stMMR: accurate and robust spatial domain identification from spatially resolved transcriptomics with multi-modal feature representation

Daoliang Zhang[1,†], Na Yu[1,†], Wenrui Li[2], Xue Sun[1], Qi Zou[1], Xiangyu Li[3], Zhiping Liu[1], Zhiyuan Yuan[4,*], Wei Zhang[1,*], Rui Gao[1,*]

[1] Center of Intelligent Medicine, School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China

[2] MOE Key Lab of Bioinformatics and Bioinformatics Division of BNRIST, Department of Automation, Tsinghua University, Beijing 100084, China

[3] School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

[4] Center for Medical Research and Innovation, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, China

[†]These authors contributed equally

*To whom correspondence should be addressed. Tel: +86 0531 88364052; Email: gaorui@sdu.edu.cn

Correspondence may also be addressed to Wei Zhang. Email: zw@sdu.edu.cn

Correspondence may also be addressed to Zhiyuan Yuan. Email: zhiyuan@fudan.edu.cn

Abstract

Deciphering spatial domains using spatially resolved transcriptomics (SRT) is of great value for the characterizing and understanding of tissue architecture. However, the inherent heterogeneity and varying spatial resolutions present challenges in the joint analysis of multi-modal SRT data. We introduce a multi-modal geometric deep learning method, named stMMR, to effectively integrate gene expression, spatial location and histological information for accurate identifying spatial domains from SRT data. stMMR uses graph convolutional networks (GCN) and self-attention module for deep embedding of features within unimodal and incorporates similarity contrastive learning for integrating features across modalities. Comprehensive benchmark analysis on various types of spatial data shows superior performance of stMMR in multiple analyses, including spatial domain identification, pseudo-spatiotemporal analysis, and domain-specific gene discovery. In chicken heart development, stMMR reconstruct the spatiotemporal lineage structures indicating accurate developmental sequence. In breast cancer and lung cancer, stMMR clearly delineated the tumor microenvironment and identified marker genes associated with diagnosis and prognosis. Overall, stMMR is capable of effectively utilizing the multi-modal information of various SRT data to explore and characterize tissue architectures of homeostasis, development and tumor.

Introduction

The advancement in spatially resolved transcriptomics (SRT) technologies has opened new avenues for a deeper understanding of the spatial architecture and functionality of tissues. By employing these advanced SRT techniques, we can conduct a precise examination of the spatial landscapes and transcriptional profiles for complex tissues. Currently, many SRT technologies have been developed, such as imaging-based and sequencing-based methods (1–7). Among these, techniques such as 10x Genomics Visium not only provide the spatial location and gene expression data for each spot but also acquire high-resolution hematoxylin and eosin (H&E) stained histology images of the tissue section, revealing richer information about the tissue organization. These technological advancements offer new insights into characterization of tissue architecture, enabling a more comprehensive understanding of tissue development and disease pathogenesis (7–9).

For SRT technologies capable of providing both gene expression data and histology images, the information from these different modalities reflects the structural information of tissues at various levels. Gene expression profiles reflect the difference of cell state between spots (10). Spatial location information provides the precise location of each spot, illustrating the complex landscape of tissue structure. Histological images display morphological features of cells, such as size, shape, and intercellular relationships, offering crucial clues for distinguishing functional domains (11). Although each of these modalities has its own strength, they complement each other, together forming a more comprehensive picture of tissue architecture. For instance, changes in gene expression are reflected not only at the molecular level but may also manifest in histological images as morphological alterations (12). Furthermore, the issues of sparsity and dropout in SRT data can be effectively addressed through integrating histological image data (13). By leveraging the interdependence between gene expression and morphological features, as well as the similarity in gene expression patterns among adjacent spots, we can enhance spatial signals and characterize tissue structure. Therefore, achieving a joint representation of multi-modal features in SRT is

key to improving data utilization, domain recognition accuracy, as well as functional interpretability.

However, the joint representation of features from different modalities in SRT is challenging. Firstly, these different modalities inherently possess significant heterogeneity. For instance, transcriptomic data are typically high-dimensional, quantified gene expression information, reflecting the gene activity in different spots or cells. In contrast, histology image is two-dimensional visual data, depicting the morphological and structural information of cells at different spots. This fundamental difference makes the direct fusion of these two types of modal difficult. Secondly, the disparity in data scale and resolution is also a crucial issue. Transcriptomic data reveals unique patterns of gene expression within spots or cells from a microscopic perspective. Conversely, histology image provides more macroscopic information on organization and morphology. This difference in scale complicates the establishment of spatial correspondence, thereby posing challenges in comparing and integrating these two categories of data.

Recently, a variety of cutting-edge computational methods have been developed to effectively address the challenge of joint representation of multi-modal SRT data. Specifically, BASS, BayesSpace and Giotto leverage spatial neighborhood information for enhancing the resolution of SRT data (14–16). CellCharter and PRECAST incorporate spatial contexts to correct batch effect for a better domain identification (17, 18). MENDER is a recently proposed multi-range cell context decipherer for ultra-fast tissue structure identification (19). CCST, STAGATE, SpaceFlow and GraphST utilize Graph Neural Networks (GNN) to integrate gene expression data with spatial information, achieving effective clustering of spots (20–23). However, these methods do not employ histology images, failing to fully enhance the interpretability of gene expression data through these images. This limitation often results in subsequent analyses that are less accurate and lack robustness. In contrast, recent pioneering studies like stLearn and DeepST have shown more significant progress (24, 25). These methods effectively integrate gene expression data with spatial neighborhood information and

morphological features extracted from histology images, demonstrating a stronger potential for application. Despite these methods demonstrating some capability in processing and interpreting multi-modal information in SRT data, they give less consideration to the complex global spot similarity across high-resolution, content-rich, and distinctive spatial multi-modal features. This limitation impedes their ability to accurately characterize spatial patterns and discover functional biological contexts in tissue.

To achieve precise identification of spatial domains, we introduce stMMR, a geometric deep learning method for effective representing multi-modal information in SRT data. stMMR utilizes spatial location information as a bridge to establish adjacency relationships between spots. It encodes gene expression data and morphological features extracted from histological images using GCN. stMMR proposed a novel strategy to achieve joint learning of intra-modal and inter-modal features. Within a certain modality, stMMR employs self-attention mechanisms to learn the relationships of different spots. For integrating cross-modal information, stMMR innovatively utilizes similarity contrastive learning along with the reconstruction of gene expression features and adjacency information.

To assess the capability of stMMR in representing multi-modal information, we conducted comprehensive tests on various SRT datasets, including samples profiled by 10x Visium, NanoString technology as well as Spatial Transcriptomics (ST) technology. We evaluated the effectiveness of stMMR with several state-of-the-art techniques in domain recognition tasks. The results showed that stMMR achieved significant success in domain identification. Additionally, the analysis of domain-specific genes across multiple datasets revealed that stMMR significantly achieves the enhancement of gene expression data and facilitates the understanding of domain-specific genes. Finally, we conducted thorough analyses on breast cancer datasets based on 10x Visium technology and lung cancer datasets based on NanoString technology (26). Our studies found that stMMR accurately identify tumor edges and tumor-infiltrating regions, demonstrating outstanding continuity in regional identification across different slices, thus proving its

potential value in clinical research. Overall, the stMMR method exhibits exceptional capability in the multi-modal feature representation of SRT, providing a powerful new tool for accurate and robust domain identification.

Materials and methods

Overview of stMMR

The multi-modal joint representation process of stMMR primarily consists of the following three steps: multi-modal feature embedding, feature fusion and feature reconstruction. The overall workflow of stMMR is illustrated in Figure 1.

Multi-modal feature embedding

The stMMR initially performs embedding on gene expression, spatial location, and histology image information. We begin by assuming the presence of SRT data comprising N spots. For tissue histological images, we extract pixel features corresponding to each spot using a pre-trained Vision Transformer (ViT) model (27), resulting in a feature matrix $\boldsymbol{H} \in \mathbb{R}^{N \times M}$, where M is the output dimension of the pre-trained ViT model (Supplementary Section 1.3). For gene expression data, we employ SeuratV3 to filter high variance genes and perform a log transformation on the expression levels of these genes, denoted as $\boldsymbol{G} \in \mathbb{R}^{N \times P}$, where P represents the number of high variance genes identified. Additionally, we encode the spatial location information of each spot, resulting in a position encoding matrix corresponding to each spot. Specifically, we used an undirected weighted graph to present SRT data. For any two spots, we posit that the closer their spatial distance, the greater their similarity. Consequently, we define the adjacency matrix $\boldsymbol{A}$ between any two spots i and j as follows:

$$\boldsymbol{A_{ij}} = exp\left(-\frac{d(i,j)^2}{2l^2}\right) \tag{1}$$

where $d(i,j)$ represents the Euclidean distance between spots i and j, and l is a

hyperparameter controlling the relationship between weight and distance. A larger value of l implies a faster decay of weight with increasing distance.

GCN is a geometric deep learning module frequently used in graph representation learning in recent years (28). They effectively integrate information from neighboring nodes to achieve efficient representation of target nodes. Subsequently, we employ encoders with two layers of GCNs to perform message passing and aggregation on pixel features and gene expression features, as shown in Eq.2:

$$\boldsymbol{E}^{(k)} = \widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{E}^{(k-1)}\boldsymbol{W}^{(k-1)} \tag{2}$$

where $\boldsymbol{E}^{(k)}$ and $\boldsymbol{E}^{(k-1)}$ represent the input and output of the GCN module, respectively. $\boldsymbol{E}^{(0)}$ corresponds to the pixel features $\boldsymbol{H}$ or gene expression features $\boldsymbol{G}$ for each spot. $\widetilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ denotes the adjacency matrix of the undirected graph, where $\boldsymbol{I}$ is the identity matrix. $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{W}^{(k-1)}$ are the weighted degree matrix and trainable parameter respectively. The pixel features and gene expression features obtained after passing through the encoder are denoted as $\boldsymbol{E}_H$ and $\boldsymbol{E}_G$.

Feature fusion

To effectively aggregate multi-modal information, we propose a novel feature fusion strategy. First, stMMR uses a normalized attention module to learn the relationships between spots in a single modality, as shown in Eq.3:

$$\boldsymbol{E}_A = \boldsymbol{E} \cdot softmax\left(\frac{\boldsymbol{E} \cdot \boldsymbol{E}^T}{\sqrt{\boldsymbol{d}}}\right) \tag{3}$$

where $\boldsymbol{E}$ represents the pixel features $\boldsymbol{E}_H$ or gene expression features $\boldsymbol{E}_G$ obtained in the previous step. The new features obtained through the attention module are $\boldsymbol{E}_{AH}$ and $\boldsymbol{E}_{AG}$. It is noteworthy that we use a nonlinear activation function and the Euclidean distance matrix $\boldsymbol{d}$ to normalize the weights. This approach effectively avoids the issue of local optima caused by excessively large weights for certain spots (29).

For cross-modal information, stMMR adopts a contrastive learning approach for feature fusion. Previous research indicates that histology information and gene

expression information share both similarities and complementary relationships (30–32). stMMR emphasizes the consistency between multiple modalities by constructing a cross-modal contrastive learning strategy. Specifically, stMMR maps the latent features of multiple modalities, $\boldsymbol{E}_H$ and $\boldsymbol{E}_G$, to a space using two fully connected neural networks, thereby obtaining hierarchical representations for both modalities, $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_G$, as shown in Eq.4:

$$\boldsymbol{Q} = Relu(\boldsymbol{W}_Q \boldsymbol{E} + \boldsymbol{b}_Q). \tag{4}$$

In Eq.4, $\boldsymbol{E}$ represents the pixel features $\boldsymbol{E}_H$ or gene expression features $\boldsymbol{E}_G$, and $\boldsymbol{Q}$ corresponds to $\boldsymbol{Q}_H$ or $\boldsymbol{Q}_G$. $\boldsymbol{W}_Q$ and $\boldsymbol{b}_Q$ are the parameters of the fully connected network.

After obtaining the low-dimensional features $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_G$ for the two modalities, we further employ a fully connected neural network to fuse these two modalities, as shown in Eq.5:

$$\boldsymbol{E}_{\boldsymbol{Q}} = \boldsymbol{W}_E \cdot concat(\boldsymbol{Q}_G, \boldsymbol{Q}_H) + \boldsymbol{b}_E \tag{5}$$

where $\boldsymbol{W}_E$ and $\boldsymbol{b}_E$ are the parameters of the fully connected network.

To enhance the consistency between $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_G$, we use a constraint as shown in Equation 6, replacing the loss of traditional contrastive learning:

$$L_{con} = \left\| \widetilde{\boldsymbol{Q}_G} \widetilde{\boldsymbol{Q}_G}^T - \widetilde{\boldsymbol{Q}_H} \widetilde{\boldsymbol{Q}_H}^T \right\|_2^2. \tag{6}$$

In this equation, $\widetilde{\boldsymbol{Q}_G}$ and $\widetilde{\boldsymbol{Q}_H}$ are the normalization matrices of $\boldsymbol{Q}_G$ and $\boldsymbol{Q}_H$, respectively.

Finally, we further integrate modality specific features $\boldsymbol{E}_{AH}$ and $\boldsymbol{E}_{AG}$ obtained from Eq.3 with the cross-modality features $\boldsymbol{E}_{\boldsymbol{Q}}$ obtained from Eq.5 to get the multi-modal feature representation $\boldsymbol{Z}$, as shown in the following equation:

$$\boldsymbol{Z} = \alpha \boldsymbol{E}_{\boldsymbol{Q}} + \beta \boldsymbol{E}_{AH} + \gamma \boldsymbol{E}_{AG} \tag{7}$$

In this equation, α, β, and γ are hyperparameters for adjusting the importance of features.

Feature reconstruction

Finally, to ensure that the embeddings $\boldsymbol{Z}$ learned by our model contain biologically information, we designed a reconstruction phase using $\boldsymbol{Z}$ to reconstruct the original features. Here, we primarily focus on reconstructing the adjacency matrix $\boldsymbol{A}$ and the gene expression profiles.

SRT data are characterized by high sparsity, discreteness, and variance greater than the mean, specifically manifested as a high number of genes expressed at zero (zero inflation) (33). Previous research has found that the zero-inflated negative binomial (ZINB) distribution can effectively characterize gene expression in SRT (34). Therefore, stMMR also adopts the ZINB to describe gene expression information. In simple terms, we estimate the parameters of the ZINB distribution for each gene (i.e., μ, θ, and π, Supplementary Section 1.4) through three different fully connected networks, as follows:

$$\boldsymbol{\Pi} = Sigmoid\left(\boldsymbol{W}_{\pi} f_D(\boldsymbol{Z})\right),$$

$$\boldsymbol{\Theta} = exp(\boldsymbol{W}_{\theta} f_D(\boldsymbol{Z})),$$

$$\boldsymbol{M} = exp(\boldsymbol{W}_{\mu} f_D(\boldsymbol{Z})),$$

where $\mathbf{M}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Pi}$ are the matrix forms of μ, θ, and π, representing the mean, dispersion, and dropout probability of the output from network respectively. The dropout probability ranges between 0 and 1. Due to the non-negative nature of the mean $\mathbf{M}$ and dispersion $\boldsymbol{\Theta}$, we use the exponential function, and for the dropout probability $\boldsymbol{\Pi}$, the Sigmoid function is used. f_D is a decoder with a fully connected layer.

Based on the aforementioned reconstruction of gene expression information, we further designed the loss function for the ZINB decoder, as shown below:

$$L_{ZINB} = -\log\left(ZINB(\boldsymbol{G}|\boldsymbol{M}, \boldsymbol{\Theta}, \boldsymbol{\Pi})\right) \tag{8}$$

For spatial neighborhood relationships, we adopted the concept of a graph auto-encoder to directly estimate the adjacency matrix (35, 36), as shown in Eq.9:

$$\boldsymbol{A}' = Sigmoid\left(\frac{\boldsymbol{Z} \cdot \boldsymbol{Z}^{\boldsymbol{T}}}{\|\boldsymbol{Z}\|_2 \cdot \|\boldsymbol{Z}^{\boldsymbol{T}}\|_2}\right) \tag{9}$$

Subsequently, we defined a function to calculate the regularization loss between the reconstructed matrix and the adjacency matrix, as in Eq.10:

$$L_{rec} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\boldsymbol{A}_{ij} - \boldsymbol{A}'_{ij}\right)^2 \tag{10}$$

This function quantifies the difference between the original adjacency matrix $\boldsymbol{A}$ and its reconstructed version $\boldsymbol{A}'$, thereby facilitating the accurate reconstruction of spatial relationships.

Objective function

Finally, we integrated Eq.6, Eq.8, and Eq.10 to formulate the final objective function, as shown in Eq.11:

$$L = a * L_{con} + b * L_{ZINB} + c * L_{rec} \tag{11}$$

In this equation, a, b, and c are hyperparameters that control the contribution of the different loss terms.

For detailed information on the training process and parameter settings, please refer to the Supplementary Section 1.5 and 1.6.

Datasets

All the dataset used in this study are listed in Supplementary Table S1. For detailed descriptions and processing procedures, please refer to Supplementary Section 2. Processed datasets are also available at SODB and can be loaded by PySODB (37, 38).

Benchmark methods

To demonstrate the effectiveness of the multi-modal feature representation in SRT data, we selected 7 different state-of-the-art methods for benchmarking comparison.

These methods include SCANPY (39), which utilizes only gene expression data; CCST, STAGATE, GraphST, and SpaceFlow, which employ both gene expression and spatial location information (20–23); and stLearn and DeepST, which incorporate all three modalities (24, 25). Methods that have already been compared in previous works are not included in our analysis (40–42).

Results

stMMR enhances detection of stratified architectural patterns in human dorsolateral prefrontal cortex (DLPFC) tissue

The spatial structure of the brain is closely related to its function, particularly evident in the layered organization of the human brain cortex. Cells in different cortical layers not only exhibit unique gene expression patterns but also vary in morphology, physiology, and connectivity (43). To explore the spatial structure arrangement of brain , we collected a 10x Visium dataset containing 12 dorsolateral prefrontal cortex (DLPFC) sections (44). Maynard et al. manually annotated different layers (layer 1-6) and white matter (WM) of the DLPFC based on morphological features and gene markers (Figure 2A) (44). This annotation served as the ground truth for comparing and analyzing the effectiveness of stMMR and other advanced spatial domain identification methods.

We initially compared the Adjusted Rand Index (ARI) levels of various methods across 12 slices of the DLPFC dataset, as shown in Figure 2B. The results reveal that stMMR outperformed other methods, achieving the highest ARI compared to manual annotations. This result also demonstrates that stMMR possesses the smallest variance across all slices. Notably, the results from STAGATE, CCST, and SpaceFlow show differences in the ARI across different slices, indicating that these methods are more sensitive to the difference of domains. Scanpy uses only gene expression information and shows the poorest performance. Methods like stlearn and DeepST, which integrate

histological information, are outperformed by GraphST and STAGATE. This underperformance might stem from insufficient integration of transcriptomic and imaging data.

Next, we conducted a detailed analysis for each slice (Figure 2C and Supplementary Figure S1). To demonstrate the results, we used slice 151509 as an example (Figure 2C-E). The results shows that DeepST struggle with rough segmentation between layers. CCST, SpaceFlow, and stLearn have issues with erroneous region identification. Although GraphST and STAGATE accurately discern the arrangement of different regions, these methods exhibit biases in identifying the boundaries between distinct domains. In this specific case, stMMR demonstrates exceptional domain identification results. We further utilize UMAP for low-dimensional visualization analysis of the results obtained from different methods (Figure 2D), to verify whether the embeddings can accurately encompass information on regional arrangement and boundaries. The analysis reveals that techniques such as stMMR, CCST, STAGATE, and stLearn effectively separate different domains. In contrast, GraphST, SpaceFlow and DeepST exhibit noticeable issues in layer boundaries. For instance, the boundaries between layers 2, 3, and 4 are confused.

Further, we conducted a detailed trajectory inference using the PAGA algorithm (45) for these methods (Figure 2E). The PAGA graphs indicates that stMMR, STAGATE, CCST, and stLearn performs well in predicting trajectory between adjacent layers. The other methods display confused results in this analysis.

Moreover, we identified domain-specific genes in the human brain cortex using differential expressed gene analysis. Genes such as AQP4 and HPCAL1 are recognized as layer-specific genes (Supplementary Table S2). These genes are enriched in multiple layers and have been confirmed through multiplex single-molecule fluorescent in situ hybridization (44).

Combining the insights from these various analyses, it is evident that stMMR remarkably effective in several critical tasks, including domain identification, pseudo-

spatiotemporal analysis and domain-specific genes discovery. These results adequately demonstrate effective capability of stMMR in integrating transcriptomic and histological data.

stMMR enhances spatial gene expression profiling and structural characterization

In SRT, the analysis of domain-specific genes holds significant importance. The functionality of complex tissues is closely related to distribution domain-specific genes. However, in SRT research, identifying domain-specific genes which have relationships with histological structures is challenging. This is primarily due to the presence of substantial noise in the gene expression profiles generated by SRT techniques. The noise mainly arises from technical issues during sequencing, such as the dropout event (46–48). To validate whether stMMR can enhance gene expression data through histological information, we analyzed domain-specific gene expression in 151509 slices of the DLPFC dataset. Inspired by the previous study (49), we validated the effectiveness of the stMMR in enhancing gene expression data by comparing the relationship between the original gene expression profile and the profile reconstructed through the ZINB decoder with manually annotated regions.

Compared to using original gene expression data, employing reconstructed gene expression data facilitates the identification of a larger number of domain-specific genes across all brain layers (Supplementary Table S2). For instance, when using the original gene expression data, the CACNA2D2 and ADCYAP1 were not identified as domain-specific genes in layer 3. However, after enhancement with stMMR, we were able to accurately detect the CACNA2D2 and ADCYAP1 as domain-specific genes in layer 3. Notably, previous research has found that in layer 3 of primates, the CACNA2D2 gene exhibits differential expression and is closely associated with several biological pathways, including calcium signaling and synaptic long-term depression (50). ADCYAP1 has also been proved to be a domain-specific gene in former study

(51). This suggests that the expression patterns after stMMR enhancement are more consistent with known neurobiological functions and pathological states.

We also conducted a more detailed analysis by combining gene expression levels with their spatial locations (Figure 3). We found that after enhancing gene expression with stMMR, more distinct expression patterns of domain-specific genes can be observed between brain layers. Specifically, Figure 3A demonstrates a clear spatial representation of domain-specific marker genes (ADAYAP1, CACNA2D2, CALB1, MARC1, MB and LPL) after data enhancement. For example, in the original data, the expression pattern of genes such as ADCYAP1, CACNA2D2 and MB are sparse, and the boundaries in spatial regions are blurred, making it difficult to discern a clear expression pattern (Figure 3A and B). However, after enhancement with stMMR, we can observe that CACNA2D2 and CALB1 exhibit much clearer expression patterns in layers 3 and 4. Additionally, the enrichment of MARC1, MB, and LPL in the white matter regions become more pronounced (Figure 3A and C). These results reflect that stMMR not only improves the spatial resolution of gene expression patterns using histological information but also enhances our understanding of the subtle differences in gene expression across different regions of the brain.

stMMR deciphers evolving cell lineage structures in chicken heart ST dataset

Analyzing temporal SRT data can reveal the dynamic domain changes during the development of tissue organs. This is crucial for our understanding of complex biological processes such as tissue development and disease progression. We collected the chicken heart SRT dataset to further investigate the effectiveness of stMMR in the integrated representation of multi-modal features (52). This dataset includes 12 tissue slices, collected on day 4 (5 slices), day 7 (4 slices), day 10 (2 slices), and day 14 (1 slices), documenting four key stages of the Hamburger-Hamilton ventricular developmental stages (52).

We annotated the slices of different developmental stages using labels provided by the original research (Figure 4A) (52). Subsequently, we employed the embeddings from stMMR and SpaceFlow to identify domains of chicken heart across these four distinct stages. Figure 4B indicate that the regions detected by stMMR largely coincide with manual annotation. For instance, in all tissue sections, major regions of the chicken heart, such as atrial cells and the inter-ventricular septum, are accurately identified. This discovery is significant for a deeper understanding of the spatial structure of cardiac tissues. Notably, stMMR also detects domains that are hard to identified (Figure 4B). For example, in the data from days 7, 10, and 14, the epicardium, a thin layer surrounding the outer side of the chicken heart, is clearly identified by stMMR. Although there are some instances of misclassification in the characterization of spot features using stMMR in a few regions, the identification of the epicardium is quite clear (Figure 4B).

Next, we adopted a method similar to previous study to analyze the pseudo-spatiotemporal map (pSM) (23). In brief, we mapped the spot features obtained through stMMR and SpaceFlow on the pseudo-temporal axis (23, 39, 53). These points reflect the relative positions of cells in their developmental trajectory or functional state. As clearly visible in Figure 4C, within the D7 to D14, the valve structures can be distinctly identified through the pSM values. Moreover, the representation of the myocardium in ventricles, as indicated by the pSM values, appears more uniform (yellow area) compared to the regional segmentation results in Figure B. According to related research (54), the endocardium, the inner layer of the heart, is one of the early events in cardiac formation. The endocardial tubes are fundamental to cardiac development, eventually merging to form the primitive heart tube. As the heart tube forms, myocardial development commences, followed closely by the development of the atria. In our analysis, we observe that the myocardium in ventricles (yellow area in Figure 4C) consistently shows higher pSM values compared to other areas in the same stage, indicating a later pseudo-temporal ordering of the ventricular myocardium (23). Additionally, the pseudo-temporal ordering of the atria (marked in teal) follows that of

the valves, suggesting that the development of the atria occurs after the valves. Therefore, the pSM derived from stMMR accurately displays the developmental sequence of the chicken heart. We further identified domain-specific genes through differential expression analysis across regions. For instance, we observed that MYH7 is highly specifically expressed in the Atria. This finding aligns with previous reports on the analysis of Atria and Ventricles specific proteins (55).

In addition, we also employed various algorithms to perform a comparative analysis of the pSM on the human DLPFC dataset. As shown in Supplementary Figure S2, we observe that in the pSM analysis, methods like GraphST, DeepST, stLearn, and SCANPY fail to clearly reflect the layered spatial organization of the tissue. In contrast, the stMMR, STAGATE, SpaceFlow and CCST are able to distinctly reveal the layered pattern of the pSM, displaying clear and smooth color gradients. This result from stMMR not only mirrors the correct internal and external developmental sequence of the cortical layers but also demonstrates the layered spatial organization of tissue, aligning with the findings of previous studies (23).

stMMR accurate identifies tumor region in human breast cancer

Breast cancer is a major type of cancer worldwide (56). We collected human breast cancer SRT data from the 10x Visium platform to conduct an in-depth study of the microenvironment in breast cancer. This dataset includes 3,798 spatial spots and 36,601 genes. Experienced pathologists have annotated these SRT data using H&E images and signature genes of breast cancer, categorizing them into 20 distinct regions (Figure 5A).

First, we applied different methods for domain identification. From the results presented in Figure 5B, it is observed that stMMR shows the most outstanding performance in category labeling. In terms of regional continuity, stMMR also demonstrates superior performance among different methods. Taking the IDC_5 area in the upper left corner as an example, this area occupies a significant portion in

invasive ductal carcinoma, with a notable increase in cancer cells compared to normal tissue or non-tumorous areas. Previous studies have also indicated that cells originating from solid tumors are primarily concentrated in the IDC area (22). However, only stMMR accurately identified the entire IDC_5 area, demonstrating higher precision compared to other methods. Additionally, stMMR also exhibits higher continuity in predicting the Tumor_edge area, whereas the results of other methods appear more dispersed in this aspect.

Next, we conducted a comprehensive analysis of domain-specific genes between merged tumor and normal regions (Supplementary Section 2.3). We utilized the DisGeNET to delve into the differentially expressed genes in the breast cancer tumor regions (57). Our analysis revealed that these domain-specific genes are enriched in several breast cancer related terms (Figure 5C left panel). For instance, C0024305 is a term related to non-Hodgkin lymphoma. Studies have shown that the development of breast cancer significantly increases the risk of non-Hodgkin lymphoma, particularly follicular lymphoma and mature T/NK cell lymphomas (58). This risk is notably more pronounced in patients undergoing hormone therapy and in younger patients (58). Importantly, genes were also enriched in an inflammation-related term (Figure 5C left panel). Numerous studies have indicated that inflammation plays a regulatory role in the development of cancer and its response to treatment (59–61). To further validate our research findings, we conducted a systematic analysis of the transcriptional regulatory network using TRRUST (62). The analysis results indicated that multiple top-ranked terms are closely associated with breast cancer (Figure 5C right panel). For instance, TRR01419 emerges as the fourth most significant term, with TP53 is identified as its key regulatory factor. TP53 plays a crucial role in both cancer-related systemic inflammation and the progression of cancer (63). Additionally, the key regulatory factors for the top three terms - TRR01256 (regulated by SP1), TRR00875 (regulated by NFKB1), and TRR01158 (regulated by RELA) - have also been confirmed in previous studies to play pivotal roles in the development and progression of breast cancer (64–68).

stMMR dissects cell type differences in a lung cancer SRT dataset based on NanoString technology

To further validate the generalization ability and applicability of stMMR, we tested its effectiveness using the single-cell SRT dataset generated by NanoString CosMx SMI. This dataset comprises lung cancer tissue samples from 20 fields of view (FOVs) (26), involving 982 genes and 83,621 cells, covering eight major cell types, including lymphocytes, neutrophils, mast cells, endothelial cells, fibroblasts, epithelial cells, myeloid cells, and tumor cells. For ease of observation, we displayed the ground truth of one of the FOVs in Figures 6A.

We employed the benchmarking methods to conduct a detailed analysis of the spatial organization within 20 FOVs, as shown in Figures 6B, D-F. Figure 6B revealed that stMMR closely aligns with the original study in detecting the spatial distribution of cell types. Particularly in identifying tumor cells, stMMR demonstrates high precision, accurately detecting tumor cells distributed across different locations in the tissue sections (Figure 6B). In the overall analysis of the 20 sections, the performance of stMMR is superior to other methods (Figure 6D). Furthermore, we conducted a cell type-specific gene analysis based on the cell annotations in one slice. We observe that different tissue cells exhibit unique expression patterns (Figure 6C). For instance, Igkc transcripts, previously reported to be upregulated in myeloid progenitor populations, is also confirmed in our study (69). The genes COL3A1 and COL1A1 shows significant positive correlations with neutrophils (70, 71). Additionally, the oncogene SOX4 is prominently featured in our differential analysis of tumor cells (72). These genes are also identified as diagnostic or prognostic biomarkers in previous studies (70, 73–76). Notably, some cell types also share similar gene expression patterns (Figure 6C). For example, epithelial cells and tumor cells in this lung cancer dataset exhibit expression similarities. Multiple studies using single-cell transcriptomics analysis have revealed that lung cancer cells share characteristics similar to those of Type 1 (AT1) and Type 2 (AT2) alveolar epithelial cells (77, 78). This similarity may be related to lung cancer

cells maintaining epithelial cell functions, such as cell adhesion and migration (79, 80).

We also conducted a visualization analysis comparing the results of stMMR applied to 20 tissue sections with the actual division of tissue regions. The analysis demonstrates that stMMR effectively identifies tissue regions across multiple sections (Figure 6E). Notably, even in regions bisected by section boundaries, stMMR maintains smooth and continuous (Figure 6E and F). These findings indicate that the joint representation of stMMR not only effectively eliminates noise from different data types but also maintains excellent performance in the recognition of tissue regions across multiple slices.

Discussion

SRT technology enables us to deeply understand the spatial structure of tissues within biological systems from multiple dimensions, including gene expression profiles, spatial positioning, and histological information. Through comprehensive integration of these modalities, we can obtain an informative joint representation. However, the inherent data heterogeneity along with the varying spatial resolutions presents challenges in the integration of these modalities. To overcome this problem, we propose a novel computational framework, stMMR. This framework aims to harmonize and unify multi-modal data as well as achieve effective joint representation for multi-modal SRT data.

stMMR effectively unifies gene expression profiles and histological information by utilizing spatial location as a connecting link. This method automates the construction of adjacency relationships between neighboring spots. Then, GCN is employed to extract features from both gene expression profiles and histological information. Furthermore, stMMR adopts an innovative strategy for representing intra-modal and inter-modal features. Initially, it employs an attention mechanism for an in-depth learning within a single modality. It then integrates cross-modality features through a combination of similarity contrastive learning, along with the reconstruction of gene expression and adjacency relationship. By applying stMMR to SRT data of

various tissues and resolutions, we have validated its exceptional performance in multiple analyses, including domain identification, pseudo-spatiotemporal analysis, gene expression data enhancement as well as the identification of domain-specific genes.

The remarkable performance of stMMR can be attributed to several innovative designs. The most crucial aspect is the integration of histological information with gene expression data through spatial location. In SRT, gene expression data suffers from issues of sparsity and zero inflation, which are key factors that interfere with downstream analysis (33, 81). Previous research has shown that histological information can predict gene expression data (30–32). Therefore, compared to methods that rely on gene expression information solely, stMMR integrates imaging information and exhibits superior performance in spot characterization. Secondly, unlike other methods that construct spatial transcriptomic data as undirected, unweighted graphs, stMMR builds undirected weighted graphs inversely proportional to Euclidean distances between spots, better reflecting the influence of spatial distance on message passing and aggregation. Furthermore, the consideration of relationships within and between modalities is also crucial. Sole reliance on gene expression data for correlation analysis may result in information loss. In contrast, methods that incorporate imaging information, such as DeepST, focus primarily on the integration of multi-modal data, overlooking the relationships within individual modalities. To fully leverage the relationships within and between modalities, stMMR not only uses similarity contrastive learning for integrating features across modalities but also incorporates a self-attention module for deep embedding of features within a modality. Additionally, the reconstruction modules for gene expression and adjacency matrix further encourage the model to retain as much original information as possible. This encoder-decoder structure improves the ability of stMMR to recover information also endows stMMR with denoising capabilities and robustness.

It is noteworthy that the stMMR framework, distinguished by its exceptional feature embedding capabilities, also exhibits a remarkable ability to handle data derived

from diverse experimental techniques. Beyond the previously mentioned datasets, our investigation also extended to the analysis of a mouse brain dataset derived using 10x Visium technology and a human pancreatic ductal adenocarcinoma dataset obtained through ST technology (Supplementary Figure S3 and S4). In these tests, the stMMR consistently achieved optimal results.

stMMR also has model scalability. In the design of the model, we considered that gene expression of each spot corresponds to a small area in space. Therefore, before embedding histological information, we also tailor it to match each spot aera. This means that the input information of stMMR is always associated with each spot area. Recently, the advancement of spatial multi-omics technologies has provided new data for analyzing the spatial distribution and functions of cells in tissues from multiple perspectives, such as simultaneous observations of transcriptomes and proteomes, transcriptomes and epigenomics (82–85). The stMMR framework can be easily expanded to support these types of data. The researchers only need to duplicate the gene expression module and then apply similar methods for embedding features within and between modalities. Moreover, the joint representation obtained through stMMR can also be applied to other tasks, such as cell type deconvolution (86–89). This application requires a process similar to methods like scaden (90), where a neural network is connected to the joint representation. Subsequently, the generation of simulation data and model training can be conducted using annotated single-cell data.

There is still room for the improvement of stMMR. Currently, stMMR employs Euclidean distance in the construction of spot adjacency matrices. However, in practical scenarios, it may be more rational to utilize different distance metrics for graph construction based on modal features. For instance, considering gene expression data, the use of Pearson Correlation Coefficients or K-L divergence might be more appropriate to measure expression similarity between spots. In contrast, for spatial imaging data, either Euclidean distance or staining similarity can serve as the distance metric. Under these circumstances, the constructed graph transitions from being a homogenous graph to a heterogeneous one. For such heterogeneous graphs with

multiple types of edges, we can apply methods like metapath2vec or multi-view learning to achieve embedding and integration of different modalities (91–94).

In this paper, we introduce a robust and accurate tool, stMMR, for the integration of gene expression data, spatial information, and histological information from SRT data. Compared to existing methods, stMMR demonstrates a significant advantage in integrating multi-modal data, particularly excelling in domain identification, pseudo-spatiotemporal analysis, and domain-specific gene analysis. Overall, as an effective and user-friendly tool, stMMR enhances the multi-modal joint analysis of SRT data, providing substantial support for research in relevant fields.

Data availability

All datasets used in this paper are publicly available. The descriptions and download address are listed in Supplementary Section 2.1 and Supplementary Table S1. Processed datasets are also available at SODB (https://gene.ai.tencent.com/SpatialOmics/) and can be loaded by PySODB (https://protocols-pysodb.readthedocs.io/en/latest/). The source code for stMMR can be downloaded from Github (https://github.com/nayu0419/stMMR).

Supplementary data

Supplementary Data are available at NAR online.

Acknowledgements

Author contributions: R.G., W.Z. and Z.Y. conceived and supervised the project. D.Z. and N.Y. designed the model and developed the stMMR software. D.Z., N.Y. and W.Z. wrote the manuscript. X.S. and Q.Z. collected and constructed the benchmark datasets. W.L. X.L. and Z.L. conducted biological interpretation. R.G., W.Z. and Z.Y. reviewed the manuscript. All authors read and approved the final manuscript.

Funding

National Natural Science Foundation of China [Nos. U1806202, 62303271]; Natural Science Foundation of Shandong Province [ZR2023QF081]. Funding for open access charge: National Natural Science Foundation of China [Nos. U1806202, 62303271].

Conflict of interest statement

None declared.

References

1. Asp,M., Bergenstråhle,J. and Lundeberg,J. (2020) Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays*, **42**, 1900221.
2. Chen,K.H., Boettiger,A.N., Moffitt,J.R., Wang,S. and Zhuang,X. (2015) Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, **348**, aaa6090.
3. Eng,C.-H.L., Lawson,M., Zhu,Q., Dries,R., Koulena,N., Takei,Y., Yun,J., Cronin,C., Karp,C., Yuan,G.-C., *et al.* (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, **568**, 235–239.
4. Wang,X., Allen,W.E., Wright,M.A., Sylwestrak,E.L., Samusik,N., Vesuna,S., Evans,K., Liu,C., Ramakrishnan,C., Liu,J., *et al.* (2018) Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, **361**, eaat5691.
5. Ståhl,P.L., Salmén,F., Vickovic,S., Lundmark,A., Navarro,J.F., Magnusson,J., Giacomello,S., Asp,M., Westholm,J.O., Huss,M., *et al.* (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, **353**, 78–82.
6. Rodriques,S.G., Stickels,R.R., Goeva,A., Martin,C.A., Murray,E., Vanderburg,C.R., Welch,J., Chen,L.M., Chen,F. and Macosko,E.Z. (2019) Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, **363**, 1463–1467.
7. Marx,V. (2021) Method of the Year: spatially resolved transcriptomics. *Nat Methods*, **18**, 9–14.
8. Guo,T., Yuan,Z., Pan,Y., Wang,J., Chen,F., Zhang,M.Q. and Li,X. (2023) SPIRAL: integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biology*, **24**, 241.
9. Huo,Y., Guo,Y., Wang,J., Xue,H., Feng,Y., Chen,W. and Li,X. (2023) Integrating multi-modal information to detect spatial domains of spatial transcriptomics by graph attention network. *Journal of Genetics and Genomics*, **50**, 720–733.
10. Son,C.G., Bilke,S., Davis,S., Greer,B.T., Wei,J.S., Whiteford,C.C., Chen,Q.-R., Cenacchi,N. and Khan,J. (2005) Database of mRNA gene expression profiles of multiple human organs. *Genome Res*, **15**, 443–450.
11. Hannig,J., Schäfer,H., Ackermann,J., Hebel,M., Schäfer,T., Döring,C., Hartmann,S., Hansmann,M.-L. and Koch,I. (2020) Bioinformatics analysis of whole slide images reveals significant neighborhood preferences of tumor cells in Hodgkin lymphoma. *PLoS Comput Biol*, **16**, e1007516.

12. Haghighi,M., Caicedo,J.C., Cimini,B.A., Carpenter,A.E. and Singh,S. (2022) High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. *Nat Methods*, **19**, 1550–1557.
13. Li,Z., Chen,X., Zhang,X., Jiang,R. and Chen,S. (2023) Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics. *Genome Res.*, 10.1101/gr.277891.123.
14. Li,Z. and Zhou,X. (2022) BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biology*, **23**, 168.
15. Zhao,E., Stone,M.R., Ren,X., Guenthoer,J., Smythe,K.S., Pulliam,T., Williams,S.R., Uytingco,C.R., Taylor,S.E.B., Nghiem,P., *et al.* (2021) Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol*, **39**, 1375–1384.
16. Dries,R., Zhu,Q., Dong,R., Eng,C.-H.L., Li,H., Liu,K., Fu,Y., Zhao,T., Sarkar,A., Bao,F., *et al.* (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology*, **22**, 78.
17. Varrone,M., Tavernari,D., Santamaria-Martínez,A., Walsh,L.A. and Ciriello,G. (2024) CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat Genet*, **56**, 74–84.
18. Liu,W., Liao,X., Luo,Z., Yang,Y., Lau,M.C., Jiao,Y., Shi,X., Zhai,W., Ji,H., Yeong,J., *et al.* (2023) Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with PRECAST. *Nat Commun*, **14**, 296.
19. Yuan,Z. (2024) MENDER: fast and scalable tissue structure identification in spatial omics data. *Nat Commun*, **15**, 207.
20. Li,J., Chen,S., Pan,X., Yuan,Y. and Shen,H.-B. (2022) Cell clustering for spatial transcriptomics data with graph neural networks. *Nat Comput Sci*, **2**, 399–408.
21. Dong,K. and Zhang,S. (2022) Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun*, **13**, 1739.
22. Long,Y., Ang,K.S., Li,M., Chong,K.L.K., Sethi,R., Zhong,C., Xu,H., Ong,Z., Sachaphibulkij,K., Chen,A., *et al.* (2023) Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun*, **14**, 1155.
23. Ren,H., Walker,B.L., Cang,Z. and Nie,Q. (2022) Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat Commun*, **13**, 4076.
24. Pham,D., Tan,X., Xu,J., Grice,L.F., Lam,P.Y., Raghubar,A., Vukovic,J., Ruitenberg,M.J. and Nguyen,Q. (2020) stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. 10.1101/2020.05.31.125658.
25. Xu,C., Jin,X., Wei,S., Wang,P., Luo,M., Xu,Z., Yang,W., Cai,Y., Xiao,L., Lin,X., *et al.* (2022) DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, **50**, e131–e131.
26. He,S., Bhatt,R., Brown,C., Brown,E.A., Buhr,D.L., Chantranuvatana,K., Danaher,P., Dunaway,D., Garrison,R.G., Geiss,G., *et al.* (2022) High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol*, **40**, 1794–1806.
27. Dosovitskiy,A., Beyer,L., Kolesnikov,A., Weissenborn,D., Zhai,X., Unterthiner,T., Dehghani,M., Minderer,M., Heigold,G., Gelly,S., *et al.* (2021) An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 10.48550/arXiv.2010.11929.

28. Kipf,T.N. and Welling,M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. 10.48550/ARXIV.1609.02907.
29. Brauwers,G. and Frasincar,F. (2023) A General Survey on Attention Mechanisms in Deep Learning. *IEEE Trans. Knowl. Data Eng.*, **35**, 3279–3298.
30. Markey,M., Kim,J., Goldstein,Z., Gerardin,Y., Brosnan-Cashman,J., Javed,S.A., Juyal,D., Padigela,H., Yu,L., Rahsepar,B., *et al.* (2023) Abstract B010: Spatially-resolved prediction of gene expression signatures in H&E whole slide images using additive multiple instance learning models. *Molecular Cancer Therapeutics*, **22**, B010–B010.
31. Bergenstråhle,L., He,B., Bergenstråhle,J., Abalo,X., Mirzazadeh,R., Thrane,K., Ji,A.L., Andersson,A., Larsson,L., Stakenborg,N., *et al.* (2022) Super-resolved spatial transcriptomics by deep data fusion. *Nat Biotechnol*, **40**, 476–479.
32. Zeng,Y., Wei,Z., Yu,W., Yin,R., Yuan,Y., Li,B., Tang,Z., Lu,Y. and Yang,Y. (2022) Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. *Briefings in Bioinformatics*, **23**, bbac297.
33. Covert,I., Gala,R., Wang,T., Svoboda,K., Sümbül,U. and Lee,S.-I. (2023) Predictive and robust gene selection for spatial transcriptomics. *Nat Commun*, **14**, 2091.
34. Yu,Z., Lu,Y., Wang,Y., Tang,F., Wong,K.-C. and Li,X. (2022) ZINB-Based Graph Embedding Autoencoder for Single-Cell RNA-Seq Interpretations. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 4671–4679.
35. Kipf,T.N. and Welling,M. (2016) Variational Graph Auto-Encoders. 10.48550/ARXIV.1611.07308.
36. Tang,M., Yang,C. and Li,P. (2022) Graph Auto-Encoder Via Neighborhood Wasserstein Reconstruction. 10.48550/ARXIV.2202.09025.
37. Yuan,Z., Pan,W., Zhao,X., Zhao,F., Xu,Z., Li,X., Zhao,Y., Zhang,M.Q. and Yao,J. (2023) SODB facilitates comprehensive exploration of spatial omics data. *Nat Methods*, **20**, 387–399.
38. Lin,S., Zhao,F., Wu,Z., Yao,J., Zhao,Y. and Yuan,Z. (2023) Streamlining spatial omics data analysis with Pysodb. *Nat Protoc*, 10.1038/s41596-023-00925-5.
39. Wolf,F.A., Angerer,P. and Theis,F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, **19**, 15.
40. Li,B., Zhang,W., Guo,C., Xu,H., Li,L., Fang,M., Hu,Y., Zhang,X., Yao,X., Tang,M., *et al.* (2022) Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods*, **19**, 662–670.
41. Cheng,A., Hu,G. and Li,W.V. (2023) Benchmarking cell-type clustering methods for spatially resolved transcriptomics data. *Briefings in Bioinformatics*, **24**, bbac475.
42. Zhu,J., Shang,L. and Zhou,X. (2023) SRTsim: spatial pattern preserving simulations for spatially resolved transcriptomics. *Genome Biology*, **24**, 39.
43. Shang,L. and Zhou,X. (2022) Spatially aware dimension reduction for spatial transcriptomics. *Nat Commun*, **13**, 7203.
44. Maynard,K.R., Collado-Torres,L., Weber,L.M., Uytingco,C., Barry,B.K., Williams,S.R., Catallini,J.L., Tran,M.N., Besich,Z., Tippani,M., *et al.* (2021) Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci*, **24**, 425–436.
45. Wolf,F.A., Hamey,F.K., Plass,M., Solana,J., Dahlin,J.S., Göttgens,B., Rajewsky,N., Simon,L. and Theis,F.J. (2019) PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, **20**, 59.

46. Gao,J., Zhang,F., Hu,K. and Cui,X. (2022) Hexagonal Convolutional Neural Network for Spatial Transcriptomics Classification. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.pp. 200–205.
47. Avşar,G. and Pir,P. (2023) A comparative performance evaluation of imputation methods in spatially resolved transcriptomics data. *Mol. Omics*, **19**, 162–173.
48. Lopez,R., Li,B., Keren-Shaul,H., Boyeau,P., Kedmi,M., Pilzer,D., Jelinski,A., Yofe,I., David,E., Wagner,A., *et al.* (2022) DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol*, **40**, 1360–1369.
49. Wang,Y., Song,B., Wang,S., Chen,M., Xie,Y., Xiao,G., Wang,L. and Wang,T. (2022) Sprod for de-noising spatially resolved transcriptomics data based on position and image information. *Nat Methods*, **19**, 950–958.
50. Arion,D., Enwright,J.F., Gonzalez-Burgos,G. and Lewis,D.A. (2022) Differential gene expression between callosal and ipsilateral projection neurons in the monkey dorsolateral prefrontal and posterior parietal cortices. *Cereb Cortex*, **33**, 1581–1594.
51. Arnsten,A.F.T., Woo,E., Yang,S., Wang,M. and Datta,D. (2022) Unusual molecular regulation of dorsolateral prefrontal cortex layer III synapses increases vulnerability to genetic and environmental insults in schizophrenia. *Biol Psychiatry*, **92**, 480–490.
52. Mantri,M., Scuderi,G.J., Abedini-Nassab,R., Wang,M.F.Z., McKellar,D., Shi,H., Grodner,B., Butcher,J.T. and De Vlaminck,I. (2021) Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. *Nat Commun*, **12**, 1771.
53. Haghverdi,L., Büttner,M., Wolf,F.A., Buettner,F. and Theis,F.J. (2016) Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods*, **13**, 845–848.
54. Martinsen,B.J. (2005) Reference guide to the stages of chick heart embryology. *Developmental Dynamics*, **233**, 1217–1237.
55. Lu,Z.Q., Sinha,A., Sharma,P., Kislinger,T. and Gramolini,A.O. (2014) Proteomic Analysis of Human Fetal Atria and Ventricle. *J. Proteome Res.*, **13**, 5869–5878.
56. Siegel,R.L., Miller,K.D., Fuchs,H.E. and Jemal,A. (2022) Cancer statistics, 2022. *CA Cancer J Clin*, **72**, 7–33.
57. Piñero,J., Queralt-Rosinach,N., Bravo,À., Deu-Pons,J., Bauer-Mehren,A., Baron,M., Sanz,F. and Furlong,L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)*, **2015**, bav028.
58. Kang,D., Yoon,S.E., Shin,D., Lee,J., Hong,Y.S., Lee,S.K., Lee,J.E., Park,Y.H., Ahn,J.S., Guallar,E., *et al.* (2021) Risk of non-Hodgkin lymphoma in breast cancer survivors: a nationwide cohort study. *Blood Cancer J.*, **11**, 1–8.
59. Berger,E., Delpierre,C., Hosnijeh,F.S., Kelly-Irving,M., Portengen,L., Bergdahl,I.A., Johansson,A.-S., Krogh,V., Palli,D., Panico,S., *et al.* (2018) Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts. *Sci Rep*, **8**, 10805.
60. Zhao,H., Wu,L., Yan,G., Chen,Y., Zhou,M., Wu,Y. and Li,Y. (2021) Inflammation and tumor progression: signaling pathways and targeted intervention. *Sig Transduct Target Ther*, **6**, 1–46.
61. McAndrew,N.P., Bottalico,L., Mesaros,C., Blair,I.A., Tsao,P.Y., Rosado,J.M., Ganguly,T., Song,S.J., Gimotty,P.A., Mao,J.J., *et al.* (2021) Effects of systemic inflammation on relapse in early breast cancer. *npj Breast Cancer*, **7**, 1–10.

62. Han,H., Shim,H., Shin,D., Shim,J.E., Ko,Y., Shin,J., Kim,H., Cho,A., Kim,E., Lee,T., *et al.* (2015) TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep*, **5**, 11432.
63. Wellenstein,M.D., Coffelt,S.B., Duits,D.E.M., van Miltenburg,M.H., Slagter,M., de Rink,I., Henneman,L., Kas,S.M., Prekovic,S., Hau,C.-S., *et al.* (2019) Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. *Nature*, **572**, 538–542.
64. Gao,Y., Gan,K., Liu,K., Xu,B. and Chen,M. (2021) SP1 Expression and the Clinicopathological Features of Tumors: A Meta-Analysis and Bioinformatics Analysis. *Pathol Oncol Res*, **27**, 581998.
65. Wang,W., Nag,S.A. and Zhang,R. (2015) Targeting the NFκB Signaling Pathways for Breast Cancer Prevention and Therapy. *Curr Med Chem*, **22**, 264–289.
66. Kanzaki,H., Chatterjee,A., Hossein Nejad Ariani,H., Zhang,X., Chung,S., Deng,N., Ramanujan,V.K., Cui,X., Greene,M.I. and Murali,R. (2021) Disabling the Nuclear Translocalization of RelA/NF-κB by a Small Molecule Inhibits Triple-Negative Breast Cancer Growth. *Breast Cancer (Dove Med Press)*, **13**, 419–430.
67. Kim,G.-C., Kwon,H.-K., Lee,C.-G., Verma,R., Rudra,D., Kim,T., Kang,K., Nam,J.H., Kim,Y. and Im,S.-H. (2018) Upregulation of Ets1 expression by NFATc2 and NFKB1/RELA promotes breast cancer cell invasiveness. *Oncogenesis*, **7**, 1–15.
68. Jeong,Y.J., Oh,H.K. and Choi,H.R. (2019) Methylation of the RELA Gene is Associated with Expression of NF-κB1 in Response to TNF-α in Breast Cancer. *Molecules*, **24**, 2834.
69. Mincarelli,L., Uzun,V., Wright,D., Scoones,A., Rushworth,S.A., Haerty,W. and Macaulay,I.C. (2023) Single-cell gene and isoform expression analysis reveals signatures of ageing in haematopoietic stem and progenitor cells. *Commun Biol*, **6**, 1–11.
70. Zhang,H., Ding,C., Li,Y., Xing,C., Wang,S., Yu,Z., Chen,L., Li,P. and Dai,M. Data mining-based study of collagen type III alpha 1 (COL3A1) prognostic value and immune exploration in pan-cancer. *Bioengineered*, **12**, 3634–3646.
71. Ren,J., Da,J. and Hu,N. (2022) Identification of COL1A1 associated with immune infiltration in brain lower grade glioma. *PLoS One*, **17**, e0269533.
72. Moreno,C.S. (2020) SOX4: The Unappreciated Oncogene. *Semin Cancer Biol*, **67**, 57–64.
73. Tang,M., Liu,P., Wu,X., Gong,J., Weng,J., Gao,G., Liu,Y. and Gan,L. (2021) COL3A1 and Its Related Molecules as Potential Biomarkers in the Development of Human Ewing's Sarcoma. *Biomed Res Int*, **2021**, 7453500.
74. Geng,Q., Shen,Z., Li,L. and Zhao,J. (2021) COL1A1 is a prognostic biomarker and correlated with immune infiltrates in lung cancer. *PeerJ*, **9**, e11145.
75. Walter,R.F.H., Mairinger,F.D., Werner,R., Ting,S., Vollbrecht,C., Theegarten,D., Christoph,D.C., Zarogoulidis,K., Schmid,K.W., Zarogoulidis,P., *et al.* (2015) SOX4, SOX11 and PAX6 mRNA expression was identified as a (prognostic) marker for the aggressiveness of neuroendocrine tumors of the lung by using next-generation expression analysis (NanoString). *Future Oncol*, **11**, 1027–1036.
76. Srivastava,M., Khurana,P. and Sugadev,R. (2012) Lung Cancer Signature Biomarkers: tissue specific semantic similarity based clustering of Digital Differential Display (DDD) data. *BMC Research Notes*, **5**, 617.
77. Wang,Z., Li,Z., Zhou,K., Wang,C., Jiang,L., Zhang,L., Yang,Y., Luo,W., Qiao,W., Wang,G., *et al.* (2021) Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nat Commun*, **12**, 6500.

78. Zuo,W., Rostami,M.R., Shenoy,S.A., LeBlanc,M.G., Salit,J., Strulovici-Barel,Y., O'Beirne,S.L., Kaner,R.J., Leopold,P.L., Mezey,J.G., *et al.* (2020) Cell-specific expression of lung disease risk-related genes in the human small airway epithelium. *Respiratory Research*, **21**, 200.
79. Janiszewska,M., Primi,M.C. and Izard,T. (2020) Cell adhesion in cancer: Beyond the migration of single cells. *J Biol Chem*, **295**, 2495–2505.
80. Millar,F.R., Janes,S.M. and Giangreco,A. (2017) Epithelial cell migration as a potential therapeutic target in early lung cancer. *European Respiratory Review*, **26**.
81. Liu,Z., Wu,D., Zhai,W. and Ma,L. (2023) SONAR enables cell type deconvolution with spatially weighted Poisson-Gamma model for spatial transcriptomics. *Nat Commun*, **14**, 4727.
82. Deng,Y., Bartosovic,M., Kukanja,P., Zhang,D., Liu,Y., Su,G., Enninful,A., Bai,Z., Castelo-Branco,G. and Fan,R. (2022) Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science*, **375**, 681–686.
83. Zhang,D., Deng,Y., Kukanja,P., Agirre,E., Bartosovic,M., Dong,M., Ma,C., Ma,S., Su,G., Bao,S., *et al.* (2023) Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature*, **616**, 113–122.
84. Xie,Y., Ruan,F., Li,Y., Luo,M., Zhang,C., Chen,Z., Xie,Z., Weng,Z., Chen,W., Chen,W., *et al.* (2023) Spatial chromatin accessibility sequencing resolves high-order spatial interactions of epigenomic markers. *eLife*, **12**.
85. Deng,Y., Bartosovic,M., Ma,S., Zhang,D., Kukanja,P., Xiao,Y., Su,G., Liu,Y., Qin,X., Rosoklija,G.B., *et al.* (2022) Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*, **609**, 375–383.
86. Zhang,W., Xu,H., Qiao,R., Zhong,B., Zhang,X., Gu,J., Zhang,X., Wei,L. and Wang,X. (2022) ARIC: accurate and robust inference of cell type proportions from bulk gene expression or DNA methylation data. *Briefings in Bioinformatics*, **23**, bbab362.
87. Sun,D., Liu,Z., Li,T., Wu,Q. and Wang,C. (2022) STRIDE: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Research*, **50**, e42–e42.
88. Song,Q. and Su,J. (2021) DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Briefings in Bioinformatics*, **22**, bbaa414.
89. Yan,L. and Sun,X. (2023) Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics*, **39**, btac805.
90. Menden,K., Marouf,M., Oller,S., Dalmia,A., Magruder,D.S., Kloiber,K., Heutink,P. and Bonn,S. (2020) Deep learning–based cell composition analysis from tissue expression profiles. *Sci. Adv.*, **6**, eaba2619.
91. Dong,Y., Chawla,N.V. and Swami,A. (2017) metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Halifax NS Canada, pp. 135–144.
92. Li,X., Chen,W., Chen,Y., Zhang,X., Gu,J. and Zhang,M.Q. (2017) Network embedding-based representation learning for single cell RNA-seq data. *Nucleic Acids Research*, **45**, e166–e166.
93. Guo,T., Chen,Y., Shi,M., Li,X. and Zhang,M.Q. (2022) Integration of single cell data by disentangled representation learning. *Nucleic Acids Research*, **50**, e8–e8.
94. Wu,G., Li,X., Guo,W., Wei,Z., Hu,T., Shan,Y. and Gu,J. (2022) JEBIN: analyzing gene co-expressions across multiple datasets by joint network embedding. *Briefings in Bioinformatics*, **23**, bbab603.

Figure captions:

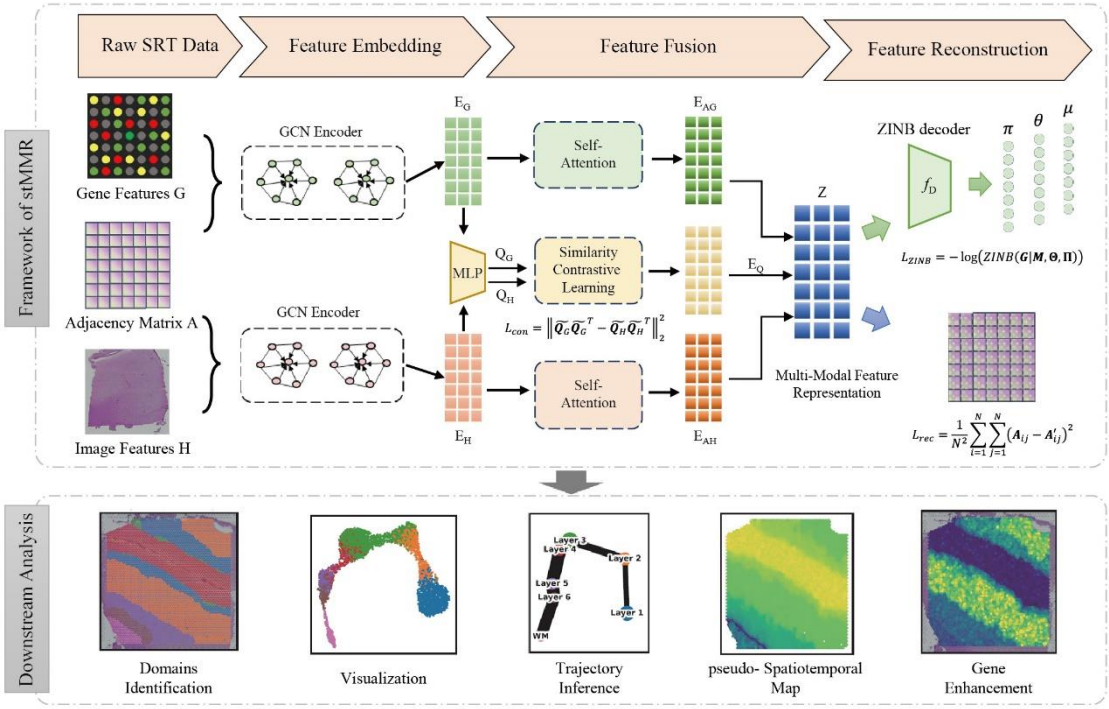

Figure 1. Schematic overview of stMMR for the joint representation of features from different modalities. Gene expression and histology image information are embedded using GCN module based on adjacent matrix. Then, the relationships between different modalities are captured through similarity contrast learning, followed by feature fusion. Finally, the original features are reconstructed from the multi-modal feature representation. This representation can be used for downstream analysis directly.

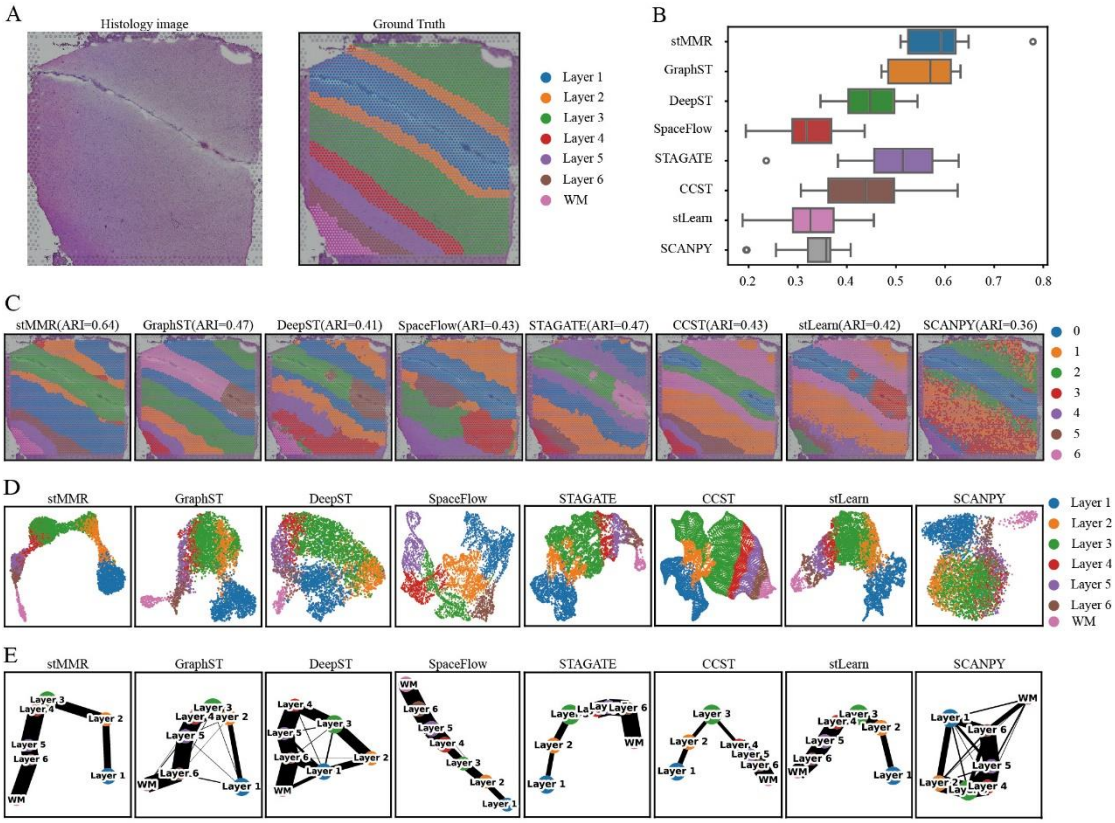

Figure 2. Performance comparisons of different methods on DLPFC datasets. (A) The histology image and manually annotated brain regions of slice 151509. (B) The overall performance of 8 different methods across 12 slices. (C) The domain recognition results on slice 151509. (D) The UMAP visualization results of the embeddings from 8 different methods on slice 151509. (E) The inferenced trajectories on slice 151509.

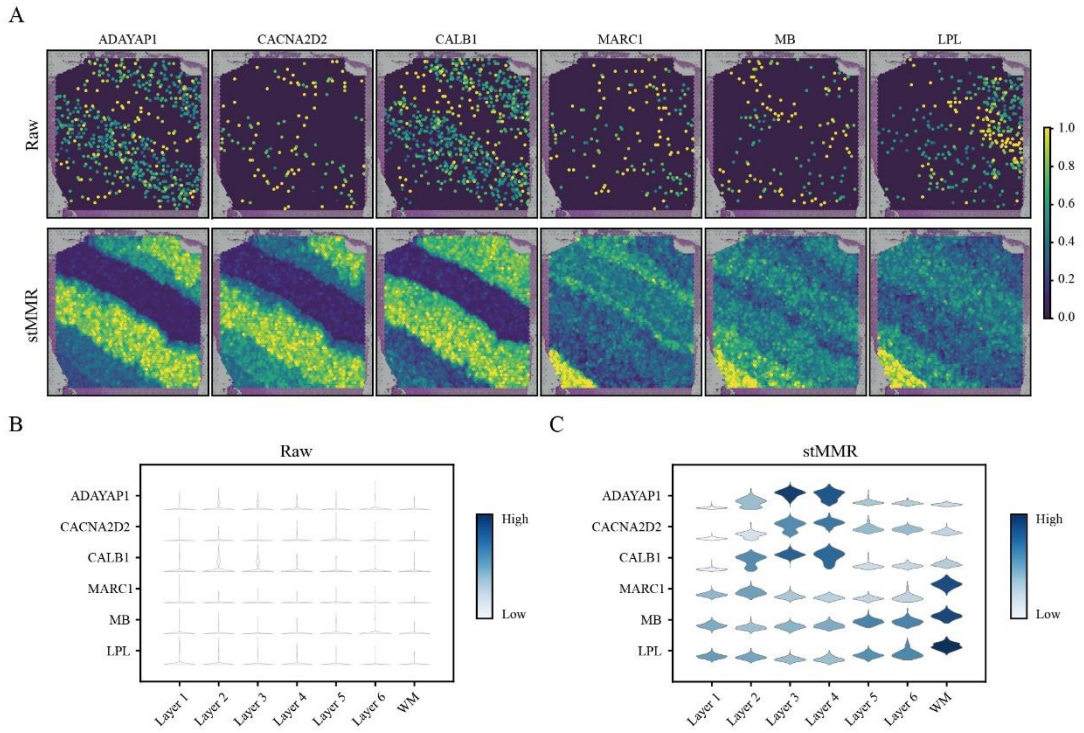

Figure 3. stMMR enhances spatial gene expression profiles and spatial structural characterization. (A) Spatial representation of layer-specific marker genes before and after data enhancement. (B) Gene expression level before and after data enhancement.

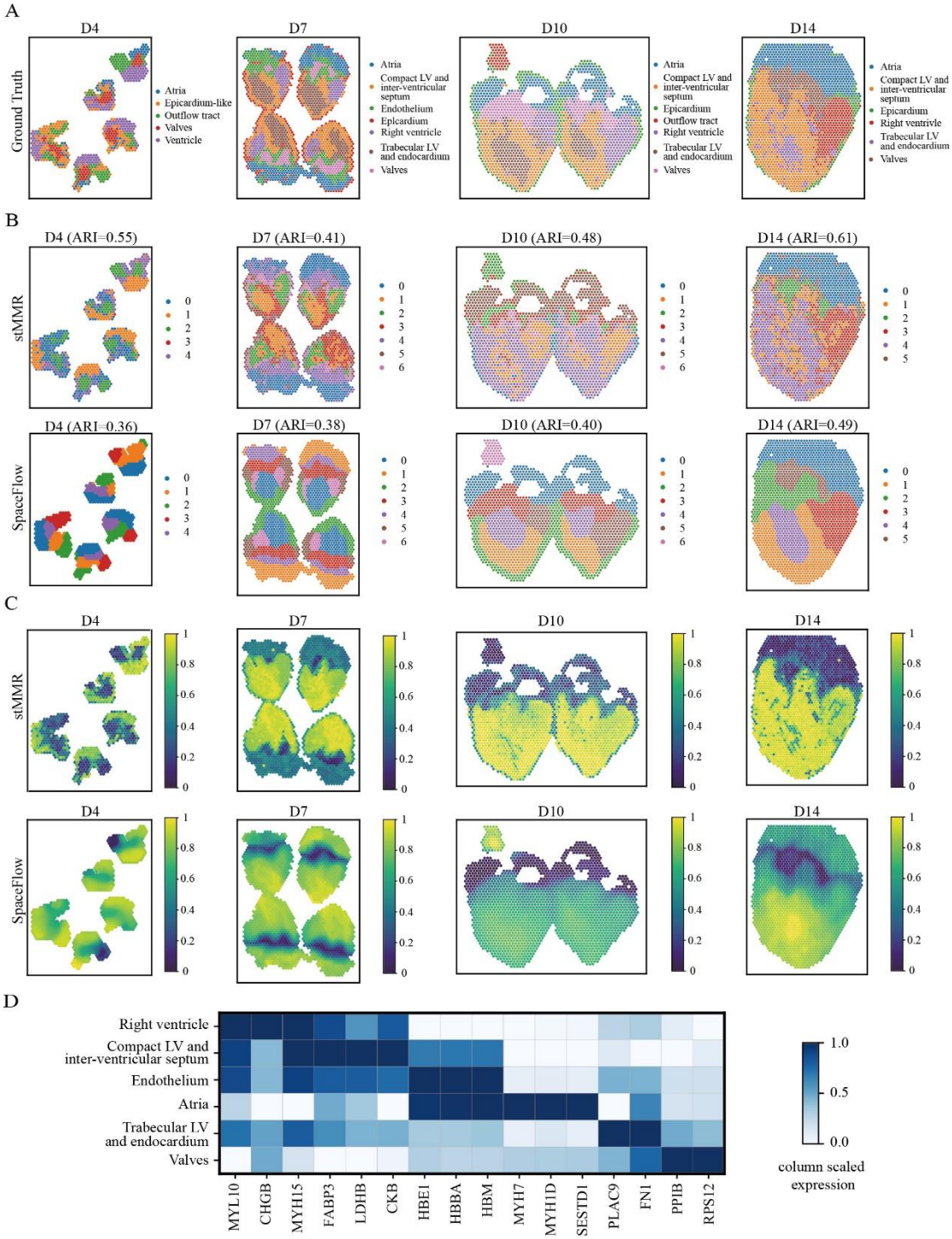

Figure 4. stMMR reveals cell lineage structures during chicken heart development. (A) The ground truth label provided by the original data. (B) The domains recognized by stMMR and SpaceFlow. (C) The plots of pSM value from stMMR and SpaceFlow for illustrating pseudo-temporal developmental trajectory. (D) The differentially expressed marker genes discovered by stMMR.

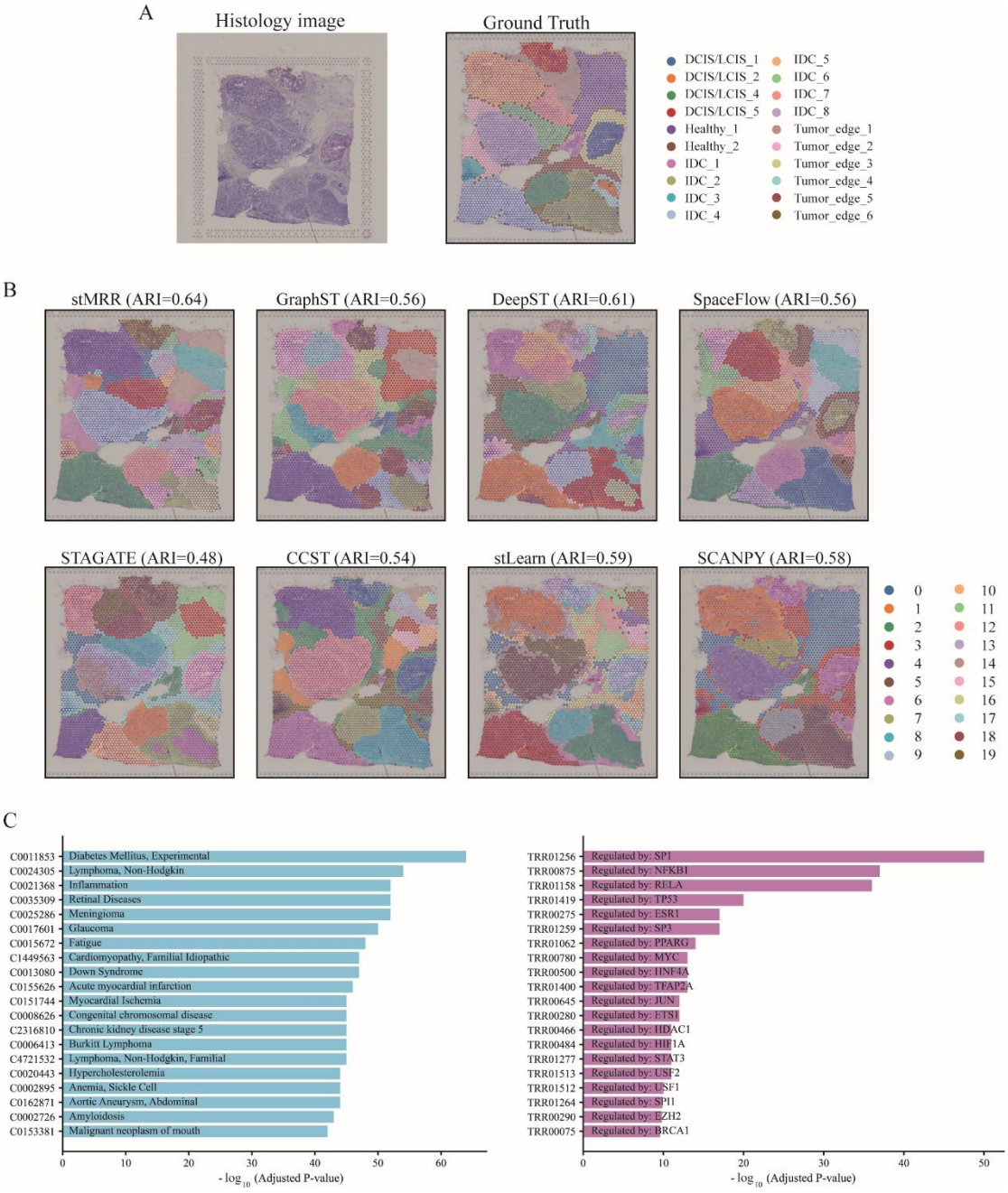

Figure 5. stMMR identifies tumor region in human breast cancer dataset. (A) The H&E images and the manually annotated regions. (B) The annotation results from different methods. (C) Top 20 differentially expressed gene enriched terms identified by DisGeNET (left panel) and TRRUST (right panel).

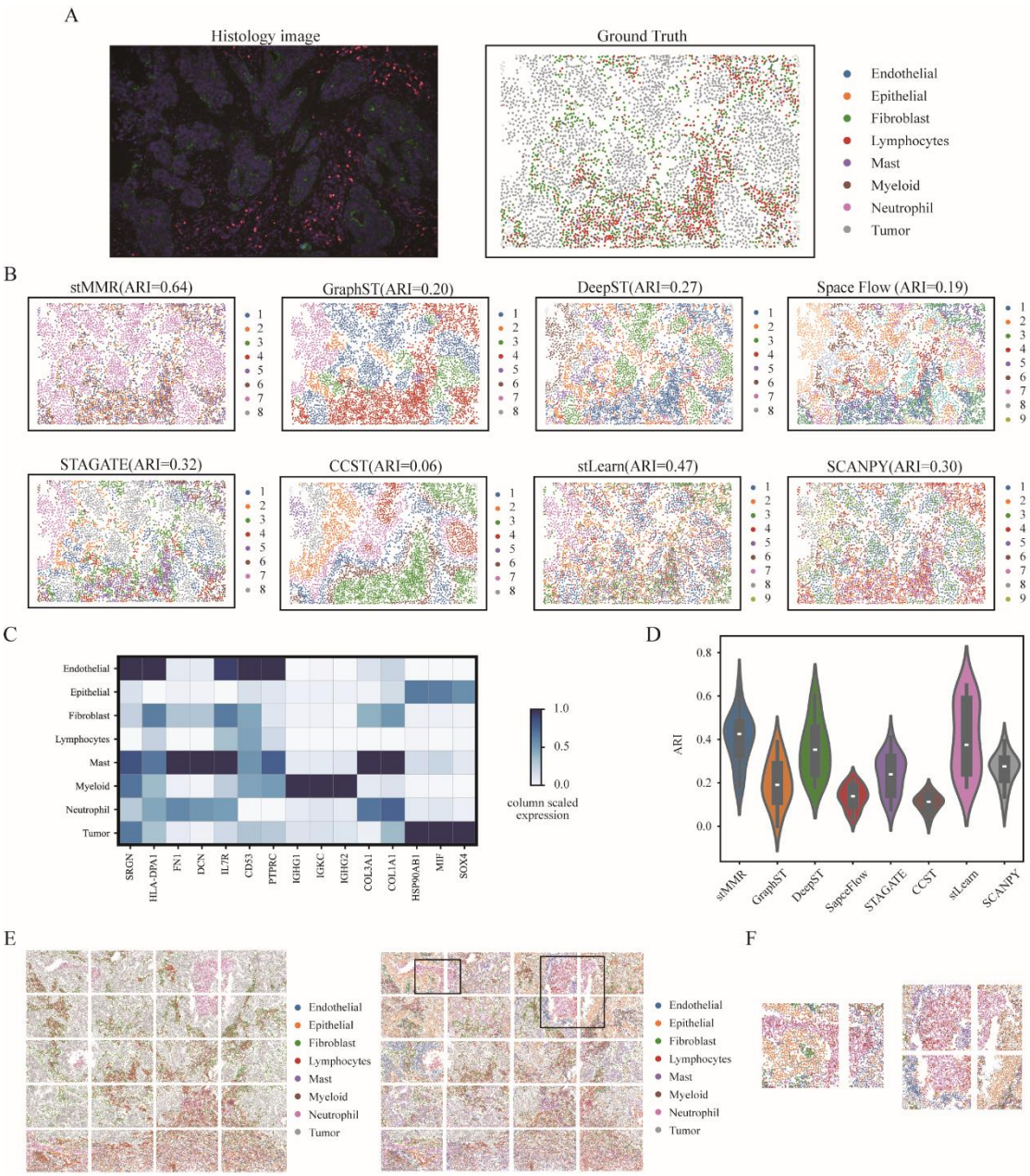

Figure 6. stMMR recognizes cell type differences in lung cancer dataset. (A) One FOV of the lung cancer SRT data. (B) Cell types identified by different methods. (C) Expression pattern of marker genes for different cell types. (D) The overall performance of different methods across 20 FOVs. (E) Cell types annotated manually in 20 FOVs. (F) Cell types annotated by stMMR in 20 FOVs. (G) The zoomed-in results of boundaries between adjacent FOVs identified by stMMR.