OXFORD

Briefings in Bioinformatics, 2025, **26(1)**, bbae608
https://doi.org/10.1093/bib/bbae608
Problem Solving Protocol

MCGAE: unraveling tumor invasion through integrated multimodal spatial transcriptomics

Yiwen Yang[1,‡], Chengming Zhang[2,‡], Zhaonan Liu[3], Kazuyuki Aihara[2], Chuanchao Zhang[4,5,*], Luonan Chen[4,5,6,*], Wu Wei[1,*]

[1]Lingang Laboratory, Building 8, 319 Yueyang Road, Xuhui District, Shanghai 200031, China
[2]International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[3]Department of Biliary-Pancreatic Surgery, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 160 Pujian Road, Pudong District, Shanghai 200127, China
[4]Key Laboratory of Systems Health Science of Zhejiang Province, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 1 Sub-Lane Xiangshan Road, West Lake District, Hangzhou 310024, China
[5]Guangdong Institute of Intelligence Science and Technology, Hengqin, Zhuhai, Guangdong 519031, China
[6]Key Laboratory of Systems Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, 320 Yueyang Road, Xuhui District, Shanghai 200031, China

*Corresponding authors. Wu Wei, E-mail: wuwei@lglab.ac.cn; Luonan Chen, E-mail: lnchen@sibcb.ac.cn; Chuanchao Zhang, E-mail: chuanchaozhang@ucas.ac.cn.
‡Yiwen Yang and Chengming Zhang contributed equally to this work as co-first authors.

Abstract

Spatially Resolved Transcriptomics (SRT) serves as a cornerstone in biomedical research, revealing the heterogeneity of tissue microenvironments. Integrating multimodal data including gene expression, spatial coordinates, and morphological information poses significant challenges for accurate spatial domain identification. Herein, we present the Multi-view Contrastive Graph Autoencoder (MCGAE), a cutting-edge deep computational framework specifically designed for the intricate analysis of spatial transcriptomics (ST) data. MCGAE advances the field by creating multi-view representations from gene expression and spatial adjacency matrices. Utilizing modular modeling, contrastive graph convolutional networks, and attention mechanisms, it generates modality-specific spatial representations and integrates them into a unified embedding. This integration process is further enriched by the inclusion of morphological image features, markedly enhancing the framework's capability to process multimodal data. Applied to both simulated and real SRT datasets, MCGAE demonstrates superior performance in spatial domain detection, data denoising, trajectory inference, and 3D feature extraction, outperforming existing methods. Specifically, in colorectal cancer liver metastases, MCGAE integrates histological and gene expression data to identify tumor invasion regions and characterize cellular molecular regulation. This breakthrough extends ST analysis and offers new tools for cancer and complex disease research.

Keywords: spatially transcriptomics; multi-view integration; multimodal integration; tumor microenvironment analysis

Introduction

In the study of multicellular organism tissues, a nuanced comprehension of biological processes and molecular dynamics within spatial context, necessitates the employment of advanced biotechnologies characterized by high spatial resolution [1, 2]. Recent advancements in spatial transcriptomics (ST), including in situ hybridization (ISH) techniques such as osmFISH [3], MERFISH [4, 5], and seqFISH [6], alongside in situ sequencing technologies like FISSEQ [7] and STARmap [8], have facilitated the precise identification of cells with high sensitivity and resolution. Barcode-based ST, as exemplified by 10× Visium [9], Slide-seq [10], and Stereo-seq [11], capture cDNA sequences in situ and perform high-throughput sequencing post-elution, thus determining all expressed genes in a spatially resolved manner. For certain tumor ST data, image information provides essential modality information for more accurately delineating tumor domains [12]. Hence, integrating multimodal information is essential for a holistic cellular landscape, given the diversity of information derived from ST.

In ST, identifying spatial domains is crucial, involving clustering to assign structural labels to spots or cells. Traditional non-spatial algorithms like Seurat [13] use linear dimensionality reduction for clustering based on expression data alone. Spatial-aware algorithms introduce advanced methods: Giotto [14] uses a Hidden Markov Random Field model with spatial priors, and BayesSpace [15] applies Bayesian methods to optimize relationships between adjacent points, improving spatial structure detection. STAGATE [16] incorporate graph-based and attention mechanisms, respectively, to enhance spatial domain identification. SGCAST [17] adopts a symmetric graph convolutional autoencoder to learn latent embeddings via integrating the gene expression similarity and the spatial proximity of the spots. SpaceFlow [18], GraphST [19], and conST [20] combine graph neural networks (GNNs)/variational graph autoencoders and contrastive learning framework for clustering. STMGCN [21] employ GNNs with contrastive learning and multiple adjacency matrices to detect spatial domains effectively. SpaGCN [22] integrates image data into a neighborhood graph using graph

Received: April 30, 2024. **Revised:** October 16, 2024. **Accepted:** November 7, 2024

convolutional networks (GCNs). stLearn [23] introduces a within-tissue normalization technique that normalizes gene expression using morphological distance, derived from characteristics collected from morphology images [e.g. by hematoxylin and eosin (H&E) staining] and spatial locations. DeepST [24], similar to stLearn, enhances gene expression by leveraging the similarity between morphological features, spatial location, and gene expression, and then employs a GNN autoencoder and a denoising autoencoder to generate a latent representation. TIST [25] constructs modality-specific networks from histopathology, gene expression, and spatial data, then subsequently fuses them into a multimodality similarity network (TIST-net) for cell cluster identification and gene expression enhancement.

In general, non-spatial clustering algorithms often overlook vital spatial relationships. In contrast, GNN-based methods enhance data analysis by converting spatial positions into adjacency matrices, thereby improving algorithmic adaptability to diverse datasets across various sequencing platforms. However, these methods tend to rely on predefined specific similarity metrics, creating challenges in selecting the most effective metric due to the resultant variation in graph structures [21]. Additionally, most of these methods do not efficiently utilize spatial image information, resulting in a lack of a comprehensive description of the cellular landscape from multiple viewpoints. [26]. This oversight is particularly critical in medical imaging, where the complexity of histological features, such as lesions and microvessels, offers invaluable insights that can significantly enhance the spatial domain identification process.

To address these challenges, we propose a multi-view contrastive graph autoencoder (MCGAE) framework, named MCGAE, for spatial domain identification. MCGAE constructs varied enhanced gene expression representations and multiple topological graphs from spatial coordinates. Simultaneously, MCGAE segments images based on such spatial coordinates and utilizes deep convolutional networks to extract image features. Notably, MCGAE is capable of functioning effectively even in the absence of image information. By integrating gene expression, spatial positions, and image information through a self-attention mechanism, MCGAE generates a multimodal, multi-view biological representation that is robust to noise and variability. Furthermore, for continuous slices, MCGAE enhances the similarity between adjacent slices by calculating the adjacency distance of neighboring batches, ultimately reconstructing 3D spatial domains.

MCGAE has been rigorously tested on diverse datasets, including 10× Visium, Stereo-seq [11], and Slide-seq datasets, as well as simulated datasets. Consistently, MCGAE has demonstrated superior performances over nine existing algorithms, exhibiting high precision and robustness. Notably, when integrated with tumor imaging data, MCGAE has shown enhanced accuracy in identifying spatial domains and achieving precise clustering. Through its capabilities in data denoising, identification of spatially variable genes (SVGs), and extraction of 3D spatial domain features, MCGAE facilitates in-depth exploration of tumor invasion areas, thereby advancing research and treatment in medical oncology. This enhanced capability holds significant promise for clinical applications, bolstering confidence in decision-making and contributing to improved clinical outcomes.

Material and methods

Overview of multi-view contrastive graph autoencoder approach

MCGAE is a novel deep computational framework designed for comprehensive analysis of ST data across platforms, integrating morphological images. Given spatial multi-modal transcriptomics data, MCGAE commences by initially acquiring the original gene expression matrix alongside the adjacency matrix, which is computed based on spatial coordinates. The framework's multi-view construction is facilitated through a modular modeling approach, granting users the flexibility to select from a variety of enhancement methods, including but not limited to simple autoencoders, for the purpose of obtaining enhanced views of gene expression matrix X. Furthermore, the construction of multiple views of adjacency matric A is achievable by employing diverse similarity metrics, thereby enriching the data analysis spectrum by capturing a multitude of perspectives and relationships inherent within the ST data (Fig. 1A).

In the ensuing phase, modality-specific spots representations are obtained through contrastive graph convolutional neural networks coupled with attention modules. Specifically, MCGAE keeps X fixed and uses it with different views of A, where each pair (X_1, A_j) is processed through a GCN to extract multi-view representations that are specifically pertinent to X_1. These representations are then fused into a comprehensive embedding Z^X using an attention mechanism, which is utilized for the reconstruction of the original X_1. Similarly, by keeping A_1 constant and varying X_i, MCGAE follows the same process to garner view-specific representations for A_1, which are aggregated into Z^A. The pair (X_1, A_1) is designated as the base graph, encapsulating the original expression data. The refinement of their biological representations is further achieved through the adoption of self-supervised contrastive learning. In instances where morphological images are accessible, MCGAE leverages a pre-trained ResNet50 [27] for the extraction of image features, resulting in the image embedding Z^{morph}, thereby enhancing the model's capability in processing multimodal data [24, 28] (Fig. 1B).

During the terminal fusion phase, MCGAE employs an attention mechanism to combine Z^X, Z^A, and Z^{morph}, creating the ultimate composite embedding Z (Fig. 1C). This embedding is further refined through an unsupervised deep iterative clustering strategy to enhance its compactness, which is then applied to downstream analytical tasks such as spatial domain identification, data denoising, SVGs identification, trajectory inference and extraction of 3D spatial domain (Fig. 1D). By integrating multi-view contrastive GNNs, attention mechanisms, and deep iterative clustering, MCGAE achieves precise and customized embeddings, significantly enhancing the reconstruction of spatial structures and the representation of gene expression patterns. This approach adeptly handles the complexities of ST, providing essential insights into tissue heterogeneity and proving to be of immense value in advanced biomedical research.

Multi-view graph construction

Spatial transcriptomics revolutionizes our understanding of tissue architecture by leveraging spatial information to correlate cellular states with their physical locations, thereby elucidating tissue substructures. The crux of ST lies in its capacity to employ spatial information for identifying similar cellular states situated in proximate locations, facilitating a nuanced delineation of tissue substructures. To fully leverage spatial information, the input gene expression matrix X and the similarity matrix A between spots are represented as an undirected graph $G = (X, A)$, where $X \in \mathbb{R}^{n \times p}$ represents the normalized gene expression matrix, $A \in \mathbb{R}^{n \times n}$ is the spatial adjacency matrix for n spots, and p is the number of filtered genes.

To fully exploit spatial information and enhance our understanding of cellular organization, we constructed a multi-view

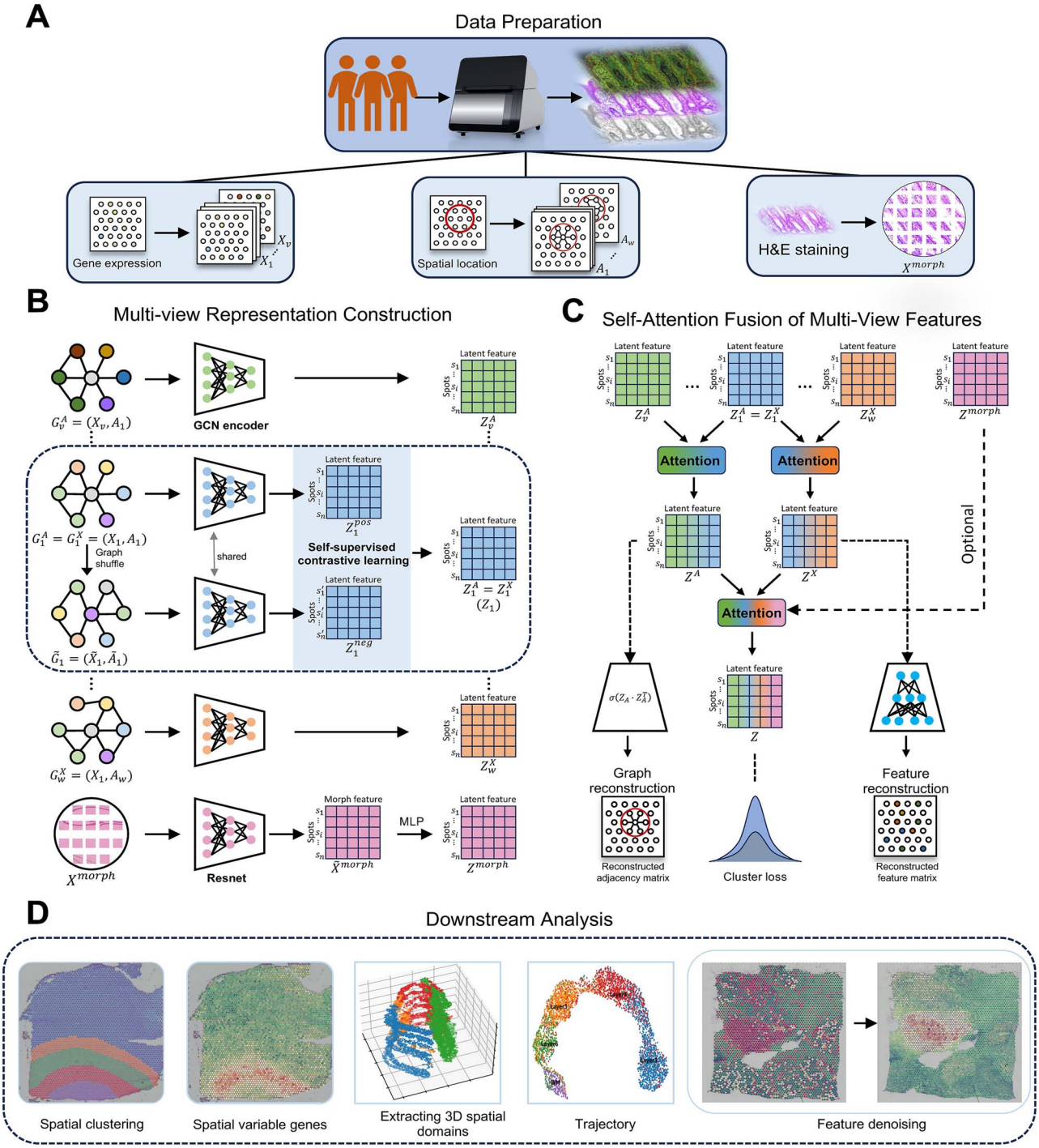

Figure 1. Overview of MCGAE. (A) Spatial transcriptomics techniques generate diverse data modalities, including gene expression profiles, cellular positioning, and morphological images. We process these modalities to create multiple views, denoted X_1,X_2,...,X_v and a_1,a_2,...,a_w. In the case of morphological images, we segment them based on cell coordinates, often H&E stained, to extract relevant features. (B) For each view, we use GCNs to integrate gene expression and spatial coordinates to learn spot representations. Simultaneously, we employ self-supervised contrastive learning on the base graph (G_1^a = G_1^X) to capture genuine biological signals (Z_1). For morphological images, we extract their representations using a ResNet50 network and map them into a spot-like representation space with an MLP. (C) We fuse representations of multiple views invariant to X into Z^X using selfattention, and representations fixed to a into Z^A. These are fused with the Z^morph view of the image into the final representation Z. We optimize the model with unsupervised iterative clustering loss. Additionally, we reconstruct A and X based on the fused representations (Z^A,Z^X). (D) The final representation Z generated by MCGAE, along with the reconstructed expression data, can be applied to various downstream analysis tasks, including spatial domain analysis, trajectory inference, identification of SVGs, data denoising, and the elucidation 3D spatial domains.

graph. This process entailed the generation of multi-view representations for both X and A, where X represents an augmented version of the original expression matrix, and A denotes the adjacency matrix between spots, inferred from a variety of metrics. This multi-view graph construction enables a comprehensive analysis of spatial context and cellular states. (See Supplementary Note 1 for detailed multi-view building methods).

Contrastive graph convolutional neural networks

GCNs excel in utilizing graph structures, enabling the processing of the constructed graph $G = (X, A)$ by integrating graph topology with node features. Framed within a multi-view learning approach, our model performs graph convolutions across each graph to derive view-specific embeddings seamlessly.

Graph convolutional neural networks

For simplicity, we omit the superscripts denoting views, representing the input data as $G = (X, A)$. We employ a GCN [29] as the encoder, iteratively aggregating representations of neighbors to learn the latent representation z_i for each spot i. The representation at the l-th layer in the encoder can be expressed as:

$$H^{(l+1)} = \mathbf{ReLU}\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right), \tag{1}$$

where $W^{(l)}$ is the weight matrix for the l-th layer of spatial convolution, with the initial $H^{(0)} = X$, and ReLU represents the activation function. Specifically, we define $\widetilde{A} = A + I$, and $\widetilde{D}$ is the diagonal degree matrix of $\widetilde{A}$. $H^{(l)}$ denotes the output representation of the l-th layer, with $H^{(0)}$ being the original input gene expression X. The final output of the encoder, denoted as Z, where each row z_i represents the latent embedding of spot i.

Self-supervised contrastive learning

To enhance latent spot representations, we employ Deep Graph Infomax [30], an unsupervised contrastive learning method that maximizes mutual information between node representations and the graph's global representation. We aggregate local neighborhood representations into g_i for each spot i, capturing its microenvironment. Following the strategy of GraphST, our readout function uses the sigmoid of neighbor representations' average, focusing on local context similarity. In our model, a spot's representation z_i and its local information vector g_i form positive pairs, while representations from perturbed graphs form negative pairs, aiming to maximize mutual information for positive pairs and minimize it for negative pairs. We apply binary cross-entropy for loss calculation:

$$\mathcal{L}_{\mathrm{scl}} = \frac{1}{2n}\left(\sum_{i=1}^{n}\mathbb{E}_{(X,A)}\left[\log\sigma\left(h_i^T M g_i\right)\right] + \sum_{i=1}^{n}\mathbb{E}_{\left(\tilde{X},\tilde{A}\right)}\left[\log\left(1-\sigma\left(\tilde{h}_i^T M g_i\right)\right)\right]\right). \tag{2}$$

Here, h_i and $\tilde{h}_i$ are the authentic and perturbed graph representations, respectively, g_i is the local microenvironment summary, and M is a trainable matrix. Perturbed data is generated by randomizing gene expression profiles while keeping the network topology constant.

Histological image embedding extraction

For ST data with morphological information, we segment an image (H&E staining tiles) based on the coordinates of each spot to obtain its partial image. We then apply torchvision.transforms [31] for various transformations and augmentations on these partial images, including normalization, rotation, sharpness adjustment, and other operations. Subsequently, we extract the high-level features of each spot tile using a pretrained convolutional neural network model, specifically ResNet50, capable of transforming each spot image into 2048-dimensional latent variables. Furthermore, to enhance the representation of spot morphology, we conduct principal component analysis to extract the first 50 principal components as latent characteristics. Finally, we map these features into a spot-like representation space Z^{morph} using a multilayer perceptron (MLP).

Embedding fusion with attention mechanism

By applying graph convolution operations on various graphs, we obtained multiple view-specific embeddings for spots. To adaptively fuse these embeddings into a unified representation based on their importance, we employed an attention mechanism to optimize the embedding fusion process.

Given a series of embeddings $z_1, \ldots, z_m$ for each spot, our goal is to use the attention mechanism to calculate the weight coefficients $\boldsymbol{\alpha}$ for each embedding and fuse them accordingly. The attention mechanism is defined as follows:

$$\boldsymbol{\alpha} = att\left(z_1, \ldots, z_m\right), \tag{3}$$

where $\boldsymbol{\alpha}$ is the vector of self-attention coefficients, with $\boldsymbol{\alpha}_i$ representing the self-attention coefficient of the target spot in the i-th embedding. Specifically, we first perform a linear transformation on each spot embedding to obtain attention values, i.e., $v_i = q^T.z_i$, where q is a shared attention vector and z_i represents the embedding of the i-th view. Then, we normalize these attention values using the softmax function to ensure comparability of coefficients across different spots. The attention coefficient for a given spot in the i-th embedding is obtained as follows:

$$\alpha_i = \mathbf{softmax}\left(v_i\right) = \frac{\exp\left(v_i\right)}{\sum_{i=1}^{m}\exp\left(v_i\right)}. \tag{4}$$

Ultimately, through these attention coefficients, we can fuse all embeddings into a comprehensive embedding z for downstream analysis:

$$z = \sum\nolimits_{i=1}^{m}\boldsymbol{\alpha}_i.z_i. \tag{5}$$

In our framework, we first conduct a preliminary fusion of embeddings from different graphs, including those generated by varying A while keeping X constant $\left(Z_1^X, \ldots, Z_v^X\right)$ and those generated by varying X while keeping A constant $\left(Z_1^A, \ldots, Z_w^A\right)$, before finally fusing them with Z^{morph}. Specifically, we use an attention mechanism to fuse $Z_1^X, \ldots, Z_v^X$ into Z^X, which is then used for the reconstruction of X: the latent representation Z^X is input into a decoder to restore it to the original gene expression dimension. Unlike the encoder, the decoder employs an MLP for gene expression reconstruction, denoted as $\widetilde{X}$:

$$\mathcal{L}_{\mathrm{rec}}^X = \left\|\widetilde{X} - X\right\|_2^2, \tag{6}$$

where $\tilde{X}$ denotes the reconstructed gene expression, the output of the decoder. Similarly, $Z_1^A, \ldots, Z_w^A$ are fused into Z^A for the reconstruction of the adjacency matrix A.

$$\mathcal{L}_{\text{rec}}^{\text{A}} = \left\| Z^{\text{A}} \left(Z^{\text{A}}\right)^{\text{T}} - \text{A} \right\|_2^2, \tag{7}$$

$$\mathcal{L}_{\text{rec}} = \mathcal{L}_{\text{recon}}^{\text{X}} + \mathcal{L}_{\text{recon}}^{\text{A}}. \tag{8}$$

Here X and A refer to X_1 and A_1 of the base graph, respectively. Subsequently, we utilize the attention mechanism to adaptively fuse Z^X, Z^A, and Z^{morph} into the final embedding Z for downstream analysis.

Self-optimizing deep embedded clustering

To effectively identify spatial domains, we employ an unsupervised deep embedding clustering framework to iteratively assign spots to their respective domains (See Supplementary Note 2 for the detailed method of calculating q_{ij} and p_{ij}). After obtaining the soft assignment distribution q_{ij} and the auxiliary target distribution p_{ij}, we optimize the model based on the Kullback–Leibler (KL) divergence loss between the two distributions:

$$\mathcal{L}_{\text{cluster}} = \text{KL}\left(P \| Q\right) = \sum_i \sum_j p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) \tag{9}$$

Post-training, the class assignment for spot i can be determined by $\arg\max_j q_{ij}$. By identifying the index j corresponding to the maximum value in the q_i vector, we assign spot i to class j.

Optimization objectives and training of multi-view contrastive graph autoencoder

During training, the contrastive loss $\mathcal{L}_{\text{scl}}$, reconstruction loss $\mathcal{L}_{\text{rec}}$ and clustering loss $\mathcal{L}_{\text{cluster}}$ are jointly optimized. The final training objective of MCGAE is defined as:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{scl}} + \beta \mathcal{L}_{\text{rec}} + \gamma \mathcal{L}_{\text{cluster}}, \tag{10}$$

where α, β, and γ are weight factors used to balance the impact of different losses. Through ablation study, we demonstrated the necessity of each component in our model (Supplementary Note 3). Empirically, we set $\alpha = 0.1$, $\beta = 0.1$, and $\gamma = 1$.

The model was trained with the Adam optimizer [32]. The MCGAE architecture was implemented using PyTorch [33] (v.1.13.0) and run using a GPU running CUDA (v.11.6).

Results

Multi-view contrastive graph autoencoder improved spatial domain detection on human dorsolateral prefrontal cortex data with 10× Visium, enhancing known layer identification

We initially assessed the spatial clustering performance of MCGAE using the LIBD human dorsolateral prefrontal cortex (DLPFC) dataset [34]. This dataset included spatially resolved transcriptomic profiles from 12 DLPFC slices, each depicting four to six layers of the cortex and white matter (WM). This evaluation focused on the unsupervised capacity of MCGAE and competing methods to recover annotated anatomical layers (Supplementary Note 4). MCGAE demonstrated superior performance across all slices, achieving a median score of 0.51 and an average Adjusted Rand Index (ARI) of 0.52, surpassing all other methods (Supplementary Note 5). Notably, STAGATE also exhibited commendable performance, with a median score of 0.44 and an average ARI of 0.41 (Fig. 2A). We also evaluated the performance of MCGAE in the mouse forebrain, STARmap [8] (Supplementary Fig. S1), image-based technology (MERFISH [5]) (Supplementary Fig. S2) and two sets of simulated datasets (Supplementary Note 6 and Figs S3 and S4), achieving optimal results (Fig. 2G–I).

Subsequently, we used one specific slice (#151669) to exemplify our results (Fig. 2B). Visual inspection revealed that most algorithms faced challenges in differentiating between layer 3 and layer 4, as well as in demarcating the boundary between layer 4 and layer 5. MCGAE, however, distinguished itself by accurately stratifying these layers and clearly identifying the boundary between them. A quantitative evaluation utilizing the ARI underscored MCGAE's superior performance, with a leading score of 0.63, followed by the BayesSpace algorithm at 0.46 (Fig. 2C and Supplementary Fig. S5).

The low-dimensional embeddings and denoised expression data generated by MCGAE were instrumental for subsequent analysis. For trajectory inference (TI), we utilized the final representationz obtained from our model and performed TI using Scanpy's 'sc.tl.paga' function. Although we did not explicitly constrain z for TI, the refinedz effectively preserved biological structures, leading to accurate visualization for TI. For the identification of SVGs, we employed a strategy similar to that used in SpaGCN, integrated within the MCGAE framework, which leveraged both the clusters derived from z and the original or denoised expression data (Fig. 2D and E and Supplementary Figs S4B and S5). Employing uniform manifold approximation and projection (UMAP) visualization of these embeddings, as an example for slice #151669, highlighted the distinct manifold structures. Unlike most methods, which struggled to differentiate between WM and layer 3, MCGAE successfully depicted a clear pseudo-trajectory from WM to layer 3 (Fig. 2D and Supplementary Fig. S5).

In an in-depth exploration of the identified spatial domains, our methodology paralleled that of SpaGCN [22] for SVGs detection and comparison. To better quantify the differences between the two methods, we employed Moran's I, a measure of spatial autocorrelation that quantifies the degree to which a variable is similar to itself in nearby locations. This analysis indicated that the median Moran's I values [35] of SVGs derived from the original gene expression data by MCGAE closely aligned with those identified by SpaGCN, with values of 0.13 and 0.12, respectively (Fig. 2E, Supplementary Fig. S6 and Supplementary Table 1). Remarkably, SVGs from denoised data demonstrated significantly higher spatial correlation, with a median Moran's I value of 0.76. Furthermore, spatial domain patterns identified from the denoised data were markedly more distinct and concentrated compared to those derived from the original expression data (Fig. 2F).

To demonstrate the necessity of each component in our model, we conducted ablation studies (Supplementary Note 3 and Fig. S6A and B), which confirmed the benefits of incorporating multi-view learning strategies, attention mechanisms, and the spatial clustering module in MCGAE. Additionally, we developed a fine-tuned version of MCGAE for image feature extraction to better capture image information (Supplementary Note 7). We also compared the computational cost and memory usage of different methods. By sampling datasets, we evaluated runtime and RAM usage across varying numbers of spots, finding that MCGAE's performance is moderate in both metrics. For example, processing 20 000 spots took under six minutes with a RAM

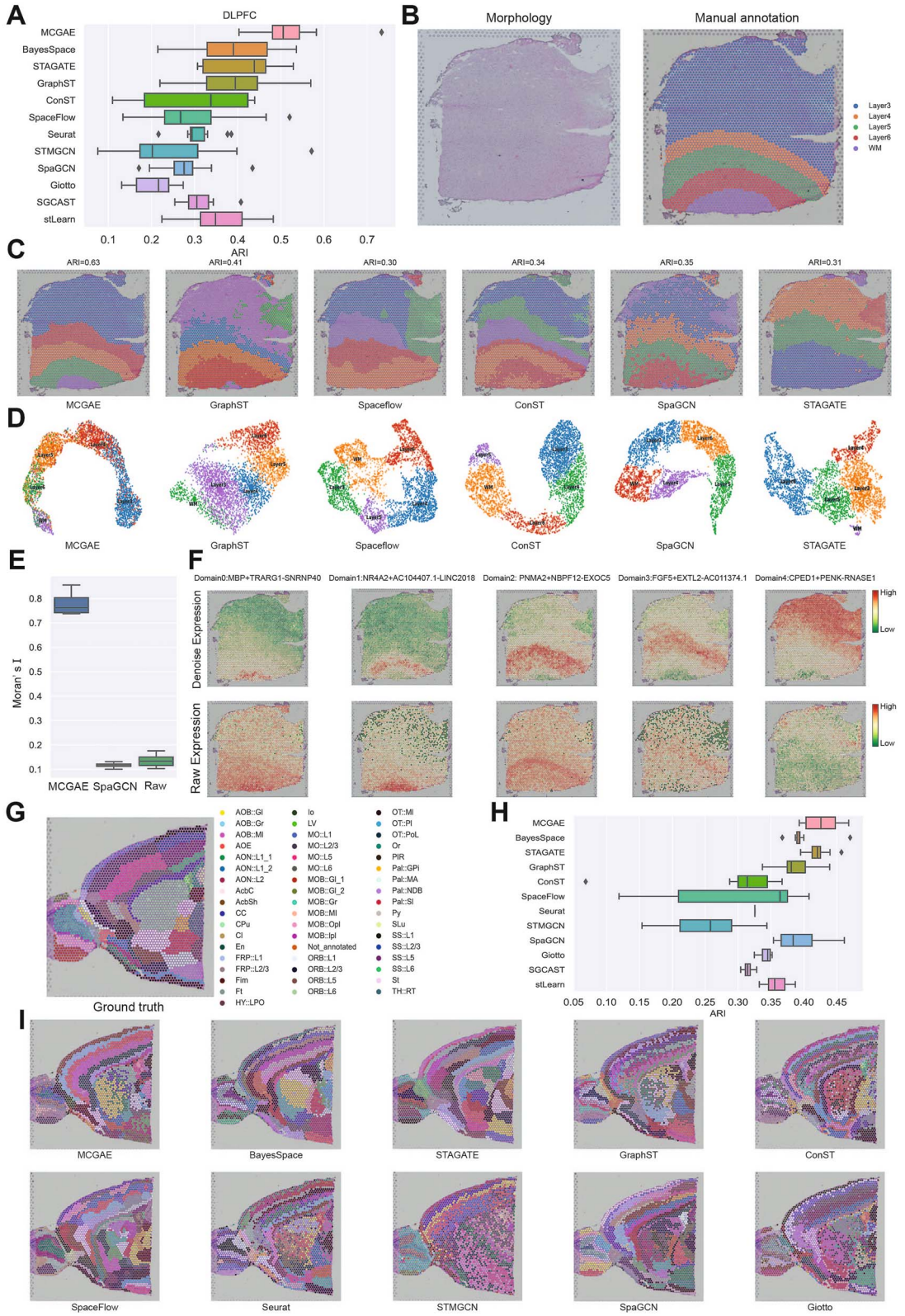

Figure 2. MCGAE has shown improvements in spatial domain recognition on both DLPFC and mouse brain anterior datasets. (A) Boxplot displays the ARI scores of the twelve methods applied to the 12 DLPFC slices. In the boxplot, the central line represents the median score, the box boundaries represent the upper and lower quartiles, and the whiskers extend up to 1.5 times the interquartile range, similar to what is shown in (E) and (H). (B) H&E image and manual annotation from the original study. (C) The clustering results for slice 151,669 of the DLPFC dataset were obtained using MCGAE, GraphST, Spaceflow, ConST, SpaGCN, and STAGATE. Clustering results for other algorithms can be found in Supplementary Fig. S5. (D) Trajectory inference was generated using embeddings from MCGAE, GraphST, Spaceflow, ConST, SpaGCN, and STAGATE for the DLPFC section 151669. (E) Boxplot is used to display the Moran's I values for SVGs detected by MCGAE, SpaGCN and raw (raw expression), and for slice 151,669. (F) the visualization of denoised and raw expression patterns of layer marker genes. (G) Spatial domain clustering based on the annotation labels. (H) Boxplot displays the ARI scores of the twelve methods applied to the mouse brain anterior dataset. (I) The clustering results of the mouse brain anterior dataset were obtained using MCGAE, BayesSpace, STAGATE, GraphST, ConST, SpaceFlow, Seurat, STMGCN.

consumption of approximately 768.95 MB, which is comparable to SGCAST (Supplementary Fig. S7D and E).

Multi-view contrastive graph autoencoder accurately delineates spatial domains within the human breast cancer and colorectal cancer datasets utilizing histological images

In the Human Breast Cancer (HBC) dataset, we investigated MCGAE's performance when incorporating image information and validated the advantages of denoised data. Importantly, integrating tumor images augments this analysis, markedly improving the precision in identifying spatial domains within the tumor microenvironment. Utilizing existing H&E stained images and annotation information (Fig. 3A), we assessed the performance of each algorithm using ARI. Notably, integrating H&E stained images with MCGAE resulted in a significant increase in the ARI score, rising from an average of 0.53–0.64, marking an improvement of 20.75%. Additionally, SpaGCN and STAGATE also demonstrated robust performance, with ARI averages of 0.56 and 0.57, respectively (Fig. 3B and Supplementary Fig. S8). Domain clustering visualizations reveal MCGAE's accurate identification of most tumor domains, while without image integration, the identification of tumor domains is less precise. The GraphST method accurately identifies tumor border regions, with correct identification of IDC_4 and DCIS/LCIS_4 (Fig. 3C). Subsequently, we visualized specific tumor domains characterized by marker genes, including the marker BMERB1 within the IDC_4 region. Through the reconstruction of the raw gene expression data, MCGAE achieved more precise domain patterns of marker genes (Fig. 3D).

To assess MCGAE's tumor data analysis capabilities, we compared it with SpaGCN, GraphST, and STAGATE using data from colorectal cancer patients [36]. Annotations from an oncology pathologist highlighted a small tumor domain on an H&E stained slide with low-resolution issues in Region X, showing a clustered distribution of darker-stained cells and vacuolar changes in tumor cells post-chemotherapy (Fig. 3E). MCGAE effectively identified tumor areas and grouped disputed regions with lymphatic clusters, suggesting their lymphatic nature. In contrast, STAGATE misidentified muscle tissue as lymphatic structures, and MCGAE without image data failed to distinguish tumor domains accurately. With image integration, MCGAE precisely separated lymphatic structures from tumor domains and showed superior accuracy in recognizing other tissue structures like vessels (Fig. 3F and Supplementary Fig. S9).

By selecting three well-known tumor markers, we discerned that the denoised data more effectively unveiled domain patterns (Fig. 3G). In the context of the previously ambiguous region X, which demonstrates clustering affinity with lymphoid domains, our enrichment analysis of SVGs within this region underscored a pronounced engagement with immune system functionalities, notably those pertaining to lymphocytes and immune responses. Accordingly, by synthesizing insights from both clustering patterns and pathway enrichment analyses, it becomes evident that this area is more aptly characterized as a lymphoid structural domain, thereby clarifying its biological identity and significance (Fig. 3H).

In summary, the integration of MCGAE with tumor imaging data holds promising potential for advancing our understanding of tumor biology. Through its precise identification of spatial domain structures, MCGAE empowers clinicians to make informed decisions regarding treatment strategies, paving the way for more personalized and effective therapies for cancer patients.

Multi-view contrastive graph autoencoder's spatial clustering of colorectal cancer liver metastasis data elucidated the phenomenon of tumor invasion

The spatial organization of tissue domains in tumors is complex, with lesions containing diverse cancer cells and pathological changes. Clinically, it's crucial to evaluate tumor initiation and progression. Accurately delineating tumor margins and assessing the invasive status through the microenvironment is essential. Here, we applied MCGAE to a complex dataset derived from colorectal cancer liver metastasis using the 10× platform [36]. Through annotations by an oncology pathologist, we discovered that the primary tumor area of colorectal cancer was predominantly infiltrated by tumor cells (Fig. 4A). Initially, when applying various methods to perform spatial domain clustering on this dataset, we observed that MCGAE and STAGATE exhibited more defined clustering patterns (Fig. 4B and Supplementary Fig. S10). Interestingly, nearly all methods identify a domain Y that encompasses both connective tissue and tumor, as depicted by the black dashed box in domain of Fig. 2B. This prompts us to question whether it is due to tumor invasion that they are consistently grouped into a single domain.

Initially, we performed t-SNE dimensionality reduction on all spots in the sample, colored by manual annotation and region Y (Fig. 4C and Supplementary Fig. S11). Interestingly, in the t-SNE plot, we found that the spots of region Y predominantly located in the tumor category, although a considerable number of spots were scattered among the connective tissue category (Fig. 4C). Subsequent t-SNE analysis of the spots of region Y alone revealed three main clusters (Fig. 4D). Differential gene expression and enrichment analysis revealed that Cluster 0, the most populous, mainly consisted of tumor cells, with pathways related to the collagen-containing extracellular matrix and angiogenesis—key factors in the tumor microenvironment and cell invasion (Fig. 4D and E). Cluster 1, located within the tumor region, showed enrichment in pathways linked to energy metabolism and mitochondrial functions, aligning with its location. Cluster 2, found primarily in connective tissue, displayed genes linked to both normal tissue function and tumor development, including responses to topologically incorrect proteins and enhanced defense mechanisms (Fig. 4E and F). The analysis above confirms our initial hypothesis that the development and invasion of the tumor lead to the manifestation of tumor molecular characteristics in non-tumor domains. Although these features are not conspicuously present in histological imaging, they can still be corroborated through various molecular characterizations. This further underscores the significance of integrating multiple modalities of ST information.

As described previously, the primary lesion of the patient indeed underwent invasion and metastasis, particularly severe in the liver. Next, we explored the characteristics of tumor development and invasion within the liver metastatic lesions of the patient. Through annotations by an oncology pathologist, we observed that half of the colorectal cancer liver metastasis area consists of tumor tissue, while the other half consists of liver tissue, with clear boundaries (Fig. 4G). Various algorithms effectively distinguished the boundaries between liver and tumor domains, with MCGAE identifying domains located on both sides of the boundary (Fig. 4H and Supplementary Fig. S12). The reconstructed expression data module facilitated the discernment of marker gene expression patterns (Fig. 4I). Hierarchical clustering analysis revealed that domains 2, 8, 16, and 17 exhibited clustering patterns distinct from both liver and tumor tissues (Fig. 4J).

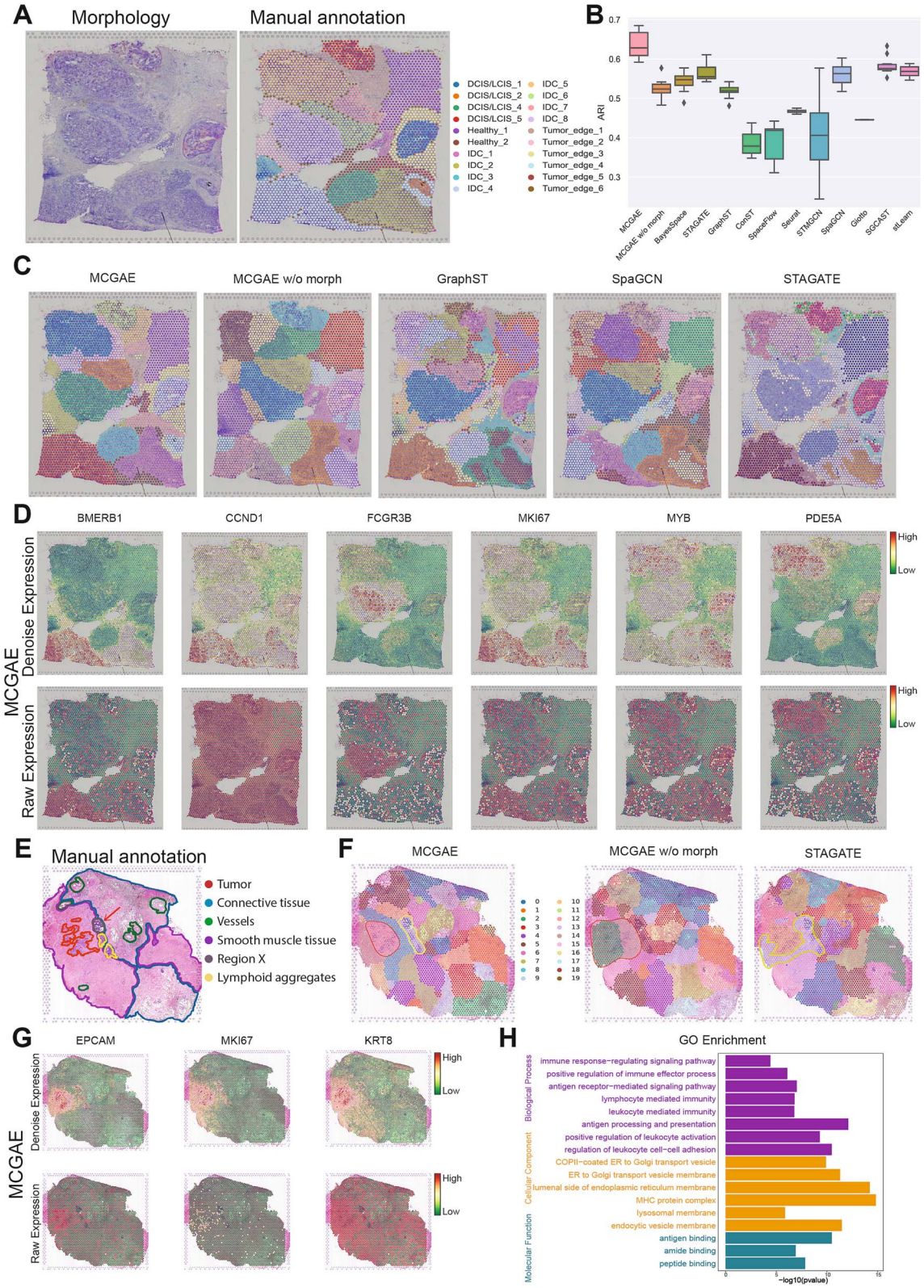

Figure 3. MCGAE accurately identifies spatial domains in the HBC and colorectal cancer datasets using histological images. (A) Morphological images and manual annotations for HBC data. (B) Boxplot displays the ARI scores of the twelve methods applied to the HBC dataset. In the boxplot, the central line represents the median score, the box boundaries represent the upper and lower quartiles, and the whiskers extend up to 1.5 times the interquartile range. (C) Clustering results for the HBC dataset were obtained using MCGAE, MCGAE without morphological information, GraphST, SpaGCN, and STAGATE, with additional algorithm results in Supplementary Fig. S8. (D) The spatial expression patterns of SVGs detected by MCGAE were visualized using both the denoised and original expression data. (E) Colorectal cancer morphological images with an oncology pathologist' annotation. (F) The clustering results for the colorectal cancer dataset were obtained using MCGAE, MCGAE without morphological information, and STAGATE, with additional algorithm results presented in Supplementary Fig. S9. The additional contour lines in the figure represent the contours of the domains of interest for each method, with annotations consistent with those in (E) (G) Visualization of marker genes for tumor using denoised and original expression data from MCGAE. (H) the GO enrichment plot for SVGs in tumors located at the boundary between tumor and connective tissue depicts biological processes, cellular components, and molecular functions using different shading patterns to distinguish the categories.

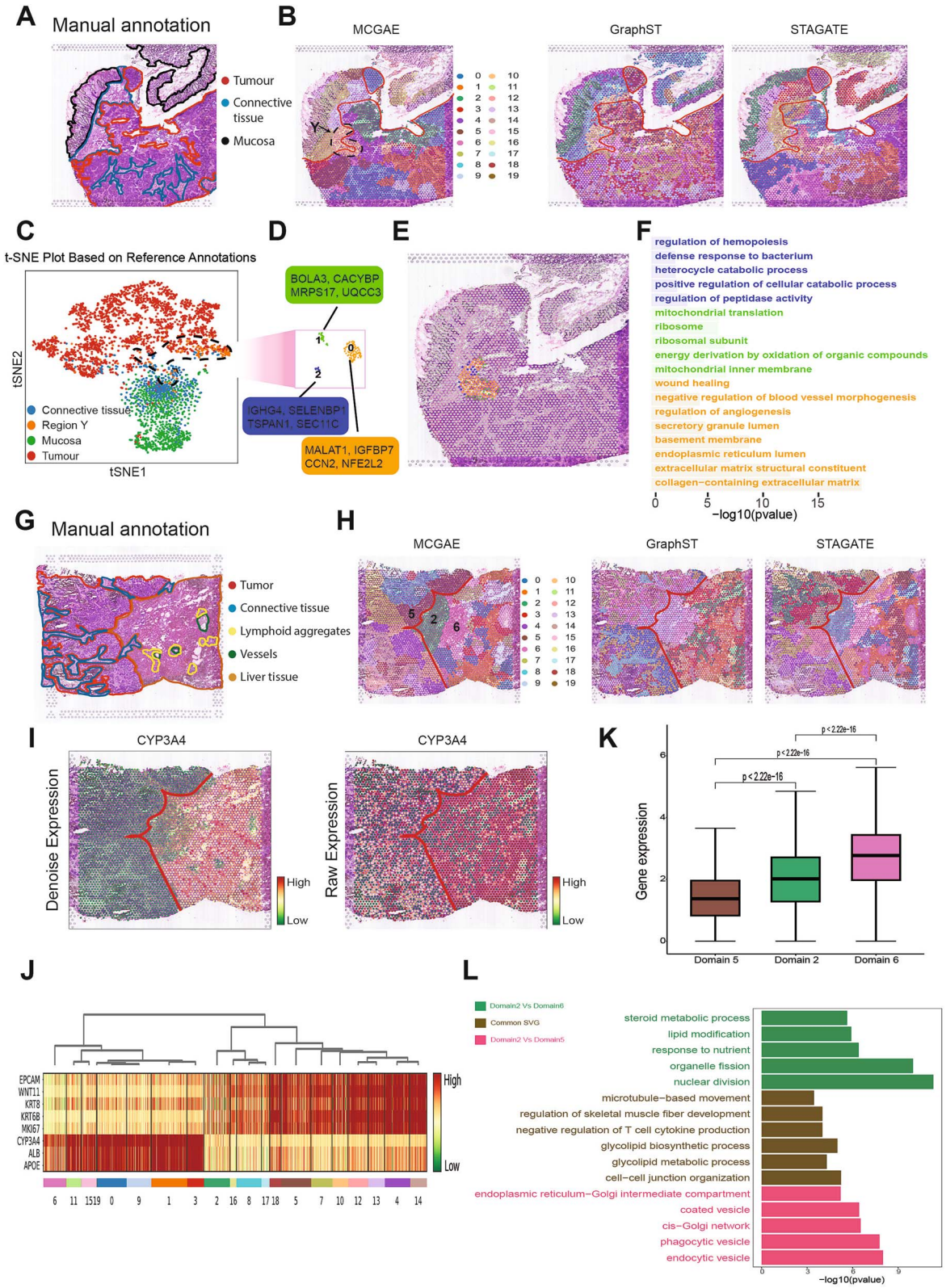

Figure 4. MCGAE facilitates the study of tumor invasiveness in colorectal cancer liver metastasis cases. (A) Colorectal cancer morphological images with an oncology pathologist's annotation. (B) Clustering results for the colorectal cancer dataset were obtained using MCGAE, GraphST, and STAGATE, with additional algorithm results in Supplementary Fig. S10. (C) A t-SNE plot of all spots in the sample, colored according to manual annotation and region Y. (D) A refined t-SNE plot of region Y, segmented into three distinct categories, highlighting the top DEGs for each category. (E) The three categories of region Y projected back onto the original ISH slides. (F) A GO enrichment plot for DEGs across the three categories in region Y, with colors corresponding to each category. (G) Liver cancer morphological images with an oncology pathologist's annotation. (H) Clustering results for the liver cancer dataset were obtained using MCGAE, GraphST, and STAGATE, with additional algorithm results in Supplementary Fig. S12. (I) Denoised and original expression patterns of metabolism-related gene CYP3A4 in liver cancer dataset. (J) The heatmap of common marker genes in liver cancer and normal liver tissue. (K) A boxplot showing the expression differences of common SVGs across domains 5, 2, and 6. (L) Enrichment analysis for SVGs in domain 2 relative to domain 5 and 6, along with their shared SVGs.

Histopathological examination revealed that domains 8, 16, and 17 were situated within the fibrous tissue layer of the tumor domain, while domain 2 was positioned at the midpoint between tumor and liver tissues, with half of it consisting of fibrous tissue.

We analyzed differential gene expression in Domain 2 versus its neighboring tumor Domain 5 and liver Domain 6, noting distinct expression patterns among shared SVGs, suggesting tumor progression variations and liver function (Fig. 4K). Pathway enrichment analysis revealed that SVGs in Domain 2 compared to Domain 5 involve processes like vesicle formation and cellular engulfment, indicating interactions within the tumor microenvironment. Conversely, SVGs between Domain 2 and Domain 6 primarily concern metabolic pathways, reflecting liver function. Common genes include those involved in tumor-relevant pathways like cell–cell junction organization and fibrous tissue processes, matching the spatial characteristics of Domain 2 (Fig. 4L).

In this intricate and detailed example of colorectal cancer liver metastasis, the utilization of MCGAE reveals a range of valuable functionalities. Firstly, MCGAE proves its effectiveness in extracting essential tumor image information, enabling a comprehensive understanding of the spatial distribution and characteristics of the metastatic lesions within the liver. Secondly, MCGAE employs expression denoising techniques to refine and enhance the accuracy of gene expression data, ensuring reliable and robust analysis of molecular profiles. Lastly, MCGAE excels in identifying domain-specific SVGs, shedding light on the unique molecular processes and interactions taking place within distinct tumor microenvironments. Together, these functionalities of MCGAE contribute to a deeper comprehension of the intricate dynamics and heterogeneity of colorectal cancer liver metastasis.

Multi-view contrastive graph autoencoder handles datasets from both stereo-seq and slide-seq technologies and constructs 3D spatial domains

Leveraging the diversity of ST technologies, we used a Stereo-seq dataset to analyze coronal sections of the mouse olfactory bulb [11]. We began by utilizing established annotations to identify key layers such as the olfactory nerve layer and external plexiform layer, among others (Fig. 5A). Our evaluation found that while MCGAE, GraphST, and STAGATE all effectively distinguished the outer layers, GraphST confused the glomerular layer with the mitral cell layer, and STAGATE missed internal structures like the rostral migratory stream (Fig. 5B and Supplementary Fig. S13). In contrast, MCGAE accurately differentiated both outer and inner layers. Validation with marker genes confirmed MCGAE's results, showing a strong match between identified clusters and known markers (Fig. 5C).

Currently, various technologies capable of producing consecutive sections, integrating and reconstructing their 3D expression domains, allow for the genuine spatial observation of biological developmental changes. We applied MCGAE to pseudo-3D spatial data from seven hippocampal datasets using Slide-seq [10], focusing on 'cord-like' structures within the hippocampus (Fig. 5D). To adapt MCGAE for 3D identification, we constructed a large adjacency matrix for all batches, using adjacency distance between sections (Fig. 5E and Supplementary Note 8). Initially, MCGAE faced challenges in 3D domain identification due to continuous slices, although most spots were discernible on the 2D UMAP plot (Fig. 5F). By integrating adjacent edges between neighboring sections, we enhanced MCGAE's ability to accurately delineate tissue structures and distinguish different regions on the UMAP plot (Fig. 5G). These adaptations demonstrate MCGAE's utility in reconstructing 3D tissue models and extracting precise 3D expression patterns.

These examples underline the flexibility and robustness of MCGAE in processing and analyze spatial genomics data across various dimensions and sequencing platforms. Its modular modeling approach not only facilitates a deeper understanding of complex biological structures and processes but also positions MCGAE as a promising universal algorithm framework for ST. This versatility opens new avenues for research in tissue architecture and function, offering insights that could lead to novel discoveries in biology and medicine.

Discussion

Biological phenomena unfold within spatial contexts where cells interact within the three-dimensional structures of tissues, crucial for disease and tissue functionality. Spatial transcriptomics captures diverse cell information, including gene expression, locations, and histological images. Yet, many methods have not fully utilized this multi-modal data. In tumor research, particularly, histological images are vital for accurately depicting tumor status and boundaries.

In this study, we introduce the Multi-View Contrastive Graph Auto-Encoder (MCGAE), a model designed for spatial domain detection in ST. By integrating multi-modal information through multi-view processing, contrastive graph convolutional neural networks, and attention modules, MCGAE improves feature embeddings and data reconstruction accuracy. The model also utilizes a pre-trained ResNet50 network for image feature extraction, enhancing its capability to handle complex tumor data. In the methods employed by SpaGCN, ConST, DeepST, stLearn, and TIST, image data plays a crucial role in enhancing the analysis of ST. SpaGCN incorporates image information into the spatial location data, expanding it into a three-dimensional space, thereby influencing the similarity/adjacency matrix A for spots. ConST utilizes pre-trained neural networks to extract image features, which are then concatenated with the original gene expression matrix X, directly affecting the spot/node representations. DeepST and stLearn construct spot similarity matrices using image data, which are subsequently combined with gene expression and spatial similarity matrices to enhance the original gene expression, thus impacting the feature matrix X. TIST, on the other hand, directly fuses different modality-specific similarity matrices to create a multi-modality similarity network for downstream analysis. Our approach similarly influences gene expression X by integrating image representations extracted using a pre-trained ResNet with other modality-derived representations.

To validate the performance of our model, we evaluated MCGAE on eight real datasets from five platforms and two simulated datasets from one platform. Experimental results demonstrate that our model exhibits competitive performance compared to existing methods. Additionally, our algorithm provides denoised expression data for downstream analysis, aiding in the identification of spatially differential genes and spatial domain expression patterns. Through multi-view modeling of expression data and spatial coordinates, MCGAE facilitates the capture of biological signals, enabling accurate identification of tumor regions even in complex medical tumor datasets. This advancement contributes to the study of tumor invasion by providing valuable insights.

However, there are some spaces to improve for our methods. Primarily, the computation of the adjacency matrix for spatial

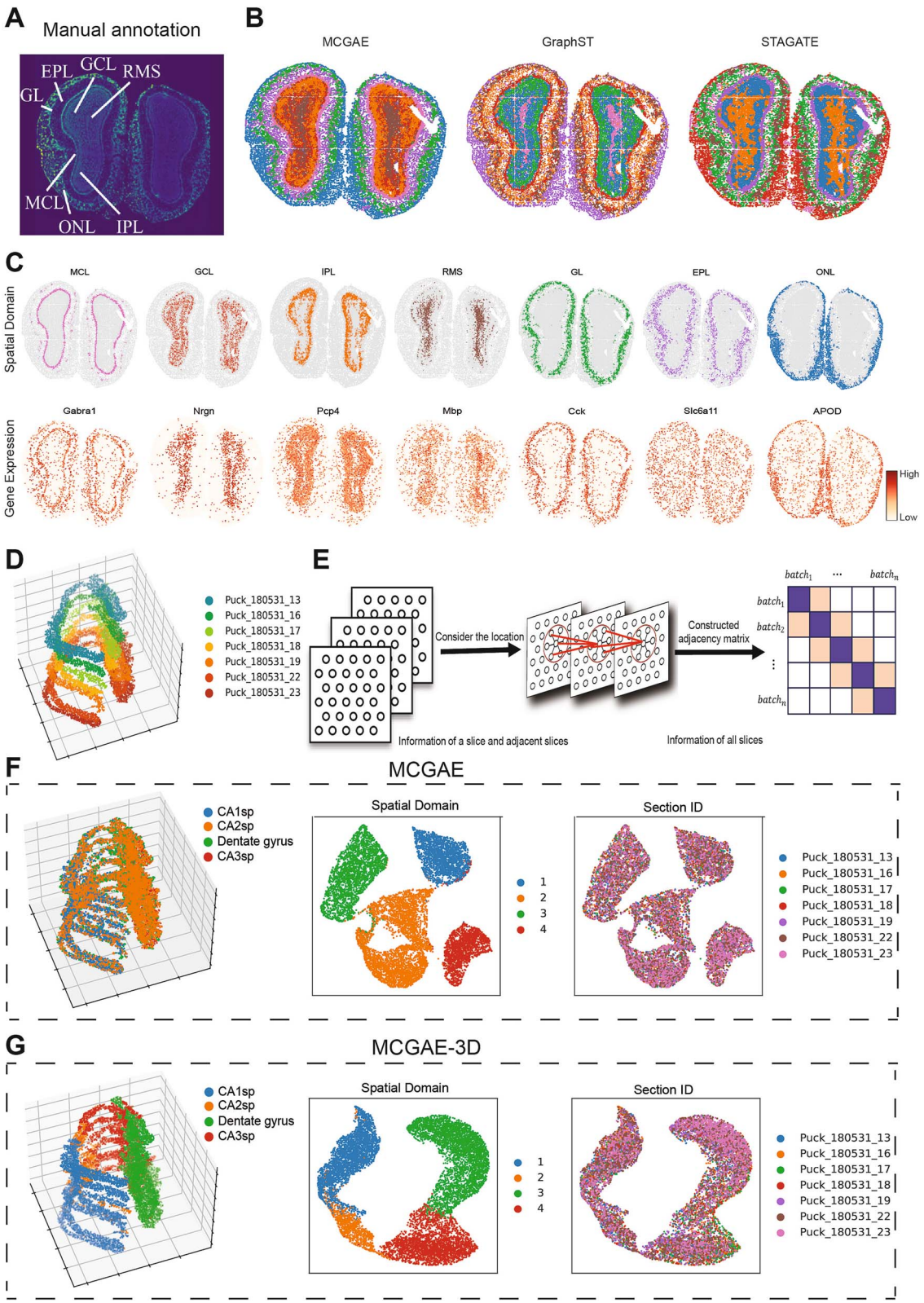

Figure 5. MCGAE handles datasets from both stereo-seq and slide-seq technologies and can construct 3D spatial domains. (A) Laminar organization of the mouse olfactory bulb annotated using the DAPI-stained image. (B) Spatial domains identified by MCGAE, GraphST and STAGATE in the mouse olfactory bulb Stereoseq data. (C) Visualization of the spatial domains identified by MCGAE and the corresponding marker gene expressions. The identified domains are aligned with the annotated laminar organization of the mouse olfactory bulb. (D) Visualization of the 3D hippocampal volume stacked by seven aligned consecutive sections profiled by slide-seq. (E) 3D spatial domain recognition is achieved by extending MCGAE, where we construct a large adjacency matrix covering all batches and focus on the adjacency distances between adjacent sections. (F) Cluster assignments using conventional adjacency matrix calculation in MCGAE and UMAP plots based on MCGAE embeddings. Spots are color-coded for identified spatial domains and section IDs. (G) Cluster assignments using extended adjacency matrix calculation in MCGAE-3D and UMAP plots based on MCGAE-3D embeddings. Spot colors represent spatial domains and section IDs.

position information incurs a computational complexity of $O(n^2)$, which may rapidly exhaust computational resources with an increasing number of samples. To address this, we could consider using a neighbor sampling strategy, where a subset of spots is selected along with their neighbors to construct the adjacency matrix. This approach would reduce computational demands while preserving the global structure, making it more feasible to handle large-scale ST data with GCNs. Secondly, although we addressed batch effects in the 3D example by considering the distance between neighboring regions, this strategy may not suffice for other complex datasets, necessitating further refinement in future research.

MCGAE stands out as a highly competitive method, and its modular modeling approach makes it scalable for future extensions. It is foreseeable that the field of spatial genomics will evolve toward multiomics, 3D spatial genomics, and spatiotemporal genomics. As the number of samples is expected to rapidly increase in the future, MCGAE will also need to address the challenge of large-scale datasets. This can be achieved by employing subgraph sampling or deep multichannel GCN strategies in future developments.

Key Points

- Comprehensive Integration of Multimodal Data: MCGAE is adept at integrating diverse types of data, including gene expression, spatial coordinates, and morphological information. This integration is critical for accurately identifying spatial domains within complex tissue environments, enhancing the understanding of tissue heterogeneity.
- Advanced Computational Framework: The MCGAE utilizes a sophisticated computational framework that employs multi-view representations, modular modeling, and contrastive graph convolutional networks. These features allow for the effective processing of spatial transcriptomics data and the generation of modality-specific spatial representations.
- Enhanced Data Processing Capabilities: By incorporating attention mechanisms and morphological image features, MCGAE significantly improves its ability to process and interpret multimodal data. This enriched framework facilitates more accurate spatial domain detection, data denoising, trajectory inference, and 3D feature extraction, outperforming traditional methods..
- Applications in Cancer Research: Specifically applied to colorectal cancer liver metastases, MCGAE successfully integrates histological and gene expression data to identify regions of tumor invasion and characterize cellular molecular regulation. This capability demonstrates MCGAE's potential as a powerful tool for advancing cancer research and understanding complex diseases.

Supplementary Data

Supplementary data is available at *Briefings in Bioinformatics* online.

Acknowledgements

We thank Professor Yingbin Liu from Renji Hospital affiliated to Shanghai Jiao Tong University for their helpful discussions and comments during the study.

Author contributions

Conceptualization: W.W., L.N.C., C.C.Z. Methodology: C.M.Z., Y.W.Y. Software: Y.W.Y., C.M.Z. Formal Analysis: Y.W.Y., Z.N.L. Writing – Original Draft: Y.W.Y., C.M.Z., C.C.Z., K.A., L.N.C., W.W. Supervision: W.W., L.N.C., C.C.Z, K.A.

Conflict of interest: The authors declare no conflict of interest.

Funding

This work was supported by National Key R&D Program of China [2021YFF0703802, 2022YFA1004800]; the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB38040400); National Natural Science Foundation of China [81870187, 81911530167, 31930022, T2350003, T2341007, 12131020, 62202120]; Science and Technology Commission of Shanghai Municipality [23JS1401300]; Major Science and Technology R&D Project of the Science and Technology Department of Jiangxi Province [20213AAG01013]] to W.W.; JST Moonshot R&D (JPMJMS2021); AMED under Grant Number JP23dm0307009; Institute of AI and Beyond of the University of Tokyo; International Research Center for Neurointelligence (WPI-IRCN) at the University of Tokyo Institutes for Advanced Study (UTIAS); JSPS KAKENHI Grant Number JP20H05921; R&D project of Pazhou Lab (Huangpu) under Grant 2023 K0602.

Data availability

All data analyzed in this paper are available in raw form from their original authors. Specifically, the DLPFC dataset is accessible within the spatialLIBD package (http://spatial.libd.org/spatialLIBD). The MouseBrain dataset is collected from the 10x Genomics website (https://support.10xgenomics.com/spatial-gene-expression/datasets). Slide-seq datasets are available at https://portals.broadinstitute.org/single_cell/study/slide-seq-study. The processed Stereo-seq data from mouse olfactory bulb tissue is accessible on https://github.com/JinmiaoChenLab/SEDR_analyses. The Human Breast Cancer dataset is collected from the 10x Genomics website (https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Breast_Cancer_Block_A_Section_1). The Colorectal Cancer Liver dataset is sourced from the mentioned article [35] (http://www.cancerdiversity.asia/scCRLM). The raw data for STARmap are available at https://www.dropbox.com/sh/f7ebheru1lbz91s/AADm6D54GSEFXB1feRy6OSASa/visual_1020/20180505_BY3_1kgenes?dl=0&subfolder_nav_tracking=1. Image-based data (MERFISH) from SpatialDB [37] [https://www.spatialomics.org/SpatialDB/].

Code availability

An open-source Python implementation of the MCGAE toolkit is accessible at https://github.com/yiwen-yang/MCGAE.

References

1. Rao A, Barkley D, Franca GS. *et al.* Exploring tissue architecture using spatial transcriptomics. *Nature* 2021;**596**:211–20. https://doi.org/10.1038/s41586-021-03634-9.
2. Armingol E, Officer A, Harismendy O. *et al.* Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* 2021;**22**:71–88. https://doi.org/10.1038/s41576-020-00292-x.

3. Ji N, van Oudenaarden A. Single molecule fluorescent in situ hybridization (smFISH) of C. Elegans worms and embryos. *WormBook* 2012;1–16. https://doi.org/10.1895/wormbook.1.153.1.
4. Chen KH, Boettiger AN, Moffitt JR. *et al.* RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;**348**:aaa6090. https://doi.org/10.1126/science.aaa6090.
5. Moffitt JR, Bambah-Mukku D, Eichhorn SW. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 2018;**362**:eaau5324. https://doi.org/10.1126/science.aau5324.
6. Lubeck E, Coskun AF, Zhiyentayev T. *et al.* Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 2014;**11**:360–1. https://doi.org/10.1038/nmeth.2892.
7. Lee JH, Daugharthy ER, Scheiman J. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* 2014;**343**:1360–3. https://doi.org/10.1126/science.1250212.
8. Wang X, Allen WE, Wright MA. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;**361**:eaat5691. https://doi.org/10.1126/science.aat5691.
9. Ji AL, Rubin AJ, Thrane K. *et al.* Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* 2020;**182**:497–514.e22. https://doi.org/10.1016/j.cell.2020.05.039.
10. Rodriques SG, Stickels RR, Goeva A. *et al.* Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;**363**:1463–7. https://doi.org/10.1126/science.aaw1219.
11. Chen A, Liao S, Cheng M. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;**185**:1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.
12. Moncada R, Barkley D, Wagner F. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020;**38**:333–42. https://doi.org/10.1038/s41587-019-0392-8.
13. Satija R, Farrell JA, Gennert D. *et al.* Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**:495–502. https://doi.org/10.1038/nbt.3192.
14. Dries R, Zhu Q, Dong R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;**22**:78. https://doi.org/10.1186/s13059-021-02286-2.
15. Zhao E, Stone MR, Ren X. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* 2021;**39**:1375–84. https://doi.org/10.1038/s41587-021-00935-2.
16. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* 2022;**13**:1739. https://doi.org/10.1038/s41467-022-29439-6.
17. Li J, Wang J, Lin Z. SGCAST: symmetric graph convolutional auto-encoder for scalable and accurate study of spatial transcriptomics. *Brief Bioinform* 2024;**25**:bbad490. https://doi.org/10.1093/bib/bbad490.
18. Ren H, Walker BL, Cang Z. *et al.* Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat Commun* 2022;**13**:4076. https://doi.org/10.1038/s41467-022-31739-w.
19. Long Y, Ang KS, Li M. *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun* 2023;**14**:1155. https://doi.org/10.1038/s41467-023-36796-3.
20. Zong Y, Yu T, Wang X. *et al.* conST: an interpretable multi-modal contrastive learning framework for spatial transcriptomicsbioRxiv. 2022. 2022.2001.2014.476408.
21. Shi X, Zhu J, Long Y. *et al.* Identifying spatial domains of spatially resolved transcriptomics via multi-view graph convolutional networks. *Brief Bioinform* 2023;**24**:bbad278. https://doi.org/10.1093/bib/bbad278.
22. Hu J, Li X, Coleman K. *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**:1342–51. https://doi.org/10.1038/s41592-021-01255-8.
23. Pham D, Tan X, Balderson B. *et al.* Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat Commun* 2023;**14**:7739. https://doi.org/10.1038/s41467-023-43120-6.
24. Xu C, Jin X, Wei S. *et al.* DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res* 2022;**50**:e131. https://doi.org/10.1093/nar/gkac901.
25. Shan Y, Zhang Q, Guo W. *et al.* TIST: transcriptome and histopathological image integrative analysis for spatial transcriptomics. *Genom Proteom Bioinform* 2022;**20**:974–88. https://doi.org/10.1016/j.gpb.2022.11.012.
26. Wang B, Luo J, Liu Y. *et al.* Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. *Brief Bioinform* 2023;**24**:bbad262. https://doi.org/10.1093/bib/bbad262.
27. He K, Zhang X, Ren S. *et al.* Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–8, 2016.
28. Zhang C, Yang Y, Tang S. *et al.* Contrastively generative self-expression model for single-cell and spatial multimodal data. *Brief Bioinform* 2023;**24**:bbad265. https://doi.org/10.1093/bib/bbad265.
29. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016. arXiv:1609.02907.
30. Veličković P, Fedus W, Hamilton WL. *et al.* Deep graph Infomax. 2018. arXiv:1809.10341.
31. Paszke A, Gross S, Massa F. *et al.* PyTorch: An imperative style, high-performance deep learning library. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 8026–37. Red Hook, NY, USA: Curran Associates Inc., 2019, Article 721.
32. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.
33. Paszke A, Gross S, Massa F. *et al.* Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inform Processing Syst* 2019;**32**.
34. Maynard KR, Collado-Torres L, Weber LM. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 2021;**24**:425–36. https://doi.org/10.1038/s41593-020-00787-0.
35. Li H, Calder CA, Cressie N. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geograp Anal* 2007;**39**:357–75. https://doi.org/10.1111/j.1538-4632.2007.00708.x.
36. Wu Y, Yang S, Ma J. *et al.* Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov* 2022;**12**:134–53. https://doi.org/10.1158/2159-8290.CD-21-0316.
37. Fan Z, Chen R, Chen X. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res* 2020;**48**:D233–d237. https://doi.org/10.1093/nar/gkz934.