains still have rough boundaries and outliers.

We further validated the performance of SpaMWGDA in detecting cancer tissue heterogeneity using the human breast cancer dataset. We compared the expression of the top differentially expressed genes (DEGs) in clusters 4 (healthy), 13 (DCIS/LCIS), 3 (IDC), and 15 (tumor margins), and found significant heterogeneity among these clusters ([Fig 5B](#)). We also performed differential expression analysis ( $|logFoldChange| \geq 2$  and  $P\text{-value} < 0.05$ ) between cluster 3 (IDC) and cluster 13 (DCIS/LCIS) to explore gene expression differences between IDC and DCIS/LCIS, and identified about 100 significant DEGs between these two clusters ([Fig 5C](#)). As shown in [Fig 5C](#), among the significant differentially expressed genes (DEGs) identified, CRISP3 [49] plays a pivotal role in breast cancer by regulating tumor cell migration, invasion,



**Fig 5. (A)** The performance comparison of spatial domain identification of SpaMWGDA and seven state-of-the-art methods (Scanpy, stlearn, SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN) on human breast cancer dataset. **(B)** Results of differentially expressed genes analysis among different clusters identified by SpaMWGDA. **(C)** Results of differentially expressed genes analysis between cluster 3 (IDC) and cluster 13 (DCIS/LCIS).

<https://doi.org/10.1371/journal.pcbi.1013667.g005>

and immune modulation. It contributes to the tumor microenvironment through inflammatory pathways and immune evasion mechanisms. Additionally, MGP (Matrix Gla Protein), a calcium-binding protein, is integral to extracellular matrix remodeling and mineralization, and its elevated expression is associated with tumor aggressiveness and adverse prognosis. According to the studies [27,50], we found that the upregulation of CRISP3, S100A13, and S100A16 in domain 3 suggests that this region possesses tumor invasiveness, metastatic potential, and an active inflammatory environment, thereby promoting tumor progression. In contrast, domain 13, with high expression of MGP, CPB1, and S100G, points to a region associated with calcification, ECM remodeling, and tumor progression, which may contribute to treatment resistance and poor clinical outcomes.

These findings highlight the critical roles of CRISP3 and MGP in breast cancer pathogenesis, positioning them as valuable prognostic biomarkers and potential therapeutic targets. The distinct molecular profiles observed in domains 3 and 13 underscore the regional heterogeneity of the tumor microenvironment and the need for region-specific therapeutic strategies. In summary, SpaMWGDA offers a more refined approach to dissect cancer tissue heterogeneity and enhances our understanding of ST data.

## Discussion

In this paper, we proposed a novel deep learning model, SpaMWGDA, combined a multi-view weighted fused graph convolutional network and data augmentation for spatial domain identification of spatial transcriptome (ST) data. By leveraging multi-view weighted fusion and data enhancement, SpaMWGDA effectively learns the relationships between spatial information and gene expression. Experimental results demonstrated that SpaMWGDA outperforms competing methods in terms of clustering accuracy and identification of biologically relevant domains, and exhibits superior performance in spatial domain identification. Additionally, SpaMWGDA successfully evaluated ST data from various platforms with different spatial resolutions. Experiment results highlighted the importance of integrating different similarity metrics for exploring spatial information, as well as reconstructing gene expression through data enhancement and weighted fusion attention mechanism.

However, SpaMWGDA still has limitations. For instance, in experiments involving model noise injection, we can employ various noise addition methods such as signal-to-noise ratio (SNR) to validate the model's robustness from different perspectives. When considering the use of different similarity measures to construct spatial neighborhood graphs, we can experiment with alternative metrics, such as the cosine similarity measure based on vectors. Utilizing spatial multi-omics data [51] will enable the model to more accurately resolve the spatial domain, which is vital for inferring the biological functions of complex tissues in organisms. Therefore, spatial domain identification methods that effectively use spatial multi-omics data should be proposed in future work.

## Materials and methods

### Data preparation

To validate the performance of the model, we used five datasets from multiple platforms. As shown in [Table 4](#), the first dataset is from 10 x Visium, with each slice containing five to seven manually annotated regions of the human dorso-lateral prefrontal cortex (DLPFC) [44]. The second dataset is the ST dataset of human breast cancer from 10 x Visium, which contains 20 domains and four major morphological types: DCIS/LCIS (Ductal Carcinoma In Situ/Lobular Carcinoma In Situ), IDC (invasive ductal carcinoma), healthy tissues, and hypomalignant tumor margins [52]. The third dataset is a mouse olfactory bulb dataset obtained from Stereo-seq [53], which annotates the RMS (rostral migratory stream), GCL (granular cell layer), IPL (internal plexiform layer), MCL (mitral cell layer), EPL (external plexiform layer), and ONL (olfactory nerve layer). The fourth dataset is a Human Pancreas dataset from 10x Visium HD [54]. It offers spatial gene expression data from human pancreas tissue, enabling analysis of gene activity across different pancreatic regions. The fifth dataset is a FFPE Human Renal Cell Carcinoma dataset from 10x Xenium [55]. It provides high-resolution spatial

**Table 4. Overview of five ST datasets used in this study.**

Datasets	Spots	Genes	Slices	Domains	Platforms
DLPFC	3460-4789	33538	12	5/7	10x Visium
Human Breast Cancer	3798	36601	1	20	10x Visium
Mouse Olfactory Bulb	19109	14376	1	7	Stereo-seq
Human Pancreas	5000-10000	2000-5000	5	5	10x Visium HD
FFPE Human Renal Cell Carcinoma	More than 10000	10000	6	7	10x Xenium

<https://doi.org/10.1371/journal.pcbi.1013667.t004>

transcriptomic data from formalin-fixed, paraffin-embedded (FFPE) human renal cell carcinoma tissue. It allows for the exploration of gene expression patterns within the tumor microenvironment, aiding in the study of tumor heterogeneity and cellular interactions in cancer.

SpaMWGDA takes gene expression and spatial location as input. To reduce technical noise, spots outside the main tissue areas were first removed. Subsequently, SCANPY [11] was used to filter genes with low-expression and low-variance, eliminate genes that were not expressed in less than 100 cells, and select top 3,000 highly variable genes (HVGs). Finally, the expression data of HVGs were normalized to the total expression level of each cell to 10,000. The formula can be defined as follows:

$$x'_{ij} = \left( \frac{x_{ij}}{\sum_j x_{ij}} \right) \times 10000 \quad (1)$$

where  $x_{ij}$  ( $1 \leq i \leq N, 1 \leq j \leq M$ ) represents the j-th gene expression value at the i-th point.

## Experiment settings

For SpaMWGDA, we employed the learning rate of 0.001, the weight decay of 5e-4, and utilized the ADAM optimization algorithm. Additionally, to optimize the nearest neighbor search, we adopted the K-D-tree algorithm and tested various k values ranging from 1 to 20. Performance scores were calculated for each k value, enabling dynamic adjustment of k to enhance the quality of the neighborhood matrix. The radius r is set based on the data resolution. All experiments are repeated 10 times, and spatial domain recognition performance is evaluated using ARI and NMI. The average of 10 runs is taken to obtain a reasonable performance assessment.

## The SpaMWGDA framework

SpaMWGDA extracts both gene expression and spatial location from ST data, leveraging a deep neural network architecture to identify spatial domains. As depicted in the Fig 1, SpaMWGDA consists of five key modules: (i) spatial neighborhood network construction module; (ii) gene expression enhancement module; (iii) multi-view weighted fusion GCN encoder; (iv) ZINB decoder; and (v) spatial regularization constraint.

In the spatial neighborhood network construction module, KNN and Radius similarity measures are used to create a normalized adjacency matrix, ensuring the model adapts to ST data with varying resolutions and fully utilizes neighborhood information. The gene expression enhancement module enhances gene expression data, yielding improved feature representations. A graph-based contrastive encoder is employed to learn effective representations of gene expression from original and enhanced features. This enhancement allows adaptive integration of key information when reconstructing gene expression. The multi-view weighted fusion GCN encoder extracts embeddings from gene expression data, spatial location data, and their combinations. These embeddings are adaptively fused via a view-level attention mechanism, which helps model minimize noise during clustering. Afterwards, the ZINB decoder reconstructs the feature matrix

to capture the global information of the spatial expression spectrum. Finally, a spatial regularization constraint is incorporated into the learning process of the model to ensure that spatial neighborhood information is preserved and the inherent spatial structure of the data is maintained.

### Spatial neighborhood networks construction

K-nearest neighbor (KNN) can effectively capture the local structure of data, while Radius helps to uncover the global structure of the data [56]. We construct two undirected neighborhood graphs  $G_s^{(n)} = (V, E^{(n)})$ ,  $n = 1, 2$  using KNN and Radius respectively. Here  $V$  represent the set of points, and  $E^{(n)}$  represent the set of edges in  $n$ -th graph. The adjacency matrix is defined as  $A_s^{(n)} \in R^{N \times N}$ ,  $N$  is the number of points.  $A_{ij}^{(1)} = A_{ji}^{(1)} = 1$  if points  $i$  and  $j$  are each other's K-nearest neighbors, and otherwise 0.  $A_{ij}^{(2)} = A_{ji}^{(2)} = 1$  if points  $i$  and  $j$  are within a radius, and otherwise 0. We employ K-D tree algorithm to determine the best  $k$  value and set the radius  $r$  according to the data resolution. The normalized adjacency matrix is calculated as follows:

$$A^{(n)} = \tilde{D}^{(n)-\frac{1}{2}} \tilde{A}_s^{(n)} \tilde{D}^{(n)-\frac{1}{2}} \quad (2)$$

where  $\tilde{A}_s^{(n)} = A_s^{(n)} + I_N$ ,  $\tilde{D}^{(n)} = \sum_j^N \tilde{A}_{ij}^{(n)}$ ,  $n=1,2$ , which  $\tilde{A}_s^{(n)}$  is an adjacency matrix with additional self-connections,  $I_N$  is a unit matrix.

### Gene expression enhancement

- 1) Feature graph construction: SpaMWGDA calculates gene expression similarity using cosine distance.  $G_f = (A_f, X)$  is graph of the gene expression matrix  $X$ ,  $A_f \in R^{N \times N}$  is the feature adjacency matrix,  $A_f^{ij} = 1$  if point  $j$  is the nearest neighbor of point  $i$ , and otherwise 0.
- 2) Data augment: We perform contrastive learning [57] on gene expression to learn latent representations from both original and augmented data. SpaMWGDA takes  $(A_f, X)$  as input and extracts spot features without changing graph structure. For contrastive learning, we generate corrupted neighborhood graph  $(A_f, X')$  on the  $(A_f, X)$ , which scrubs gene expression data while maintaining the original neighborhood graph topology.
- 3) Graph-embedded contrastive encoder: We develop a graph-embedded contrastive encoder to learn key feature representations. The model consists of three modules: graph convolutional encoder, graph deconvolutional decoder, and deep contrastive self-encoder. The process is summarized as follows:
  - Graph convolutional encoder: We utilize a GCN as the encoder to extract relevant feature representations from both original and augmented data.

$$H_0^{(l)} = \phi_0^{(l)}(W_0^{(l)} H_0^{(l-1)} * p_c(L) + b_0^{(l)}) \quad (3)$$

$$H_1^{(l)} = \phi_1^{(l)}(W_1^{(l)} H_1^{(l-1)} * p_c(L) + b_1^{(l)}) \quad (4)$$

The normalized Laplacian matrix is denoted as  $L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ , and  $D$  represents the degree matrix. The activation functions  $\phi_0^{(l)}$  and  $\phi_1^{(l)}$  for the  $l$ -th GCN encoder layer are applied to  $X$  and  $X'$ , respectively. The weight matrices of the encoder for  $X$  and  $X'$  are denoted as  $W_0^{(l)}$  and  $W_1^{(l)}$ , respectively. The bias terms for the VAE corresponding to  $X$  and  $X'$  are represented as  $b_0^{(l)}$  and  $b_1^{(l)}$ . In this model, we use the convolutional operator of GCN as the convolution operation. For convenience, we refer to the gene expression matrices  $X$  and  $X'$  as  $H_0^{(0)}$  and  $H_1^{(0)}$ , respectively.

- Graph deconvolutional decoder: Although graph convolution is adept at capturing local feature, the resulting smoothing can adversely affect the quality of data reconstruction and compromise the learnable global features. We adopt a graph deconvolution network (GDN) as a decoder to alleviate the negative effects of graph convolution and learn feature representation more effectively.

$$H_0^{(k)} = \varphi_0^{(k)}(W_0^{(k)} H_0^{(k-1)} * p_d(L) + b_0^{(k)}) \quad (5)$$

$$H_1^{(l)} = \phi_1^{(k)}(W_1^{(k)} H_1^{(k-1)} * p_d(L) + b_1^{(k)}) \quad (6)$$

where  $\varphi_0^{(k)}$  and  $\phi_1^{(k)}$  are the activation functions of the k-th GDN decoder layer for  $X$  and  $X'$ , respectively.  $W_0^{(k)}$  and  $W_1^{(k)}$  denote the weight matrix of decoder for  $X$  and  $X'$ , respectively.  $b_0^{(k)}$  and  $b_1^{(k)}$  correspond to the bias term of VAE for  $X$  and  $X'$ , respectively. And the deconvolutional operator  $p_d$  is the inverse function of  $p_c$ .

Deep contrastive self-encoder: To effectively balance local and global information, we propose a deep contrastive learning strategy as a constraint for feature learning. This approach enhances SpaMWGDA's ability to represent complex information. The loss of contrastive learning can be expressed as:

$$L_{\text{com}} = \frac{1}{N} \sum_{i=1}^N [\text{sim}(f_{\text{orig}}(A_f, X), f_{\text{aug}}(A_f, X')) - \text{sim}(f_{\text{orig}}(A_f, X), f_{\text{neg}}(A_f, X')) + \alpha] \quad (7)$$

where  $f_{\text{orig}}(A_f, X)$  and  $f_{\text{aug}}(A_f, X')$  represent the feature representation on original graph  $(A_f, X)$  and corrupted graph  $(A_f, X')$  respectively, and  $\alpha$  controls the distance difference between positive and negative samples.

### Multi-view weighted fusion GCN encoder

GCN is a powerful neural network that processes graph data by aggregating information from neighboring nodes, capturing dependencies, and generating embeddings [58]. We utilize a multi-view weighted fusion GCN encoder to extract important information [33]. This encoder consists of four main modules: spatial convolution, feature convolution, co-convolution, and weighted fusion attention mechanism.

- 1) Spatial convolution: A convolution operation is performed on to aggregate spatial information of the neighbors and multi-layer spatial convolutional network applies hierarchical propagation rules:

$$\begin{cases} Z_s^{(l+1)} = \text{ReLU}(\tilde{D}_s^{-\frac{1}{2}} \tilde{A}_s \tilde{D}_s^{-\frac{1}{2}} Z_s^{(l)} W_s^{(l)}) \\ Z_{s\_a}^{(l+1)} = \text{ReLU}(\tilde{D}_{s\_a}^{-\frac{1}{2}} \tilde{A}_{s\_a} \tilde{D}_{s\_a}^{-\frac{1}{2}} Z_{s\_a}^{(l)} W_{s\_a}^{(l)}) \end{cases} \quad (8)$$

where  $W_s^{(l)}$ ,  $W_{s\_a}^{(l)}$  are weight matrices of l-th layer in spatial convolution, and initially  $Z_s^{(0)} = Z_{s\_a}^{(0)} = X$ .  $D_s$  and  $D_{s\_a}$  are diagonal matrices of  $A_s$  and  $A_{s\_a}$  respectively.  $\tilde{A}_s = A_s + I$ ,  $\tilde{A}_{s\_a} = A_{s\_a} + I$ , and the joint embedding is  $Z_s^{(l+1)} = \frac{Z_s^{(l+1)} + Z_{s\_a}^{(l+1)}}{2}$ .

- 2) Feature convolution: To learn more comprehensive gene expression information from latent representations obtained by contrast learning, feature convolution is performed on  $A_f$  and  $X'$ :

$$Z_f^{(l+1)} = \text{ReLU}(\tilde{D}_f^{-\frac{1}{2}} \tilde{A}_f \tilde{D}_f^{-\frac{1}{2}} Z_f^{(l)} W_f^{(l)}) \quad (9)$$

where  $W_f^{(l)}$  is weight matrix of l-th layer, initially  $Z_f^{(0)} = X'$ .

3) Co-convolution: We adopt a parameter sharing strategy to extract co-embedding of gene expression and spatial distribution:

$$\begin{cases} Z_{sc}^{(l+1)} = \text{ReLU}(\tilde{D}_s^{-\frac{1}{2}} \tilde{A}_s \tilde{D}_s^{-\frac{1}{2}} Z_{sc}^{(l)} W_c^{(l)}) \\ Z_{sc\_ac}^{(l+1)} = \text{ReLU}(\tilde{D}_{s\_ac}^{-\frac{1}{2}} \tilde{A}_{s\_ac} \tilde{D}_{s\_ac}^{-\frac{1}{2}} Z_{s\_ac}^{(l)} W_c^{(l)}) \\ Z_{fc}^{(l+1)} = \text{ReLU}(\tilde{D}_f^{-\frac{1}{2}} \tilde{A}_f \tilde{D}_f^{-\frac{1}{2}} Z_{fc}^{(l)} W_c^{(l)}) \end{cases} \quad (10)$$

where  $W_c^{(l)}$  is weight matrix of l-th layer, and initially  $Z_{sc}^{(0)} = Z_{s\_ac}^{(0)} = Z_{fc}^{(0)} = X$ .  $Z_c^{(l+1)} = \frac{Z_{sc}^{(l+1)} + Z_{s\_ac}^{(l+1)} + Z_{fc}^{(l+1)}}{3}$  is joint embedding. The normalization constraint is defined as follows:

$$L_{con} = \left\| \tilde{Z}_{sc} \tilde{Z}_{sc}^T - \tilde{Z}_{s\_ac} \tilde{Z}_{s\_ac}^T \right\|_2^2 + \left\| \tilde{Z}_{sc} \tilde{Z}_{sc}^T - \tilde{Z}_{fc} \tilde{Z}_{fc}^T \right\|_2^2 \quad (11)$$

### Attention mechanism

In order to adaptively learn the importance of each latent embedding, we introduce a weighted attention mechanism. The final attention weight is obtained by calculating weighted sum of attention weights of each view.

$$att\_weight = \text{Softmax}(\text{Linear}(W_2 * \tanh(W_1 * x_i + b_1) + b_2)) \quad (12)$$

where  $x_i$  is the feature representation of each view,  $x_i \in \{Z_s, Z_c, Z_f\}$ ,  $W_1, W_2, b_1, b_2$  are weights and biases.

The weighted fusion module performs weighted fusion on the features of different views according to the weights to obtain an integrated feature representation. The weighted fusion is computed as follows:

$$att_{x_i} = att\_weight_{x_i} * x_i, x_i \in \{Z_s, Z_c, Z_f\} \quad (13)$$

The weighted merged features are accumulated into the final integrated feature representation  $Z = \sum_{i=1}^n att_{x_i}$ .

### ZINB decoder

The ZINB decoder is widely used to reconstruct gene expression matrix to capture the global information [59]. The latent low-dimensional representation  $Z$  is used as the input. Given gene expression data  $X$  of the ST data,  $\pi_{ij}, \mu_{ij}, \theta_{ij}, b_{ij}$  are parameter matrices of zero-inflation parameter, mean, decoder output discretization, and bias vector, respectively. The decoder outputs the estimated values of three parameters.

$$\begin{cases} p_{nb}(x_{ij}|b_i) = \frac{\Gamma(x_{ij}+\theta_{ij})}{\Gamma(x_{ij}+1)\Gamma(\theta_{ij})} \left( \frac{\theta_{ij}}{\theta_{ij}+\mu_{ij}} \right)^{\theta_{ij}} \left( \frac{\mu_{ij}}{\theta_{ij}+\mu_{ij}} \right)^{x_{ij}} \\ p_{zinb}(x_{ij}|b_i) = \text{ZINB}(x_{ij}|\pi_{ij}, \mu_{ij}, \theta_{ij}, b_i) = 0 + (1 - \pi_{ij})p_{nb}(x_{ij}|b_i) \\ L_{zinb} = - \sum_{j=1}^N \sum_{i=1}^N [\log \pi_{ij} p_{zinb}(x_{ij}|b_i)] \end{cases} \quad (14)$$

### Spatial regularization constraint

Spatially neighboring points should be close to each other, while spatially non-neighboring points should be far apart in the latent space [60–62]. Similarity information and spatial neighborhood information are used to compute the loss of spatial regularization constraint:

$$L_{reg} = -\frac{1}{2} \left( \frac{1}{|\varepsilon|} \sum_{(i,j) \in R_i} \log(\sigma(sim_{ij})) + \frac{1}{|N|} \sum_{(i,j) \notin R_i} \log(1 - \sigma(sim_{ij})) \right) \quad (15)$$

where  $sim_{ij}$  represents the cosine similarity between embedding vector  $x_i$  of i-th point and embedding vector  $x_j$  of j-th point in learning-based potential representation H.  $\varepsilon$  represents the set of all neighboring nodes of point i, N represents the set of all non-neighboring nodes of point i, and  $R_i$  represents the set of spatial neighbors of point i.

SpaMWGDA learns more informative and discriminative latent representations by maximizing the similarity of neighboring point pairs and minimizing the similarity of non-neighboring point pairs.

### Evaluation strategies

We compared SpaMWGDA with Scanpy [11], stlearn [25], SpaGCN [26], SEDR [28], STAGATE [29], GraphST [30], and Spatial-MGCN [32] to test the performance of SpaMWGDA. These seven methods include two shallow learning algorithms (Scanpy, stlearn) and five state-of-the-art DL-based methods (SpaGCN, SEDR, STAGATE, GraphST, and Spatial-MGCN). Scanpy (2018) is a Python library that provides data processing, dimensionality reduction, and clustering tools that can be used to identify spatial domain. stlearn (2020) integrated gene expression, spatial data, and morphological features to efficiently identify spatial domains. SpaGCN (2021) combined gene expression and spatial data using GCN to identify spatial domains with consistent gene expression patterns. SEDR (2024) used a deep autoencoder to generate unsupervised spatial embeddings by learning gene representations and embedding spatial information with a variogram autoencoder. STAGATE (2022) fused spatial and gene expression information to identify spatial domains via an adaptive graph attention autoencoder. GraphST (2023) achieved spatial domain identification by integrating graph neural networks with self-supervised contrastive learning. Spatial-MGCN (2023) utilized a multi-view GCN encoder to identify spatial domains.

All benchmark methods were executed using their default parameters, with an equal number of clusters applied during the clustering process. Two widely used metrics (ARI [63] and NMI [64]) were used to evaluate the performance of each method.

### Evaluation metrics

Adjusted Rand Index (ARI) is a measure of clustering similarity that adjusts the Rand Index (RI) for chance. It quantifies how well the predicted clusters align with the ground truth labels while accounting for random cluster assignments. The ARI is calculated as follows:

$$\begin{cases} ARI(W, V) = \frac{RI(W, V) - E[RI(W, V)]}{\max[RI(W, V) - E[RI(W, V)]]} \\ RI(W, V) = \frac{TP \cdot TN}{TP + TN + FP + FN} \end{cases} \quad (16)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Normalized Mutual Information (NMI) is a normalized version of Mutual Information (MI) that corrects for biases introduced by differing numbers of labels. It ensures that values range between 0 and 1, where 1 indicates perfect agreement between predicted clusters and ground truth labels, and 0 indicates no correlation. NMI is calculated as follows:

$$\begin{cases} NMI(W, V) = \frac{MI(W, V)}{\sqrt{H(W)H(V)}} \\ MI(W, V) = \sum_{w \in W} \sum_{v \in V} P(w, v) \log \frac{P(w, v)}{P(w)P(v)} \\ H(W) = - \sum_{w \in W} P(w) \log P(w) \\ H(V) = - \sum_{v \in V} P(v) \log P(v) \end{cases} \quad (17)$$

Let  $W$  represent the set of clustering results, and  $V$  represent the set of ground truth. Then  $P(w)$  and  $P(v)$  denote the probabilities of the ground truth  $w$  and the clustering label  $v$ , while  $P(w, v)$  represents the probabilities of  $w$  and  $v$  occurring simultaneously.  $H(W)$  and  $H(V)$  denote the entropy of the clustering results and the true labels, respectively, reflecting the uncertainty in each label set.

## Supporting information

**S1 Table. Experimental results of SpaMWGDA and seven competing methods on the noisy DLPFC dataset.** (DOCX)

**S1 Fig. Detailed results for all 12 slices in the DLPFC dataset.**

(TIF)

## Author contributions

**Conceptualization:** Lin Yuan, Boyuan Meng, Qingxiang Wang, Chunyu Hu, Cuihong Wang, De-Shuang Huang.

**Data curation:** Lin Yuan, Boyuan Meng, Cuihong Wang, De-Shuang Huang.

**Funding acquisition:** Lin Yuan, De-Shuang Huang.

**Investigation:** Lin Yuan, Boyuan Meng, Cuihong Wang, De-Shuang Huang.

**Methodology:** Lin Yuan, Boyuan Meng, Qingxiang Wang, Chunyu Hu, Cuihong Wang, De-Shuang Huang.

**Project administration:** Lin Yuan, Cuihong Wang, De-Shuang Huang.

**Software:** Boyuan Meng, Qingxiang Wang, Chunyu Hu, Cuihong Wang.

**Supervision:** De-Shuang Huang.

**Visualization:** Lin Yuan, Boyuan Meng, Qingxiang Wang, Chunyu Hu.

**Writing – original draft:** Lin Yuan, Boyuan Meng, Qingxiang Wang, Chunyu Hu, Cuihong Wang, De-Shuang Huang.

**Writing – review & editing:** Lin Yuan, Boyuan Meng, Qingxiang Wang, Chunyu Hu, Cuihong Wang, De-Shuang Huang.

## References

1. Huang Z, Luo S, Zhang Z, Wang Z, Zhou T, Zhang J. A Unified Probabilistic Framework for Modeling and Inferring Spatial Transcriptomic Data. *CBIO*. 2024;19(3):222–34. <https://doi.org/10.2174/1574893618666230529145130>
2. Wang H, Ding Y, Tang J, Guo F. Identification of membrane protein types via multivariate information fusion with Hilbert–Schmidt Independence Criterion. *Neurocomputing*. 2020;383:257–69. <https://doi.org/10.1016/j.neucom.2019.11.103>
3. Cui Y, Wei L, Wang R, Ye X, Sakurai T. Identification of Spatial Domains, Spatially Variable Genes, and Genetic Association Studies of Alzheimer Disease with an Autoencoder-based Fuzzy Clustering Algorithm. *CBIO*. 2024;19(8):765–76. <https://doi.org/10.2174/0115748936278884240102094058>
4. Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. 2019;363(6434):1463–7. <https://doi.org/10.1126/science.aaw1219> PMID: 30923225
5. Gulati GS, D'Silva JP, Liu Y, Wang L, Newman AM. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat Rev Mol Cell Biol*. 2025;26(1):11–31. <https://doi.org/10.1038/s41580-024-00768-2> PMID: 39169166
6. Wei L, He W, Malik A, Su R, Cui L, Manavalan B. Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief Bioinform*. 2021;22(4):bbaa275. <https://doi.org/10.1093/bib/bbaa275> PMID: 33152766
7. Wang Z, Ding H, Zou Q. Identifying cell types to interpret scRNA-seq data: how, why and more possibilities. *Brief Funct Genomics*. 2020;19(4):286–91. <https://doi.org/10.1093/bfgp/ela003> PMID: 32232401
8. Qi R, Wu J, Guo F, Xu L, Zou Q. A spectral clustering with self-weighted multiple kernel learning method for single-cell RNA-seq data. *Brief Bioinform*. 2021;22(4):bbaa216. <https://doi.org/10.1093/bib/bbaa216> PMID: 33003206
9. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre EJ. Fast unfolding of communities in large networks. *J Stat Mech*. 2008;2008(10):P10008.
10. Traag VA, Waltman L, Van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Scientometrics*. 2019;9(1):1–12.

11. 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
12. Satija R, Farrell JA, Gennert D, Schier AF, Regev AJ. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015;33(5):495–502.
13. Dai C, Jiang Y, Yin C, Su R, Zeng X, Zou Q, et al. scIMC: a platform for benchmarking comparison and visualization analysis of scRNA-seq data imputation methods. *Nucleic Acids Res.* 2022;50(9):4877–99. <https://doi.org/10.1093/nar/gkac317> PMID: 35524568
14. Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. 2021;22:1–31.
15. 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> PMID: 34083791
16. Xu J, Lu C, Jin S, Meng Y, Fu X, Zeng X, et al. Deep learning-based cell-specific gene regulatory networks inferred from single-cell multiome data. *Nucleic Acids Res.* 2025;53(5):gkaf138. <https://doi.org/10.1093/nar/gkaf138> PMID: 40037709
17. Li HL, Pang YH, Liu BJJN. BioSeq-BLM: a platform for analyzing DNA, RNA and protein sequences based on biological language models. *Nucleic Acids Research.* 2021;49(22):e129–e.
18. Liu J, Ma L, Ju F, Zhao C, Yu L. SpaCcLink: exploring downstream signaling regulations with graph attention network for systematic inference of spatial cell–cell communication. *J Biol B.* 2025;23(1):44.
19. Huang Z, Guo X, Qin J, Gao L, Ju F, Zhao C, et al. Accurate RNA velocity estimation based on multibatch network reveals complex lineage in batch scRNA-seq data. *BMC Biol.* 2024;22(1):290. <https://doi.org/10.1186/s12915-024-02085-8> PMID: 39696422
20. Zhao M, Li J, Liu X, Ma K, Tang J, Guo F. A gene regulatory network-aware graph learning method for cell identity annotation in single-cell RNA-seq data. *Genome Res.* 2024;34(7):1036–51. <https://doi.org/10.1101/gr.278439.123> PMID: 39134412
21. Zhang H-Q, Arif M, Thafar MA, Albaradei S, Cai P, Zhang Y, et al. PMPred-AE: a computational model for the detection and interpretation of pathological myopia based on artificial intelligence. *Front Med (Lausanne).* 2025;12:1529335. <https://doi.org/10.3389/fmed.2025.1529335> PMID: 40182849
22. Wang J, Zou Q, Lin C. A comparison of deep learning-based pre-processing and clustering approaches for single-cell RNA sequencing data. *Brief Bioinform.* 2022;23(1):bbab345. <https://doi.org/10.1093/bib/bbab345> PMID: 34472590
23. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-Pro: support bio-sequence machine for proteins. *Sci China Inf Sci.* 2024;67(11). <https://doi.org/10.1007/s11432-024-4171-9>
24. Liu T, Fang ZY, Li X, Zhang LN, Cao DS, Yin MZ. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *J Biol Inorg Chem.* 2023;24(3):bbad146.
25. Pham D, Tan X, Xu J, Grice LF, Lam PY, Raghubar A. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. 2020;2020:31.125658.
26. 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
27. 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
28. 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> PMID: 38217035
29. 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
30. Long Y, Ang KS, Li M, Chong KLK, 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> PMID: 36859400
31. Pang C, Qiao J, Zeng X, Zou Q, Wei L. Deep Generative Models in De Novo Drug Molecule Generation. *J Chem Inf Model.* 2024;64(7):2174–94. <https://doi.org/10.1021/acs.jcim.3c01496> PMID: 37934070
32. 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
33. Shi X, Zhu J, Long Y, Liang C. Identifying spatial domains of spatially resolved transcriptomics via multi-view graph convolutional networks. *Brief Bioinform.* 2023;24(5):bbad278. <https://doi.org/10.1093/bib/bbad278> PMID: 37544658
34. Guo X, Huang Z, Ju F, Zhao C, Yu L. Highly Accurate Estimation of Cell Type Abundance in Bulk Tissues Based on Single-Cell Reference and Domain Adaptive Matching. *Adv Sci (Weinh).* 2024;11(7):e2306329. <https://doi.org/10.1002/advs.202306329> PMID: 38072669
35. Qi R, Ma A, Ma Q, Zou Q. Clustering and classification methods for single-cell RNA-sequencing data. *Brief Bioinform.* 2020;21(4):1196–208. <https://doi.org/10.1093/bib/bbz062> PMID: 31271412
36. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inform Theory.* 1967;13(1):21–7. <https://doi.org/10.1109/tit.1967.1053964>
37. Bentley JL, Stanat DF, Williams EH Jr. The complexity of finding fixed-radius near neighbors. *Information Processing Letters.* 1977;6(6):209–12. [https://doi.org/10.1016/0020-0190\(77\)90070-9](https://doi.org/10.1016/0020-0190(77)90070-9)

38. Zeng Y, Yin R, Luo M, Chen J, Pan Z, Lu Y, et al. Identifying spatial domain by adapting transcriptomics with histology through contrastive learning. *Brief Bioinform.* 2023;24(2):bbad048. <https://doi.org/10.1093/bib/bbad048> PMID: 36781228
39. Qian Y, Zou Q, Zhao M, Liu Y, Guo F, Ding Y. scRNMF: An imputation method for single-cell RNA-seq data by robust and non-negative matrix factorization. *PLoS Comput Biol.* 2024;20(8):e1012339. <https://doi.org/10.1371/journal.pcbi.1012339> PMID: 39116191
40. Yu Z, Lu Y, Wang Y, Tang F, Wong K-C, Li X. ZINB-Based Graph Embedding Autoencoder for Single-Cell RNA-Seq Interpretations. *AAAI.* 2022;36(4):4671–9. <https://doi.org/10.1609/aaai.v36i4.20392>
41. Yuan H, Liu M, Qiu Y, Ching W-K, Zou Q. PLNMFG: Pseudo-label guided non-negative matrix factorization model with graph constraint for single-cell multi-omics data clustering. *PLoS Comput Biol.* 2025;21(8):e1013375. <https://doi.org/10.1371/journal.pcbi.1013375> PMID: 40825083
42. 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. *Frontiers in Genetics.* 2023;14:1202409.
43. Si Z, Li H, Shang W, Zhao Y, Kong L, Long C, et al. SpaNCMG: improving spatial domains identification of spatial transcriptomics using neighborhood-complementary mixed-view graph convolutional network. *Brief Bioinform.* 2024;25(4):bbae259. <https://doi.org/10.1093/bib/bbae259> PMID: 38811360
44. Maynard KR, Collado-Torres L, Weber LM, Uytengco C, Barry BK, Williams SR, 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> PMID: 33558695
45. Jiang J, Wang N, Chen P, Zheng C, Wang B. Prediction of Protein Hotspots from Whole Protein Sequences by a Random Projection Ensemble System. *Int J Mol Sci.* 2017;18(7):1543. <https://doi.org/10.3390/ijms18071543> PMID: 28718782
46. Kadokawa K, Sugimoto K, Yamaguchi F, Song T, Watanabe Y, Singh K, et al. Phosphohippolin expression in the rat central nervous system. *Brain Res Mol Brain Res.* 2004;125(1–2):105–12. <https://doi.org/10.1016/j.molbrainres.2004.03.021> PMID: 15193427
47. Liu X, Liu S. Cholecystokinin selectively activates short axon cells to enhance inhibition of olfactory bulb output neurons. *J Physiol.* 2018;596(11):2185–207. <https://doi.org/10.1113/JP275511> PMID: 29572837
48. Sun X, Liu X, Starr ER, Liu S. CCKergic Tufted Cells Differentially Drive Two Anatomically Segregated Inhibitory Circuits in the Mouse Olfactory Bulb. *J Neurosci.* 2020;40(32):6189–206. <https://doi.org/10.1523/JNEUROSCI.0769-20.2020> PMID: 32605937
49. Wang Y, Sheng N, Xie Y, Chen S, Lu J, Zhang Z, et al. Low expression of CRISP3 predicts a favorable prognosis in patients with mammary carcinoma. *J Cell Physiol.* 2019;234(8):13629–38. <https://doi.org/10.1002/jcp.28043> PMID: 30609035
50. Liang X, Liu P, Xue L, Chen B, Liu W, Shi W, et al. A multi-modality and multi-granularity collaborative learning framework for identifying spatial domains and spatially variable genes. *Bioinformatics.* 2024;40(10):btae607. <https://doi.org/10.1093/bioinformatics/btae607> PMID: 39418177
51. Long Y, Ang KS, Sethi R, Liao S, Heng Y, van Olst L, et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat Methods.* 2024;21(9):1658–67. <https://doi.org/10.1038/s41592-024-02316-4> PMID: 38907114
52. Buache E, Etique N, Alpy F, Stoll I, Muckensturm M, Reina-San-Martin B, et al. Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. *Oncogene.* 2011;30(29):3261–73. <https://doi.org/10.1038/onc.2011.41> PMID: 21358676
53. 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
54. Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA. Human pancreas development. *Development.* 2015;142(18):3126–37. <https://doi.org/10.1242/dev.120063> PMID: 26395141
55. Eikrem OS, Strauss P, Beisland C, Scherer A, Landolt L, Flatberg A, et al. Development and confirmation of potential gene classifiers of human clear cell renal cell carcinoma using next-generation RNA sequencing. *Scand J Urol.* 2016;50(6):452–62. <https://doi.org/10.1080/21681805.2016.1238007> PMID: 27739342
56. Cui L, Guo G, Ng MK, Zou Q, Qiu Y. GSTRPCA: irregular tensor singular value decomposition for single-cell multi-omics data clustering. *Brief Bioinform.* 2024;26(1):bbae649. <https://doi.org/10.1093/bib/bbae649> PMID: 39680741
57. Zhu Y, Xu Y, Yu F, Liu Q, Wu S, Wang L. Graph Contrastive Learning with Adaptive Augmentation. In: *Proceedings of the Web Conference 2021*, 2021. 2069–80. <https://doi.org/10.1145/3442381.3449802>
58. Kipf TN, Welling MJ. Semi-supervised classification with graph convolutional networks. 2016.
59. Gan Y, Huang X, Zou G, Zhou S, Guan J. Deep structural clustering for single-cell RNA-seq data jointly through autoencoder and graph neural network. *Brief Bioinform.* 2022;23(2):bbac018. <https://doi.org/10.1093/bib/bbac018> PMID: 35172334
60. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *Proceedings of Machine Learning Research*, 2020.
61. Stoltzfus CR, Filipek J, Gern BH, Olin BE, Leal JM, Wu Y. CytoMAP: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. 2020;31(3).
62. Yuan L, Xu Z, Meng B, Ye L. scAMZI: attention-based deep autoencoder with zero-inflated layer for clustering scRNA-seq data. *BMC Genomics.* 2025;26(1):350. <https://doi.org/10.1186/s12864-025-11511-2> PMID: 40197174
63. Hu Y, Xiao K, Yang H, Liu X, Zhang C, Shi Q. Spatially contrastive variational autoencoder for deciphering tissue heterogeneity from spatially resolved transcriptomics. *J Biol Informa Bio.* 2024;25(2):bbae016.
64. Estévez PA, Tesmer M, Perez CA, Zurada JMJ. Normalized mutual information feature selection. *JITonn.* 2009;20(2):189–201.