Efficient integration of multiple spatial transcriptomics data for 3D domain detection, matching, and alignment with stMSA

Han Shu[a,b,2], Jing Chen[c,2], Jialu Hu[a,b], Yongtian Wang[a,b], Jiajie Peng[a,b], Xuequn Shang[a,b,1], Tao Wang[a,b,1]

[a]*School of Computer Science, Northwestern Polytechnical University, 1 Dongxiang Rd., Xi'an, 710072, Shaanxi, China*
[b]*Key Laboratory of Big Data Storage and Management, Ministry of Industry and Information Technology, Northwestern Polytechnical University, 1 Dongxiang Rd., Xi'an, 710072, Shaanxi, China*
[c]*School of Computer Science and Engineering, Xi'an University of Technology, No.5 South Jinhua rd., Xi'an, 710048, Shaanxi, China*

Abstract

Spatial transcriptomics (ST) is a powerful methodology that enables the study of genes within tissue architecture by providing gene expression information along with spatial location data. With the increasing availability of ST datasets, researchers are now inclined to explore potential biological features across larger datasets simultaneously, aiming for a more comprehensive understanding. However, existing methods primarily focus on cross-batch feature learning, often overlooking the intricate spatial patterns within individual slices. This limitation poses a significant challenge in effectively integrating features across different slices while considering slice-specific patterns. To address this challenge and enhance the integration performance of multi-slice data, we propose stMSA, a deep graph contrastive-learning model that incorporates graph auto-encoder techniques. stMSA is specifically designed to generate batch-corrected representations while preserving the unique spatial patterns within each slice, simultaneously considering both inner-batch and cross-batch patterns during the integration process. Our extensive evaluations demonstrate that stMSA outperforms state-of-the-art

[1]Corresponding author. T.W.: twang@nwpu.edu.cn or X.S.: Shang@nwpu.edu.cn
[2]These authors contributed equally to this work.

methods in discerning tissue structures across different slices, even when confronted with diverse experimental protocols and sequencing technologies. Furthermore, stMSA exhibits remarkable performance in cross-slice matching and alignment for three-dimensional reconstruction. The source code for stMSA is available at `https://github.com/hannshu/stMSA`.

Keywords:
Spatial transcriptomics, 3D reconstruction, domain detection, cross-slice alignment, contrastive learning

---

1. Introduction

Spatial transcriptomics (ST) technologies have greatly advanced our understanding of gene expression within the context of tissue architecture. These technologies enable the simultaneous analysis of thousands of genes, providing insights into their spatial distribution within tissue samples [1]. Over the years, several ST methods, including imaging-based and high-throughput sequencing-based technologies, have been developed and continue to evolve, offering higher cell-capturing resolution and broader genomics coverage. For instance, early high-throughput ST technologies like Spatial Transcriptomics [2] and 10x Visium [3] measured the transcriptome of multiple mixed cells per capture site or spot, whereas recent advancements like Slide-seq V2 [4] and Stereo-seq [5] achieve even finer levels of resolution, measuring the transcriptomes of single cells or even sub-cellular structures per spot. Such advancements have paved the way for novel applications in understanding cellular interactions, identifying spatial biomarkers, and unraveling complex developmental or disease processes [6].

In parallel with the development of ST technologies, numerous computational methods have been proposed to analyze individual ST slices. One fundamental step in these methods is the identification of spatial domains, which involves clustering spots with similar gene expression profiles and spatial locations in an unsupervised manner. To achieve this, accurate integration of gene expression and spatial information is crucial. For instance, STAGATE [7] leverages self-attention mechanisms and cell type-aware spatial graphs to learn latent representations of spots, while CCST [8] adopts a Deep Graph Infomax (DGI) model, offering a general framework for multiple ST analysis tasks, including spatial domain detection and cell cycle identification. Despite the success of these methods in analyzing single ST slices,

they are limited to a two-dimensional (2D) analysis due to their single-slice focus.

With the increasing availability of ST data derived from the same tissues, organs, or embryos, it has become necessary to integrate multiple slices in three dimensions (3D) to investigate more realistic spatial structures and cellular functions [9]. Some methods, such as Scanorama [10] and Harmony [11], originally designed for single-cell data integration, have been employed to integrate ST data, albeit with limited performance due to the lack of utilization of spatial information [12]. SEDR [13] addresses batch effects across multiple ST datasets by learning latent representations for each slice using deep auto-encoders and performing batch correction using the Harmony method. While SEDR considers spatial information within individual slices, it overlooks cross-slice relationships when learning uniform representations. Recently, several methods specifically designed for integrating multiple ST slices have been proposed. STAligner [14] optimizes model parameters and integrates data from different slices using mutual nearest neighbors (MNNs), demonstrating superior batch-correction performance. Additionally, STAligner achieves 3D reconstruction by aligning the stacked slices to a uniform orientation using the Iterative Closest Point (ICP) algorithm. STitch3D [15] constructs a 3D spatial graph and employs a graph attention network to learn latent representations. It optimizes model parameters by aligning gene expression distribution patterns with reference single-cell data and mitigates specific effects between slices, spots, and genes. SPACEL [16] uses predicted cell-type deconvolution results and spatial information to train a graph neural network and employs adversarial learning to integrate gene expression data across multiple slices. SLAT [17] identifies cell relationships across slices by employing a lightweight graph-convolutional network for fast cross-slice matching, enabling the exploration of embryo developmental patterns at different time points.

While these approaches have enhanced integration performance compared to non-spatial methods, they do have certain limitations. Some methods require additional data, such as reference single-cell data or cell-type proportions, which significantly impact integration performance. Moreover, existing methods excel in specific tasks like batch-effect correction, spatial domain detection, cross-slice matching, or alignment for 3D reconstruction, lacking a comprehensive solution that can address multiple tasks simultaneously.

To address these challenges, we present stMSA, a novel graph contrastive learning-based method designed for integrating ST data from multiple sources.

stMSA employs a contrastive learning strategy, optimizing two key principles to simultaneously learn the embedding of cross-batch spots. First, within each slice, the embeddings of spots in a connected micro-environment will be similar, while those of randomly sampled spots will be dissimilar. Second, the embeddings of cross-batch spots with similar gene expression profiles will be optimized to be similar, while those of randomly sampled cross-batch spots will remain dissimilar. By adhering to these principles, stMSA captures both the inner-slice and cross-slice distribution patterns of spots. Furthermore, to optimize model parameters, stMSA reconstructs gene expression profiles using graph attention-based auto-encoder techniques and fine-tunes the latent representations through deep embedding clustering (DEC). We conducted a comprehensive evaluation of stMSA's performance, verifying its proficiency in various downstream tasks, including batch-effect correction, 3D domain detection, cross-slice matching, and multi-slice alignment. Overall, stMSA provides a comprehensive and effective solution for integrating multiple ST slices, addressing the limitations of existing methods and enabling advanced analysis of spatial transcriptomics data from multiple sources.

2. Materials and Methods

2.1. Overview of stMSA

In this study, we propose stMSA, a deep graph representation learning model specifically designed for multiple ST data integration (Figure 1). stMSA aims to learn the uniform cross-batch spot embeddings, which can be applied to 3D domain detection, cross-slice matching, and alignment for 3D reconstruction. This model takes pre-processed gene expression and spatial information of multi-slices as input. For each individual slice, stMSA initially constructs a spatial graph based on the spot coordinates. Subsequently, the stMSA leverages the gene expression profiles and spatial graphs to learn the latent representation through the following strategies.

The main idea of stMSA is to use a contrastive learning strategy to integrate multi-slice data, encoding spot features from different slices into a coherent embedding space. This strategy is designed to optimize latent representations by maximizing the similarity between spatially adjacent spots within each slice and between similar spots across different slices that share gene expression patterns, while minimizing the similarity between dissimilar spot pairs in both intra-slice and cross-slice contexts. Specifically, by increasing the similarity between centroid spots and their corresponding micro-

environmental latent representations, the model can capture comprehensive local features from spatially adjacent spots. Moreover, to preserve the distinctive expression patterns of each spot, we simultaneously train the model to optimize gene expression reconstruction using an attention-based graph auto-encoder, referred to as the gene expression reconstruction (GER) task. Additionally, we fine-tune the model by dynamically optimizing the predicted spatial domains using the deep embedding clustering (DEC) method [18]. We will introduce the details of these steps in the following context.

2.2. Spatial graph construction for ST slices

We first construct a spatial graph for each slice based on the spot coordinates. Given the spot coordinates on slice t, represented as s_t, we denote the spatial information of all k slices as set $S = \{s_1, s_2, ..., s_k\}$. stMSA initially calculates the Euclidean distances between all spots within each slice. Then stMSA uses a radius threshold r to define the edges between spots. In our experimental settings, we fine-tuned the radius threshold r to achieve spatial graphs with an approximate connectivity of 6 neighbors for each spot. The adjacency matrices for spatial graphs are denoted as $\mathbf{A} = \{\mathbf{A}^{(s_1)}, \mathbf{A}^{(s_2)}, ..., \mathbf{A}^{(s_k)}\}$, where $\mathbf{A}_{ij}^{(s_t)}$ is set as 1 only if spot i and spot j within slice s_t are connected, and it is assigned 0 otherwise.

2.3. Representation learning via parameter-shared graph auto-encoder

In order to effectively couple gene expression and spatial information, stMSA uses the technique of parameter-shared graph autoencoder to learn the latent embeddings of spots. Specifically, the graph auto-encoder consists of a graph attention network (GAT)[19] and a fully connected network (FCN)[20] for both the encoder and the decoder. Furthermore, to facilitate model convergence, stMSA shares the GAT and FCN parameters between the encoder and the decoder.

2.3.1. Encoder

stMSA adopts the GAT network to aggregate spot features within a local microenvironment on a slice. In contrast to graph convolutional neural networks, GAT uses the self-attention mechanism to dynamically compute the importance of inter-spot connections during the feature aggregation process. Specifically, the GAT layer takes the gene expression profiles and the spatial graph of each slice as inputs. Suppose the gene expression profiles on slice s_m are denoted as $\mathbf{X}^{(s_m)} = \{\mathbf{x}_1^{(s_m)}, \mathbf{x}_2^{(s_m)}, ..., \mathbf{x}_{n^{(s_m)}}^{(s_m)}\}$, where $\mathbf{x}_i^{(s_m)} \in \mathbb{R}^F$ indicates

the gene expression of spot i, and $n^{(s_m)}$ denotes the number of spots in slice s_m, the formulation of the GAT layer is described as follows:

$$\mathbf{h}_i^{(s_m)} = \sigma(\sum_{j \in N_i^{(s_m)}} \alpha_{ij}^{(s_m)} \mathbf{W}^{(GAT)} \mathbf{x}_j^{(s_m)}) \tag{1}$$

where $\mathbf{h}_i^{(s_m)}$ denotes the initial latent embedding of spot i within slice s_m, $\sigma(\cdot)$ represents the aggregation function applied to neighboring spots. Specifically, the aggregation method selected in stMSA is the summation. $N_i^{(s_m)}$ denotes the set of neighbors for spot i within slice s_m. $\mathbf{W}^{(GAT)}$ denotes the trainable weight matrix, while $\alpha_{ij}^{(s_m)}$ denotes the attention coefficient between spot i and its neighbor spot j within slice s_m, computed via the following formula:

$$\alpha_{ij}^{(s_m)} = \frac{\exp\left(\text{Sigmoid}(\mathbf{a}^T \cdot [\mathbf{W}\mathbf{x}_i^{(s_m)} || \mathbf{W}^{(GAT)} \mathbf{x}_i^{(s_m)}])\right)}{\sum_{k \in N_i^{(s_m)}} \exp\left(\text{Sigmoid}(\mathbf{a}^T \cdot [\mathbf{W}\mathbf{x}_i^{(s_m)} || \mathbf{W}^{(GAT)} \mathbf{x}_k^{(s_m)}])\right)} \tag{2}$$

where $\mathbf{a}^T$ denotes the trainable attention coefficient vector responsible for weighting the closeness between the centroid spot and its neighboring spots. The Sigmoid$(\cdot)$ is employed as an activation function. Additionally, $[\cdot||\cdot]$ denotes the concatenation operation.

Subsequently, stMSA employs a fully connected network further to compress the initial latent embedding into a lower dimension as shown in the following equation:

$$\mathbf{z}_i^{(m)} = \mathbf{W}^{(FCN)} \cdot \text{ELU}(\mathbf{h}_i^{(s_m)}) \tag{3}$$

where $\mathbf{z}_i^{(s_m)}$ denotes the latent representation learned by stMSA, $\mathbf{W}^{(FCN)}$ denotes the trainable weight matrix, and ELU$(\cdot)$ denotes the Exponential Linear Unit activation function [21].

2.3.2. Decoder

The decoder in stMSA also consists of a GAT layer and an FCN layer, both designed to reconstruct the input gene expressions. To expedite convergence and enhance efficiency, the decoder layers share the parameters learned from the encoder layers. Specifically, stMSA utilizes the transposed weight matrix of the encoder as the weight matrix for the corresponding decoder

layers. Additionally, it directly incorporates the attention coefficients from the encoder's GAT layer into the decoder's GAT layer. The formula for the decoder is as follows:

$$\hat{\mathbf{h}}_i^{(s_m)} = \sigma(\sum_{j \in N_i^{(s_m)}} \alpha_{ij}^{(s_m)} \mathbf{W}^{(FCN)^T} \mathbf{z}_j^{(s_m)}) \tag{4}$$

$$\mathbf{out}_i^{(s_m)} = \mathbf{W}^{(GAT)^T} \cdot \mathrm{ELU}(\hat{\mathbf{h}}_i^{(s_m)}) \tag{5}$$

where $\hat{\mathbf{h}}_i^{(s_m)}$ represents the reconstructed representation generated by the GAT decoder, while $\mathbf{out}_i^{(s_m)}$ denotes the output of the reconstructed gene expression profiles.

2.4. Enhancing multi-slice integration via graph contrastive learning

To create a unified embedding space for spot features extracted from various slices, stMSA employs a contrastive learning approach to remove batch effects while preserving the original spatial patterns of the spots. In contrast to conventional graph contrastive learning methods that require building a corrupted graph, our approach leverages spatial graphs constructed from multiple input ST slices. It consists of two components: inner-batch contrastive learning and cross-batch contrastive learning.

2.4.1. Inner-batch contrastive learning

In the inner-batch contrastive learning stage, the model is constrained to generate embeddings that closely resemble those of adjacent spots, and separate those non-adjacent spots. To achieve this, stMSA maximizes the embedding similarity between a centroid spot i and a spot in its adjacent micro-environment (denoted as positive pairs), while minimizing its similarity with a randomly sampled spot k (denoted as negative pairs) within the same batch (i.e. slice), and this process is formulated as follows:

$$L_{\text{inner}} = \frac{1}{N} \sum_{s_m} \sum_i \max\{||\mathbf{z}_i^{(s_m)} - \mathbf{env}_i^{(s_m)}||_2 - ||\mathbf{z}_i^{(s_m)} - \mathbf{z}_k^{(s_m)}||_2 + \gamma, 0\} \tag{6}$$

where N denotes the total number of spots across all slices, $\mathbf{z}_i^{(s_m)}$ represents the embedding of the centroid spot i on slice s_m, $\mathbf{z}_k^{(s_m)}$ refers to the

embedding of the randomly sampled spot, γ is the margin parameter, typically set to 1.0 by default. The $\mathbf{env}_i^{(s_m)}$ denotes the micro-environment embedding constructed from the embeddings of neighboring spots of spot i, which is calculated as follows:

$$\mathbf{env}_i^{(s_m)} = \sum_{j \in N_i^{(s_m)}} \mathbf{z}_j^{(s_m)} \tag{7}$$

2.4.2. Cross-batch contrastive learning

In the cross-batch contrastive learning stage, stMSA aims to generate batch-corrected embedding for each spot, which also reserves the spatial gene expression pattern. To achieve this, stMSA maximizes the embedding similarity between the centroid spot i and a spot j with similar gene expression in another slice (denoted as positive pairs), while simultaneously minimizing the similarity with a randomly sampled spot k (denoted as negative pairs) across slices. This process is depicted as follows:

$$L_{\text{cross}} = \frac{1}{N} \sum_{s_m} \sum_{s_n > s_m} \sum_{i} \max\{||\mathbf{z}_i^{(s_m)} - \mathbf{z}_j^{(s_n)}||_2 - ||\mathbf{z}_i^{(s_m)} - \mathbf{z}_k^{(s_p)}||_2 + \gamma, 0\} \tag{8}$$

where $\mathbf{z}_i^{(s_m)}$, $\mathbf{z}_j^{(s_n)}$, and $\mathbf{z}_k^{(s_p)}$ represent the embeddings of centroid spot i in slice s_m, its positive pair mate j in slice s_n, and its negative pair mate k in slice s_p, respectively. The positive pairs are selected only when spot j is the most similar spot to spot i, and spot i is the most similar spot to spot j simultaneously.

The objective function of the two contrastive learning stages can be expressed as follows:

$$L_{\text{CL}} = L_{\text{inner}} + L_{\text{cross}} \tag{9}$$

Due to the large number of spots in each batch and the relatively low neighbor count that we have set, the probability of sampling nearby spots in randomly selected negative pairs is extremely low. This ensures that the model can effectively discriminate the similarities and differences among spots. Additionally, incorporating the minimization of similarity between negative pairs helps prevent a scenario in which the embeddings of all spots become excessively similar, which will affect the performance of downstream tasks.

2.5. Gene expression reconstruction via graph autoencoder

The goal of gene expression reconstruction (GER) is to generate embeddings that can capture unique gene expression patterns, preserving their inherent characteristics while avoiding excessive similarity to neighboring spots. To accomplish this, the objective is to maximize the similarity between the input gene expression and the reconstructed output. The loss function is expressed as follows:

$$L_{\mathrm{GER}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{out}_i - \mathbf{x}_i)^2 \tag{10}$$

Where N represents the total number of spots in the input data, $\mathbf{out}_i$ and $\mathbf{x}_i$ represent the reconstructed gene expression and input gene expression of spot i, respectively.

2.6. Domain distribution tuning with deep embedding clustering

The domain distribution tuning (DDT) task aims to refine the clustering performance using the deep embedding clustering (DEC) technique[18]. The primary goal is to minimize the distance between the soft clustering distribution[22] and the auxiliary target distribution, as depicted below:

$$L_{\mathrm{DDT}} = \mathrm{KL}(\mathrm{P}||\mathrm{Q}) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{11}$$

Where Q and P denote the soft clustering distribution and the auxiliary target distribution, respectively. p_{ij} and q_{ij} represent the probability that a spot i belongs to a cluster j in distribution P and Q, respectively. The soft clustering distribution and the auxiliary target distribution are computed as follows:

$$q_{ij} = \frac{(1 + ||\mathbf{z}_i - \mu_j||^2)^{-1}}{\sum_{k=1}^{K}(1 + ||\mathbf{z}_i - \mu_k||^2)^{-1}} \tag{12}$$

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{N} q_{ij}}{\sum_{k=1}^{K}(q_{ik}^2 / \sum_{i=1}^{N} q_{ik})} \tag{13}$$

Where μ_j denotes the initial cluster centroid of a cluster j, and K represents the cluster count. The initial cluster centroids and cluster count are

generated using Louvain clustering[23], which takes gene expression as input and we set the Louvain resolution parameter to 1.0 by default.

In sum, the overall loss function is defined as follows:

$$L = \alpha \cdot L_{\mathrm{CL}} + \beta \cdot L_{\mathrm{GER}} + \gamma \cdot L_{\mathrm{DDT}} \tag{14}$$

where α, β, and γ are the weight hyper-parameter for the contrastive learning, the GER task, and the DDT task. We set α and β to 1.0 and γ to 0.01 in all experiments in our study.

2.7. 3D domain detection across ST slices

After encoding the spot features across different ST slices into a coherent embedding space, we use the unsupervised clustering technique to identify the spatial domains in the 3D space. In particular, for datasets with known labels or prior knowledge of the cluster number, stMSA employs the finite Gaussian mixture clustering model from the R package MCLUST [24] to identify spatial domains. In cases where the cluster number is unknown, stMSA utilizes the Louvain clustering algorithm for community detection [23] which has been implemented in the Python package SCANPY [25].

To evaluate the performance of domain detection, we utilize the metric of adjusted Rand index (ARI) [26], which is calculated using the following formula:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}} \tag{15}$$

where n_{ij} denotes the count of spots within the predicted cluster i that are assigned to the ground truth class j, a_i represents the count of spots within the predicted cluster i, b_j denotes the count of spots within the ground truth class j, and n represents the total count of spots. A higher ARI score indicates a more accurate detection of the domains that align with the ground truth labels. The ARI score was calculated using the Python package Scikit-learn [27].

2.8. Cross-slice matching

The objective of matching spots across two adjacent ST slices is to establish a mapping between pairs of spots from different slices, ensuring that

the matched spots exhibit similar gene expression patterns. Additionally, the mapping should preserve the spatial relationship of spots within each ST slice. To achieve this, we calculate the Euclidean distances in the embedding space between each spot s on the source slice and all other spots on the target slice. The spot t on the target slice with the smallest Euclidean distance to spot s is considered its match. To evaluate the quality of the matching, we compare the identity of the spot's label (such as cell type, tissue type) for each pair, and calculate a matching score acc as depicted below:

$$acc = \frac{1}{N}\sum_{i=1}^{N} \mathrm{I}_{\mathrm{src_i=dst_i}} \tag{16}$$

Where src_i and dst_i represent the cell types of spot i in the source slice and its corresponding spot in the target slice, respectively. The indicator function $\mathrm{I}_{src_i=dst_i}$ evaluates to 1 if src_i and dst_i represent the same cell type, and 0 otherwise. The acc score quantifies the accuracy of the matching process. We utilize the Python package FAISS[28] to improve the efficiency of matching spots.

2.9. Multi-slice alignment

Aligning multi-slice data is a key step in reconstructing 2D tissue slices into 3D. In this regard, stMSA incorporates landmark points to guide the alignment process. Given the absence of dedicated ST alignment methods, we have enhanced the Iterative Closest Point (ICP) algorithm [29] to align slices using the domain detection output from stMSA. Furthermore, we have introduced a scoring method named Ali-score to evaluate the quality of the alignment result.

2.9.1. ICP with disturbance

The alignment procedure employed in stMSA comprises three primary components: (1) integration of multi-slice data using stMSA, (2) identification of the landmark domain, and (3) derivation of the transformation matrix through a modified version of the Iterative Closest Point (ICP) algorithm.

The ICP algorithm is renowned for its efficiency and simplicity, making it one of the most extensively utilized point cloud registration methods. The objective function for the ICP alignment is stated as follows:

$$R, t = \arg\min_{R,t} \frac{1}{n}\sum_{i=1}^{n} ||p_i^{(t)} - (R \cdot p_i^{(s)} + t)||^2 \tag{17}$$

where R and t denote the rotation matrix and translation vector, respectively. $p_i^{(s)}$ and $p_i^{(t)}$ represent the source and target point clouds, respectively, and n denotes the point count.

Given the non-convex nature of the problem and the dependency on local iterative steps, the ICP algorithm is prone to converging to local minima [30]. To address this concern, stMSA incorporates a threshold, where the mean error between two iterations should be below the threshold to halt the iteration process. If the threshold is surpassed, the point cloud undergoes rotation, and the iteration persists. Empirically, this strategy effectively enables ICP to avoid local minima.

2.9.2. Alignment scoring method

Currently, there is a lack of a standardized metric for evaluating the alignment results of multi-slice data. In this study, we introduce a novel scoring system called Ali-score, specifically designed for assessing the quality of alignment. The goal of multi-slice alignment is to align the positions of different slices to a global coordinate system, ensuring that cells within the same domain are accurately positioned in corresponding 3D regions across slices.

To evaluate the effectiveness of this alignment method, we begin by identifying the nearest spot in the slice below through the calculation of the Euclidean distance between their aligned coordinates. In order to assess the alignment outcome, we compare the predicted cell type of each spot pair and calculate the percentage of correctly matched spot pairs using the following approach:

$$\mathrm{Ali-score} = \frac{1}{N}\sum_{i=1}^{N} 1_{\mathrm{src_i=dst_i}} \tag{18}$$

The Ali-score serves as a quantitative measure of alignment accuracy, where a higher Ali-score signifies a closer proximity between spots of the same cell type within the aligned regions, indicating a more effective alignment. To calculate the Ali-score, we have developed a customized Python function. Moreover, our algorithm supports the identification of the k nearest neighbors for each spot and incorporates a voting mechanism based on domain labels, allowing for the computation of a more flexible Ali-score.

2.10. The overall structure of stMSA

The dimensions of the layers in the auto-encoder are set as follows: the input data shape is 512 in the encoder GAT layer, and 512 to 30 in the

encoder FCN layer. The decoder is symmetric to the encoder. To optimize the model parameters, we use the Adam optimizer[31] with a learning rate of 0.001. We implement the stMSA model using the popular graph neural network framework PyTorch Geometric (PyG)[32; 33]. We set the number of training epochs to 500. In the cross-batch SMR task, we update the cross-batch pairs, and in the DDT task, we update the DEC p distribution every 100 epochs.

2.11. Benchmarking

The performance of stMSA in multi-slice representation learning, integration, and domain identification is compared with seven state-of-the-art methods: Scanorama [10], Harmony [11], STAGATE [7], SEDR [**?**], STAligner [14], Stitch3D [15], and SPACEL [16]. Additionally, the cross-slice matching capabilities of stMSA are assessed against SLAT [17], while its alignment performance is compared specifically with STAligner [14]. Details on the methods compared are available in Supplementary Note S1.

To evaluate the performance of integration, we calculate the Local Inverse Simpson's Index (LISI)[11] as follows:

$$\mathrm{LISI}_{\mathcal{L}}(x,k) = D_{\mathcal{L}}(k{-}\mathrm{nn}(x))^{-1} \tag{19}$$

where x represents the set of object embedding vectors, $\mathcal{L}$ denotes the corresponding label set for the vector set, k is a perplexity parameter set to 90 by default, $k{-}\mathrm{nn}(\cdot)$ represents the k-nearest neighbors algorithm, and $D_{\mathcal{L}}(\cdot)$ denotes Simpson's index, computed as follows:

$$D_{\mathcal{L}}(S) = \sum_{l\in\mathcal{L}} (\frac{n_l}{n})^2 \tag{20}$$

where S denotes the object vector set, n_l represents the count of vectors belonging to label type l, and n is the overall count of vectors in set S.

The LISI score serves as an indicator of the degree of mixing between different label types, where a higher LISI score signifies improved cohesion and enhanced elimination of batch effects. The LISI score was calculated using the Python package HarmonyPy[11].

To evaluate domain detection, we use the Adjusted Rand Index (ARI) [26], which measures the congruence between the ground truth labels and the predicted clustering results. Further details on ARI are provided in the Multi-slice domain detection section. The matching results of cross-slice

matching are assessed using a Matching Score, detailed in the Cross-slice matching section. Finally, the alignment results are evaluated using the Aliscore, with specifics available in the Alignment scoring method section.

2.12. Data availability

Multiple ST datasets derived from human and mouse tissues, organs, and embryos are used to evaluate the performance of stMSA. Specifically, the DLPFC dataset can be obtained from the Lieber Institute GitHub repository at `https://github.com/LieberInstitute/HumanPilot`. The mouse brain serial dataset can be accessed at STOmicsDB [34] (Anterior: `https://db.cngb.org/stomics/datasets/STDS0000018`, Posterior: `https://db.cngb.org/stomics/datasets/STDS0000021`). The mouse brain coronal dataset can also be found at STOmicsDB (`https://db.cngb.org/stomics/datasets/STDS0000218`). The Slide-seq V2 mouse brain bulb data can be obtained from the Broad Institute webpage (`https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seqv2#study-download`). The Stereo-seq mouse olfactory bulb dataset can be accessed from the SEDR GitHub repository (`https://github.com/JinmiaoChenLab/SEDR_analyses/tree/master/data`). The mouse organogenesis spatiotemporal transcriptomic atlas dataset can be found at STOmicsDB (`https://db.cngb.org/stomics/mosta/spatial/`). All data used in this study can be obtained from our GitHub repository: `https://github.com/hannshu/st_datasets`. A summary of the dataset information can be found in Supplementary Table S1.

2.13. Data pre-processing

To prepare the data for analysis, the raw gene expressions undergo an initial log transformation and normalization using the Python package SCANPY [25]. To overcome the zero-inflation in spatial transcriptomics data, similar to STAGATE, stMSA selects the top k highly variable genes (HVGs) as input features. For ST slices with common gene backgrounds, such as the DLPFC dataset [35], we set k as 3000 by default. For ST slices with varying genes, usually derived from different platforms, it is suggested to set k higher and use the intersecting genes as input features.

3. Results and Discussion

3.1. stMSA improves 3D domain identification across 12 human dorsolateral prefrontal cortex slices

To assess the efficacy of stMSA in identifying spatial domains across multiple slices, we utilized it on the 10x Visium dataset of the human dorsolateral prefrontal cortex (DLPFC) [35]. This dataset comprises 12 slices of the human brain cortex from three donors. Each slice is labeled with annotations for seven spatial domains, including six layers of the brain cortex and the white matter (Figure 2A). These annotations serve as the reference for evaluating the accuracy of domain identification.

We compared the performance of stMSA in domain identification with seven state-of-the-art methods: Scanorama [10], Harmony [11], STAGATE [7], SEDR [**?**], STAligner [14], Stitch3D [15], and SPACEL [16]. For each method, the spot embeddings derived from all slices were used to generate domain predictions using their default clustering methods. Subsequently, we evaluated the performance of each method in domain detection using the Adjusted Rand index (ARI) [26]. The evaluation was conducted for individual slices as well as for all four slices for each donor.

Figure 2B shows the ARI scores obtained by stMSA and seven other methods across the 12 DLPFC slices. The scores were derived from separate representation learning and clustering performed on each slice. The results demonstrated that stMSA outperformed other methods in domain identification, with an average ARI score of 0.544. Although SPACEL achieved an average score of 0.537, its performance heavily relied on the quality of the single-cell data it incorporates. For a comprehensive overview of the domain identification results and the corresponding ARI scores for all 12 slices, please refer to Supplementary Figure S1.

We visualized the results of domain identification by eight methods for a donor (DLPFC slice identifiers: 151507 - 151610) in Figure 2C. Compared to the other methods, stMSA demonstrated more coherent and reasonable domain distributions, achieving the highest ARI score of 0.508. Scanorama (ARI=0.255) and Harmony (ARI=0.222) failed to effectively generate spatial domains, as they do not take spatial information into account. Similarly, STAGATE (ARI=0.406) was unable to generate clear domain boundaries due to the absence of multi-slice integration strategies. Although SEDR (ARI=0.488), Stitch3D (ARI=0.439), and SPACEL (ARI=0.466) successfully partitioned the spots into spatial domains, they exhibited incorrect

layer shapes in layers 1 to 3. While STAligner (ARI=0.467) achieved a better spatial domain distribution compared to the other six methods, stMSA surpassed it by offering a superior ARI score and relatively clearer boundaries.

In Figure 2D and Supplementary Figure S2, we employed the Uniform Manifold Approximation and Projection (UMAP) plot [36] to visualize the latent embeddings obtained by stMSA and other methods using all 12 slices. Additionally, we used the PAGA algorithm [37] to infer the domain trajectory based on the learned embeddings. We can see that stMSA accurately depicted layers 1 to 6 and the white matter, with the linear developmental trajectory aligning well with the ground truth cortex layers. To evaluate the efficacy of stMSA in batch effect removal, we visualized the UMAP plot with the donor labels in Figure 2E. The learned embeddings of the three donors mixed well, without showing clear batch effects. This indicated that stMSA successfully integrated data from multiple donors into a cohesive embedding space. Furthermore, we quantified the integration performance by calculating the Local Inverse Simpson's Index (LISI) score [11] for each method across all 12 slices (Figure 2F). The LISI score further confirmed stMSA's ability to eliminate batch effects. In this case, stMSA achieved the highest average LISI score. This indicated the successful mitigation of batch effects and a high level of local homogeneity across batches in the integrated dataset. It is worth noting that STAGATE and Stitch3D could not be executed successfully on our server with 24GB GPU memory due to their excessive GPU memory requirements.

3.2. stMSA accurately detects cross-batch tissue structures in the mouse brain

In ST sequencing technologies such as 10x Visium, a single slice can typically cover up to a limited number of spots. However, when dealing with larger tissues like the mouse brain, sequencing the entire tissue using a single slice becomes challenging. As a result, researchers often divide the tissue into multiple sections and apply sequencing methods to each individual slice. An important task in integrating these multi-slice datasets is to identify common spatial domains across different slices and assign them consistent domain labels. To evaluate the capability of stMSA in detecting shared domains across slices, we employed stMSA on a dataset of mouse brain sagittal sections obtained through 10x Visium sequencing technology. We also compared the performance of stMSA with two highly effective methods, SEDR and STAligner (Figure 3A, B, C). Considering the substantial volume of the mouse

brain structure, the data was divided into two slices: the posterior (Figure 3B, left part) and the anterior (Figure 3B, right part).

Figure 3B highlights the domain detection performance of stMSA in comparison to the manually annotated Allen Brain Atlas (Figure 3A) and results derived by STAligner and SEDR (Figure 3C). Both stMSA and STAligner accurately depicted the structure and boundaries of cross-slice structures, such as the isocortex and thalamus. However, SEDR struggled to distinguish the six cortex layers within the isocortex. Notably, stMSA was the only method capable of accurately identifying the hippocampus and its inner subdomains across slices. In Figure 3B, stMSA precisely delineated the "hook" shaped outline of Ammon's horn, while both STAligner and SEDR failed to separate it into distinct parts (domain 3 and domain 7 in STAligner, and domain 4 and domain 12 in SEDR). Furthermore, SEDR inaccurately identified the lower anterior section of the hippocampus, assigning it to domain 13, which did not align with the expected hippocampus structure observed across slices (designated as domain 18 and domain 7).

Upon closer examination of the fine-grained structures identified by stMSA, we found that stMSA can accurately recognize subdomains within landmark brain structures, such as the multiple layers in the isocortex and hippocampus. To confirm the accuracy of subdomain partitioning, we compared the distribution of these subdomains with genes that are known to exhibit significant variability within specific domains (Figure 3D, Supplementary Figure S3A). For instance, genes like *Rprm* and *Mobp*, which are highly expressed in cortical layers 6a and 6b respectively [38; 39; 40; 41], displayed consistent expression patterns corresponding with the predicted domains. Moreover, the gene *Ckb*, known for its high expression in the pyramidal layer [42], exhibited a "curved tube" inner structure that aligned well with our domain detection results for the hippocampus. In addition, the spatial distribution of the marker gene *Prkcd* for the thalamus [43] exhibited a significant overlap with the structure identified by stMSA. Moreover, the UMAP plot and trajectory inference generated from the embeddings learned by stMSA (Figure 3E, Supplementary Figure S3B) provided further evidence of the correspondence between the relative positions of different layers of the isocortex in the feature space and their actual physical positions. These findings highlight the effectiveness of stMSA as a valuable tool for analyzing cross-slice brain structures.

3.3. stMSA achieves robust performance across diverse experimental protocols and sequencing technologies

To assess the ability of stMSA to integrate spatial transcriptomics data from various experimental protocols, we conducted an integration analysis using three 10x Visium-based mouse brain coronal slices. The original tissues were preserved using fresh frozen tissue with H&E staining, fresh frozen tissue with DAPI staining, and formalin-fixed paraffin-embedded (FFPE) tissue with H&E staining (Figure 4A). These datasets are referred to as "Frozen", "DAPI Stained", and "FFPE" respectively. Using stMSA, we compared its performance to SEDR and STAligner in jointly analyzing these slices and obtaining embeddings in the same feature space. Subsequently, the domains were predicted in the feature space but visualized separately in the slices (Figure 4B, C, D).

Figure 4B, C, and D demonstrate that both stMSA and STAligner accurately identified the same tissue structures with consistent domain labels across protocols. Conversely, SEDR struggled to consistently identify tissue structures across the slices. For instance, stMSA and STAligner assigned identical domain labels to the Dentate pyrus structure on all three slices. Specifically, stMSA labeled it as domain 1, and STAligner labeled it as domain 8. These assignments align well with the histology images and the Allen Brain Atlas annotation (Supplementary Figure S4). In contrast, SEDR exhibited inconsistencies in identifying the tissue across different protocols. Additionaly, both stMSA and STAligner consistently labeled the layer structures in the isocortex across protocols. stMSA assigned clusters 2, 8, 5, and 11 to the layers, and STAligner assigned clusters 10, 6, 12, and 14. Conversely, SEDR assigned different domain labels even for the same isocortex layer.

Furthermore, stMSA successfully distinguished between the hypothalamus and white matter (fiber tracts/oligodendrocytes), accurately identifying structures in domain 7 and domain 15. In contrast, STAligner faced challenges in differentiating these structures, resulting in misclassification into domain 1 and domain 4 respectively. Moreover, stMSA achieved the highest domain detection Adjusted Rand Index (ARI) score across all three slices, providing evidence of its superior ability to learn high-quality latent representations. Furthermore, the stMSA algorithm demonstrated its effectiveness in accurately distinguishing between the hypothalamus and white matter (specifically fiber tracts and oligodendrocytes) in domain 7 and domain 15. On the other hand, STAligner encountered difficulties in accurately differ-

entiating these structures, which led to misclassification into domain 1 and domain 4, respectively. Additionally, stMSA outperformed other methods in terms of ARI scores across all three slices.

To assess how effectively stMSA can integrate ST data obtained from different sequencing platforms, we applied stMSA to mouse olfactory bulb data acquired using both the Stereo-seq [5] and Slide-seq V2 [4] methods. Similar to the previous experimental settings, we jointly analyzed the two ST slices using each method, with domain prediction taking place in the same feature space. As depicted in Figure 4E, stMSA successfully recognized consistent structures in both slices, aligning well with the Allen Brain Atlas annotations (Supplementary Figure S5A). Supplementary Figure S5B demonstrates that STAligner achieved similar results, while SEDR encountered difficulties in generating clear domain structures.

We further evaluated the effectiveness of integration by assessing the learned representations from different methods. Specifically, we compared the LISI scores of stMSA and STAligner in both the mouse brain coronal section and the mouse olfactory bulb datasets (Figure 4F). The results consistently showed that stMSA outperformed STAligner, as it achieved higher LISI scores. This indicates that stMSA has a superior ability to encode features from various embedding spaces into a common embedding space. To be noted, the LISI score was ignored for the SEDR method, since we found that when we applied SEDR to the mouse olfactory bulb dataset, it did not learn well-clustered latent representations. This can be observed in the UMAP plot of SEDR's representation (Supplementary Figure S5C), where the latent representations of different domains are erroneously intertwined.

3.4. stMSA demonstrates effective tissue matching abilities in a mouse embryo developmental dataset

Matching spots across two adjacent ST slices is crucial for gaining insights into the developmental trajectory of cell types, tissues, and organogenesis. This process involves establishing a mapping between pairs of spots from different slices while ensuring that the matched spots exhibit similar gene expression patterns. Furthermore, it is important for the mapping to preserve the spatial relationship of spots within each ST slice. To assess the cross-slice matching capability of stMSA, we employed it on a Mouse Organogenesis Spatiotemporal Transcriptomic Atlas (MOSTA) dataset. Specifically, we utilized stMSA to match tissue types between two mouse embryo slices sectioned at embryonic day 14.5 (E14.5) and day 15.5 (E15.5). Figure 5A and

B demonstrate the overall matching between the two slices, with a majority of tissue types in mature organs, such as the brain, heart, muscle, liver, and lung, being well matched.

Moreover, this matching process allows us to reveal the specific organogenesis stages. For example, we found that cartilage at stage E15.5 originates from the cartilage and cartilage primordium at stage E14.5. This evidence was further supported by the matching results depicted in Figure 5C, which only showed the matched spot pairs between the two tissues. To validate this developmental relationship, we visualized the spatial distribution of *Col2a1*, the marker gene of cartilage [44], alongside the domain distribution of cartilage and cartilage primordium in stage E14.5. Remarkably, the expression pattern of *Col2a1* aligned well with both domain distributions.

To assess the overall matching performance of stMSA, we employed a confidence score to quantitatively evaluate the similarity between the embedded spots that were matched. This confidence score was defined as the cosine distance between the embeddings of two matched spots. In Figure 5E, we present the confidence score for the slice at stage E14.5, clearly illustrating the high degree of similarity between the matched spots. Furthermore, we conducted an analysis of the cross-slice matching between pairs of slices at other stages, namely, E9.5-E10.5, E10.5-E11.5, E11.5-E12.5, E12.5-E13.5, and E13.5-E14.5. We compared the matching process with SLAT, which is also capable of performing this task. It is worth noting that SLAT has four different initialization settings based on the input type: using all genes (Raw init), using only the highly variable genes (HVG init), using dimension-reduced gene expression profiles based on principle component analysis (PCA init), or dual PCA (DPCA init) [17]. Figure 5F presents the matching scores obtained by stMSA and the four SLAT versions (details provided in Supplementary Note S1), clearly demonstrating the superior matching performance of stMSA over SLAT in various settings.

3.5. stMSA enables the multi-slice alignment in mouse brain coronal sections

Current sequencing technologies are limited to capturing two-dimensional (2D) slices of biological tissues, which hinders researchers' ability to gain a comprehensive understanding of three-dimensional (3D) biological structures. The construction of 3D views of ST slices currently requires manual manipulation to align and orient the slices to a common coordinate system. To demonstrate the ability of stMSA in automatic ST data alignment, we applied this method to a mouse brain coronal section dataset containing 7 ST

slices (Ballester et al., 2023). It is worth noting that the dataset includes two flipped orientations labeled "Orient 1" and "Orient 2", which can potentially impact the integration and alignment performance.

For the alignment process, stMSA relies on identifying a landmark domain to guide the Iterative Closest Point (ICP) algorithm in learning a rotation matrix (see Materials and Methods). As shown in Figure 6B, stMSA initially identified spatial domains, which were well aligned with the Allen Brain Atlas annotations (Supplementary Figure S4A). Taking the pyramidal layer (domain 7) in the hippocampus as an example, this domain aligned seamlessly with the hook-like structure observed in the histology image (Figure 6A), as well as the expression pattern of the marker gene *Hpca*[45] (6C). Based on this observation, we selected the pyramidal layer as the landmark domain for guiding the rotation matrix learning.

Figure 6D (left) visually demonstrates the challenges of the alignment process by displaying the original position and orientation of the dataset, revealing noticeable misalignment. To address this issue, we employed a modified ICP algorithm (see Materials and Methods) on the "Orient 1" and "Orient 2" datasets. By utilizing this approach, stMSA effectively coordinated the rotation and flipping of individual slices, seamlessly integrating them into a cohesive positional system. This alignment process ensures a uniform orientation across all slices, thereby facilitating well-aligned visualization and interpretation of the spatial transcriptomics data in 3D views(Figure 6D, right).

To further evaluate the alignment performance of stMSA, we compared it with STAligner. The aerial views of the aligned slices are shown in Figure 6E. It is evident that stMSA successfully aligns the slices, resulting in a notable coherence in the proximity of spots belonging to the same domain. This visual coherence is depicted in the aerial views, which display domain distributions consistent with the distributions observed in individual slices. In contrast, the distribution produced by STAligner appears less coherent. Furthermore, we calculated an alignment score (see Materials and Methods) to quantitatively assess the alignment performance, and the results demonstrated that stMSA outperforms STAligner in both orientations.

4. Conclusion

The rapid advancement of sequencing methods has led to the generation of an increasing amount of high-quality spatial transcriptomics data.

Consequently, there is a growing need to integrate ST data from multiple sources which might differ in experimental protocols, sequencing platforms, size and shape, and orientations. In this work, we presented a novel contrastive learning-based method, named stMSA, for multiple ST data integration. stMSA learns cross-batch spot embeddings by optimizing two principles: (1) similarity of embeddings within a slice for spots in connected micro-environments and (2) similarity of embeddings between cross-batch spots with similar gene expression profiles. By adhering to these principles, stMSA captures the inner-slice and cross-slice distribution patterns of spots. Additionally, stMSA leverages graph attention-based auto-encoder techniques and deep embedding clustering to optimize model parameters and reconstruct gene expression profiles. We extensively evaluate stMSA's performance in various downstream tasks, including batch-effect correction, 3D domain detection, cross-slice matching, and multi-slice alignment. Our results demonstrate that stMSA provides a comprehensive solution for integrating multiple ST slices, enabling advanced analysis of spatial transcriptomics data. We strongly believe that as spatial transcriptomics sequencing technologies continue to advance, stMSA will become an invaluable tool for researchers, enabling them to uncover novel biological insights from a broader range of datasets.

5. Availability of data and materials

The implemented code is available online at `https://github.com/hannshu/stMSA`.

6. Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

7. Acknowledgments

We thank all of the contributors of the open-source datasets and freely available tools used in this study. We also appreciate the suggestive comments of reviewers. This work has been supported by the National Natural Science Foundation of China (grant number 62102319) and the Natural Science Project of Shaanxi Provincial Department of Education (grant number 23JK0562).

References

[1] Cameron G. Williams, Hyun Jae Lee, Takahiro Asatsuma, Roser Vento-Tormo, and Ashraful Haque. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, 14(1):68, June 2022.

[2] Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul Igor Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg, and Jonas Frisén. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, July 2016. Publisher: American Association for the Advancement of Science.

[3] Andrew L. Ji, Adam J. Rubin, Kim Thrane, Sizun Jiang, David L. Reynolds, Robin M. Meyers, Margaret G. Guo, Benson M. George, Annelie Mollbrink, Joseph Bergenstråhle, Ludvig Larsson, Yunhao Bai, Bokai Zhu, Aparna Bhaduri, Jordan M. Meyers, Xavier Rovira-Clavé, S. Tyler Hollmig, Sumaira Z. Aasi, Garry P. Nolan, Joakim Lundeberg, and Paul A. Khavari. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell*, 182(2):497–514.e22, July 2020.

[4] Robert R. Stickels, Evan Murray, Pawan Kumar, Jilong Li, Jamie L. Marshall, Daniela J. Di Bella, Paola Arlotta, Evan Z. Macosko, and Fei Chen. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology*, 39(3):313–319, March 2021. Number: 3 Publisher: Nature Publishing Group.

[5] Xiaoyu Wei, Sulei Fu, Hanbo Li, Yang Liu, Shuai Wang, Weimin Feng, Yunzhi Yang, Xiawei Liu, Yan-Yun Zeng, Mengnan Cheng, Yiwei Lai, Xiaojie Qiu, Liang Wu, Nannan Zhang, Yujia Jiang, Jiangshan Xu, Xiaoshan Su, Cheng Peng, Lei Han, Wilson Pak-Kin Lou, Chuanyu Liu, Yue Yuan, Kailong Ma, Tao Yang, Xiangyu Pan, Shang Gao, Ao Chen, Miguel A. Esteban, Huanming Yang, Jian Wang, Guangyi Fan, Longqi Liu, Liang Chen, Xun Xu, Ji-Feng Fei, and Ying Gu. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regenera-

tion. *Science*, 377(6610):eabp9444, September 2022. Publisher: American Association for the Advancement of Science.

[6] Mengnan Cheng, Yujia Jiang, Jiangshan Xu, Alexios-Fotios A. Mentis, Shuai Wang, Huiwen Zheng, Sunil Kumar Sahu, Longqi Liu, and Xun Xu. Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. *Journal of Genetics and Genomics*, 50(9):625–640, September 2023.

[7] Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1):1739, April 2022.

[8] Jiachen Li, Siheng Chen, Xiaoyong Pan, Ye Yuan, and Hong-Bin Shen. Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science*, 2(6):399–408, June 2022.

[9] Marie Schott, Daniel León-Periñán, Elena Splendiani, Leon Strenger, Jan Robin Licha, Tancredi Massimo Pentimalli, Simon Schallenberg, Jonathan Alles, Sarah Samut Tagliaferro, Anastasiya Boltengagen, Sebastian Ehrig, Stefano Abbiati, Steffen Dommerich, Massimiliano Pagani, Elisabetta Ferretti, Giuseppe Macino, Nikos Karaiskos, and Nikolaus Rajewsky. Open-ST: High-resolution spatial transcriptomics in 3D, December 2023. Pages: 2023.12.22.572554 Section: New Results.

[10] Brian Hie, Bryan Bryson, and Bonnie Berger. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nature Biotechnology*, 37(6):685–691, June 2019.

[11] Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, 16(12):1289–1296, December 2019. Number: 12 Publisher: Nature Publishing Group.

[12] Liangchen Yue, Feng Liu, Jiongsong Hu, Pin Yang, Yuxiang Wang, Junguo Dong, Wenjie Shu, Xingxu Huang, and Shengqi Wang. A guidebook of spatial transcriptomic technologies, data resources and analysis approaches. *Computational and Structural Biotechnology Journal*, 21:940–955, January 2023.

[13] Hang Xu, Huazhu Fu, Yahui Long, Kok Siong Ang, Raman Sethi, Kelvin Chong, Mengwei Li, Rom Uddamvathanak, Hong Kai Lee, Jingjing Ling, Ao Chen, Ling Shao, Longqi Liu, and Jinmiao Chen. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1):12, January 2024.

[14] Xiang Zhou, Kangning Dong, and Shihua Zhang. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science*, October 2023.

[15] Gefei Wang, Jia Zhao, Yan Yan, Yang Wang, Angela Ruohao Wu, and Can Yang. Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nature Machine Intelligence*, pages 1–14, October 2023. Publisher: Nature Publishing Group.

[16] Hao Xu, Shuyan Wang, Minghao Fang, Songwen Luo, Chunpeng Chen, Siyuan Wan, Rirui Wang, Meifang Tang, Tian Xue, Bin Li, Jun Lin, and Kun Qu. SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nature Communications*, 14(1):7603, November 2023. Number: 1 Publisher: Nature Publishing Group.

[17] Chen-Rui Xia, Zhi-Jie Cao, Xin-Ming Tu, and Ge Gao. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nature Communications*, 14(1):7236, November 2023. Number: 1 Publisher: Nature Publishing Group.

[18] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised Deep Embedding for Clustering Analysis, May 2016. arXiv:1511.06335 [cs].

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks, February 2018. arXiv:1710.10903 [cs, stat].

[20] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding Neural Networks Through Deep Visualization, June 2015. arXiv:1506.06579 [cs].

[21] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), February 2016. arXiv:1511.07289 [cs].

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[23] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008. arXiv:0803.0476 [cond-mat, physics:physics].

[24] Luca Scrucca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, 8(1):289–317, 2016.

[25] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, February 2018.

[26] Ka Yee Yeung and Walter L Ruzzo. Details of the Adjusted Rand indexand Clustering algorithms Supplement to the paper "An empirical study on Principal Component Analysis for clustering gene expression data" (to appear in Bioinformatics).

[27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

[28] Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. The Faiss library, January 2024.

[29] P.J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.

[30] Jiaolong Yang, Hongdong Li, Dylan Campbell, and Yunde Jia. Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2241–2254, November 2016. arXiv:1605.03344 [cs].

[31] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].

[32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[33] Matthias Fey and Jan Eric Lenssen. Fast Graph Representation Learning with PyTorch Geometric, April 2019. arXiv:1903.02428 [cs, stat].

[34] Zhicheng Xu, Weiwen Wang, Tao Yang, Ling Li, Xizheng Ma, Jing Chen, Jieyu Wang, Yan Huang, Joshua Gould, Huifang Lu, Wensi Du, Sunil Kumar Sahu, Fan Yang, Zhiyong Li, Qingjiang Hu, Cong Hua, Shoujie Hu, Yiqun Liu, Jia Cai, Lijin You, Yong Zhang, YuXiang Li, Wenjun Zeng, Ao Chen, Bo Wang, Longqi Liu, Fengzhen Chen, Kailong Ma, Xun Xu, and Xiaofeng Wei. STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Research*, 52(D1):D1053–D1061, January 2024.

[35] Kristen R. Maynard, Leonardo Collado-Torres, Lukas M. Weber, Cedric Uytingco, Brianna K. Barry, Stephen R. Williams, Joseph L. Catallini, Matthew N. Tran, Zachary Besich, Madhavi Tippani, Jennifer Chew, Yifeng Yin, Joel E. Kleinman, Thomas M. Hyde, Nikhil Rao, Stephanie C. Hicks, Keri Martinowich, and Andrew E. Jaffe. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, 24(3):425–436, March 2021. Number: 3 Publisher: Nature Publishing Group.

[36] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, February 2018.

[37] F. Alexander Wolf, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J. Theis. PAGA: graph abstraction reconciles clustering with

trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1):59, March 2019.

[38] Kaitlin E. Sullivan, Larissa Kraus, Margarita Kapustina, Lihua Wang, Tara R. Stach, Andrew L. Lemire, Jody Clements, and Mark S. Cembrowski. Sharp cell-type-identity changes differentiate the retrosplenial cortex from the neocortex. *Cell Reports*, 42(3):112206, March 2023.

[39] Joachim S. Grimstvedt, Andrew M. Shelton, Anna Hoerder-Suabedissen, David K. Oliver, Christin H. Berndtsson, Stefan Blankvoort, Rajeevkumar R. Nair, Adam M. Packer, Menno P. Witter, and Clifford G. Kentros. A multifaceted architectural framework of the mouse claustrum complex. *Journal of Comparative Neurology*, 531(17):1772–1795, 2023. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.25539.

[40] Anna Hoerder-Suabedissen, Gabriel Ocana-Santero, Thomas H Draper, Sophie A Scott, Jesse G Kimani, Andrew M Shelton, Simon J B Butt, Zoltán Molnár, and Adam M Packer. Temporal origin of mouse claustrum and development of its cortical projections. *Cerebral Cortex (New York, NY)*, 33(7):3944–3959, September 2022.

[41] Ben D. Fulcher, John D. Murray, Valerio Zerbi, and Xiao-Jing Wang. Multimodal gradients across mouse cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4689–4695, March 2019.

[42] Anna Sandebring-Matton, Michael Axenhus, Nenad Bogdanovic, Bengt Winblad, Sophia Schedin-Weiss, Per Nilsson, and Lars O. Tjernberg. Microdissected Pyramidal Cell Proteomics of Alzheimer Brain Reveals Alterations in Creatine Kinase B-Type, 14-3-3-gamma, and Heat Shock Cognate 71. *Frontiers in Aging Neuroscience*, 13, 2021.

[43] Kirti Sharma, Sebastian Schmitt, Caroline G Bergner, Stefka Tyanova, Nirmal Kannaiyan, Natalia Manrique-Hoyos, Karina Kongi, Ludovico Cantuti, Uwe-Karsten Hanisch, Mari-Anne Philips, Moritz J Rossner, Matthias Mann, and Mikael Simons. Cell type– and brain region–resolved mouse brain proteome. *Nature neuroscience*, 18(12):1819–31, December 2015.

[44] T. N. Hissnauer, A. Baranowsky, J. M. Pestka, T. Streichert, K. Wiegandt, C. Goepfert, F. T. Beil, J. Albers, J. Schulze, P. Ueblacker, J. P. Petersen, T. Schinke, N. M. Meenen, R. Pörtner, and M. Amling. Identification of molecular markers for articular cartilage. *Osteoarthritis and Cartilage*, 18(12):1630–1638, December 2010.

[45] Anastassios V. Tzingounis, Masaaki Kobayashi, Ken Takamatsu, and Roger A. Nicoll. Hippocalcin Gates the Calcium Activation of the Slow Afterhyperpolarization in Hippocampal Pyramidal Cells. *Neuron*, 53(4):487–493, February 2007.

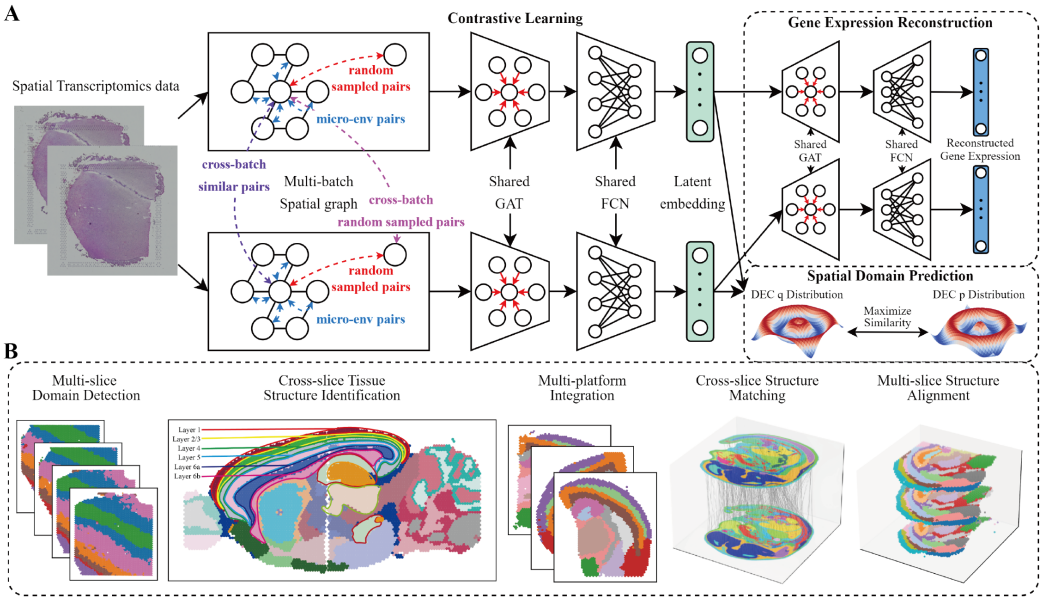

Figure 1: **Overview of the stMSA framework. (A)** stMSA leverages multi-slice gene expression and spatial coordination information as input. It utilizes an auto-encoder to learn latent batch-corrected multi-slice representation and optimizes the model through three distinct optimization tasks. **(B)** The latent embedding learned by stMSA serves various downstream tasks, including multi-slice domain detection, cross-slice tissue structure identification, multi-platform integration, cross-slice matching, and multi-slice alignment.

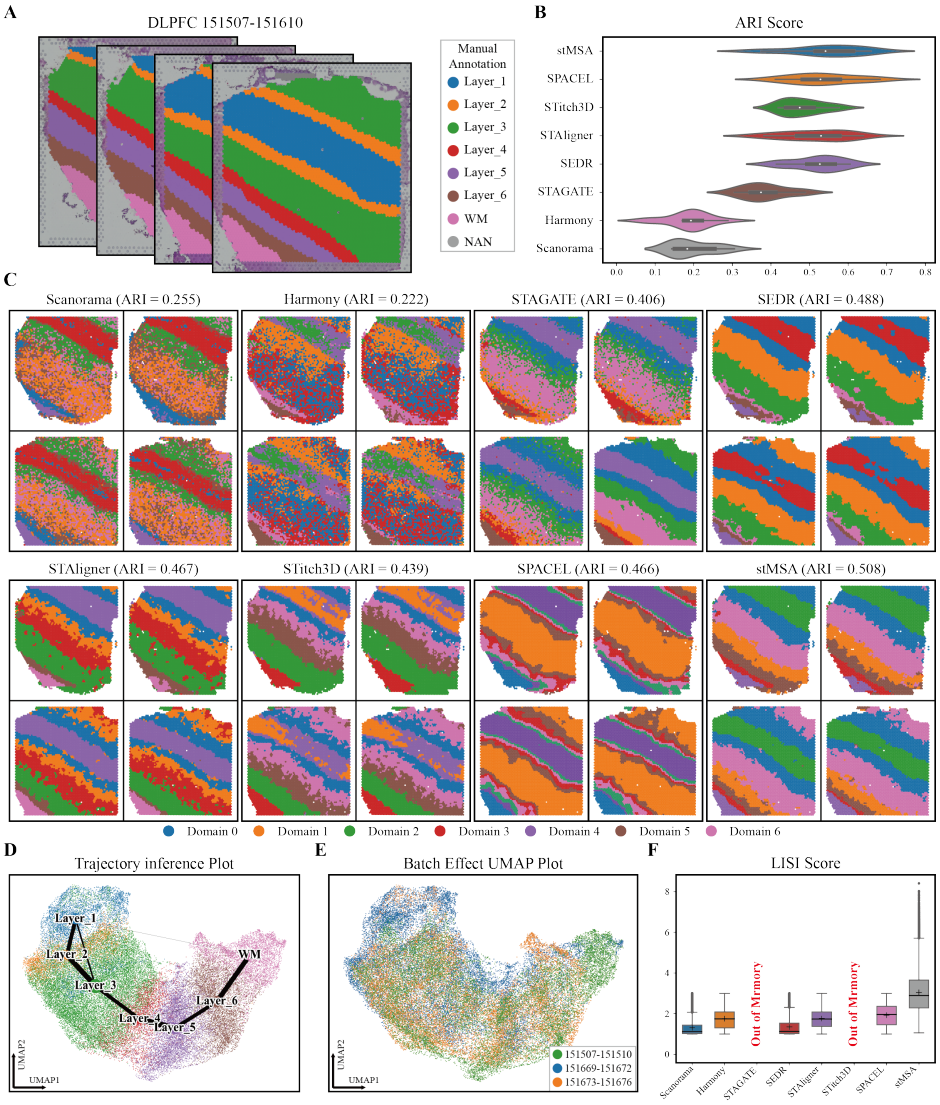

Figure 2: **stMSA improves 3D domain identification across the 12 spatial transcriptomics slices from the human dorsolateral prefrontal cortex (DLPFC) dataset. (A)** Manual annotations, ranging from the cortex layer 1 to the white matter, are provided with the histology image of a donor (slice identifiers: 151507-151610). **(B)** The violin plot illustrates the ARI scores across 12 DLPFC slices, generated by training separately for each donor. **(C)** Visualization of domain detection results for Scanorama, Harmony, STAGATE, SEDR, STAligner, Stitch3D, SPACEL, and stMSA. The domain detection processes are conducted using all four slices as inputs. **(D)** The UMAP and trajectory inference plot generated based on the embeddings learned by stMSA using all 12 DLPFC slices. The color reflects the ground truth domains. **(E)** The UMAP visualization for batch effects. The color reflects three donnors. **(F)** The LISI scores for the eight methods across all 12 slices. STAGATE and Stitch3D were unable to process all 12 slices due to the excessive demand for GPU memory.

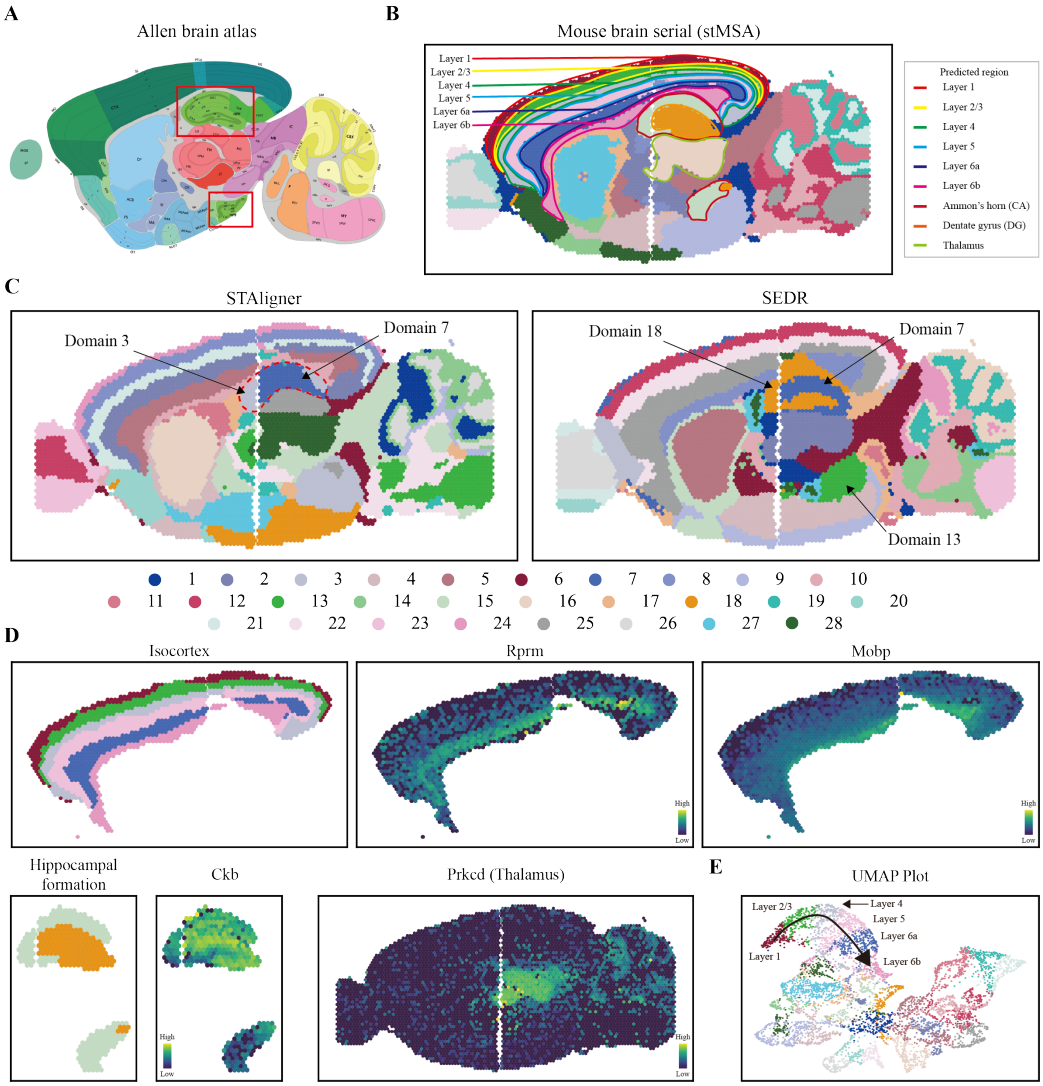

Figure 3: **stMSA can identify cross-batch tissue structures in the 10x Visium-based mouse brain dataset. (A)** The anatomical structure of sagittal mouse brain provided by the Allen Brain Atlas. **(B)** Spatial domains predicted by stMSA with manually annotated landmark domains (isocortex from layer 1 to layer 6b, hippocampus, and thalamus). **(C)** Spatial domains predicted by STAligner and SEDR. **(D)** Subdomains of the isocortex, hippocampus, and thalamus, along with the expression levels of their corresponding marker genes. **(E)** The UMAP visualization of latent embeddings derived by stMSA. The substructures of the isocortex from layer 1 to layer 6b are manually annotated on the plot.

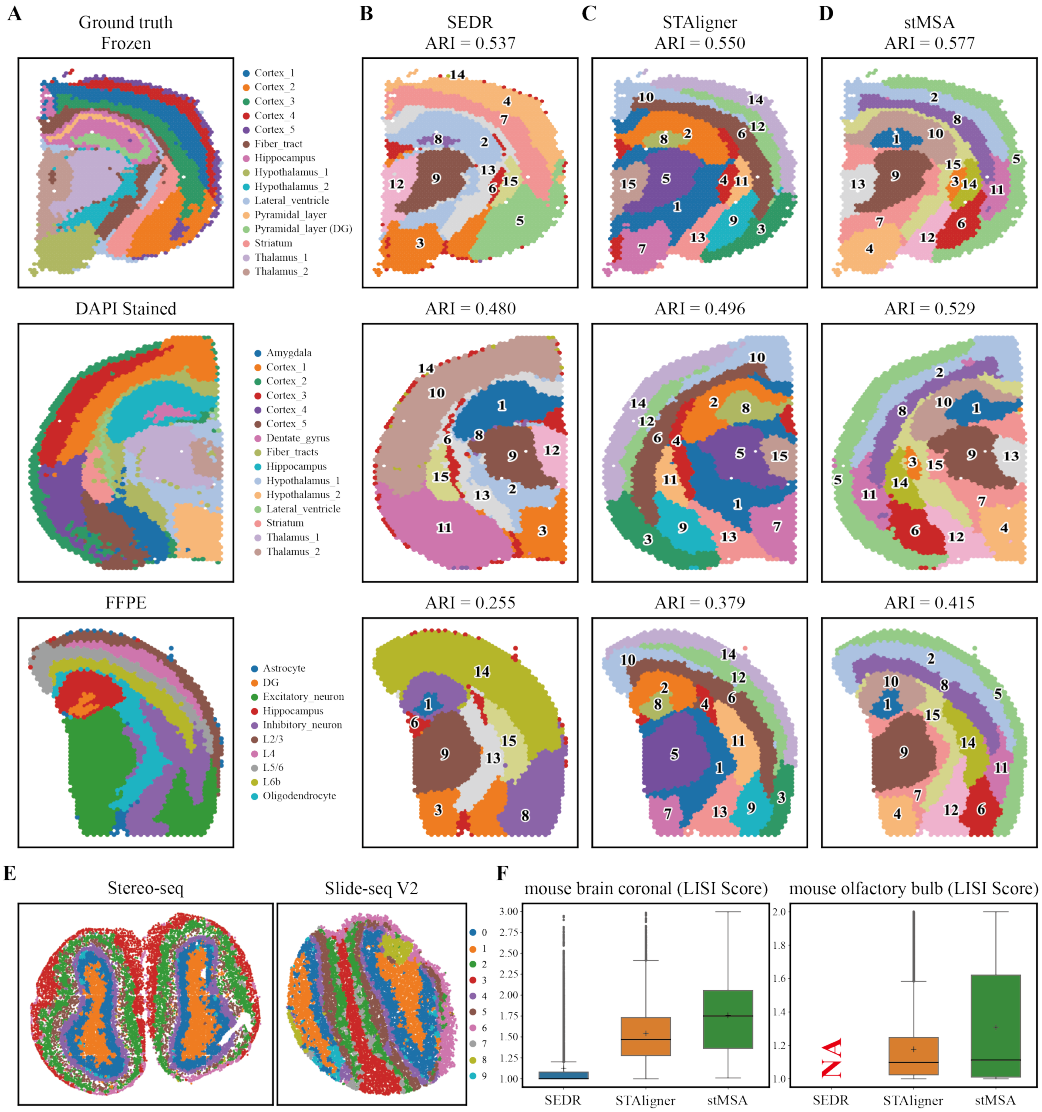

Figure 4: **stMSA achieves robust performance across diverse experimental protocols and sequencing technologies in mouse brain coronal sections.** The manual annotations **(A)** and predicted domains based on SEDR **(B)**, STAligner **(C)**, and stMSA **(D)** across mouse brain coronal sections derived from three experimental protocols.**(E)** Visualization of domains predicted by stMSA on the mouse olfactory bulb slices obtained through two sequencing technologies (Stereo-seq and Slide-seq V2). **(F)** The LISI scores measuring the integration performance for SEDR, STAligner, and stMSA in the mouse brain coronal dataset and mouse olfactory bulb dataset.

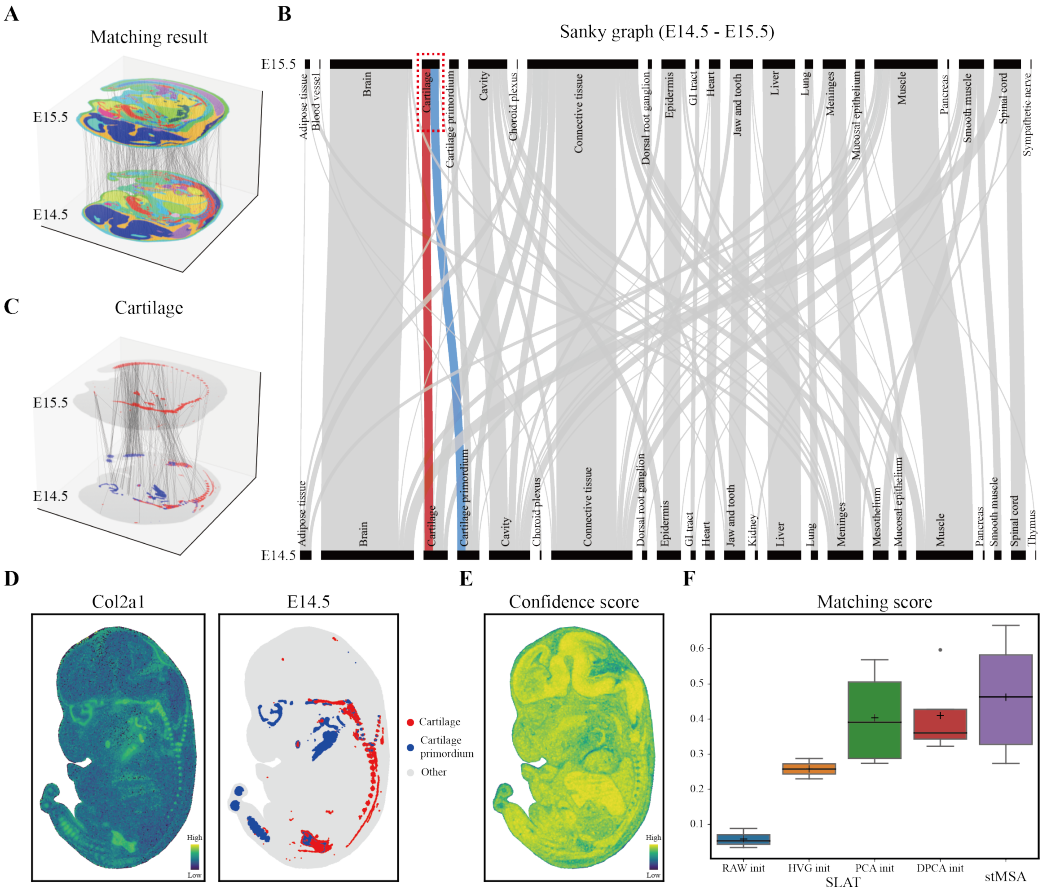

Figure 5: **stMSA demonstrates effective tissue matching abilities in the Mouse Organogenesis Spatiotemporal Transcriptomic Atlas (MOSTA) dataset. (A)** Visualization of the cross-slice matching results between E14.5 and E15.5 stages of mouse embryo. To improve the clarity of the matching results, 500 matched spot pairs were randomly selected. **(B)** The Sankey graph illustrating the tissue matching between E14.5 and E15.5 stages, with a highlighted path showcasing the matching relation between cartilage and cartilage primordium. **(C)** Matched spot pairs specifically focusing on the cartilage-cartilage primordium relationship at the E14.5 and E15.5 stages. **(D)** Marker gene expression patterns and spatial domain distributions of cartilage and cartilage primordium at E14.5 stage. **(E)** Confidence scores of the matching results, where higher scores indicate greater matching confidence, and vice versa. **(F)** Comparison of the matching scores between stMSA and SLAT under different settings (RAW, HVG, PCA, and DPCA init).

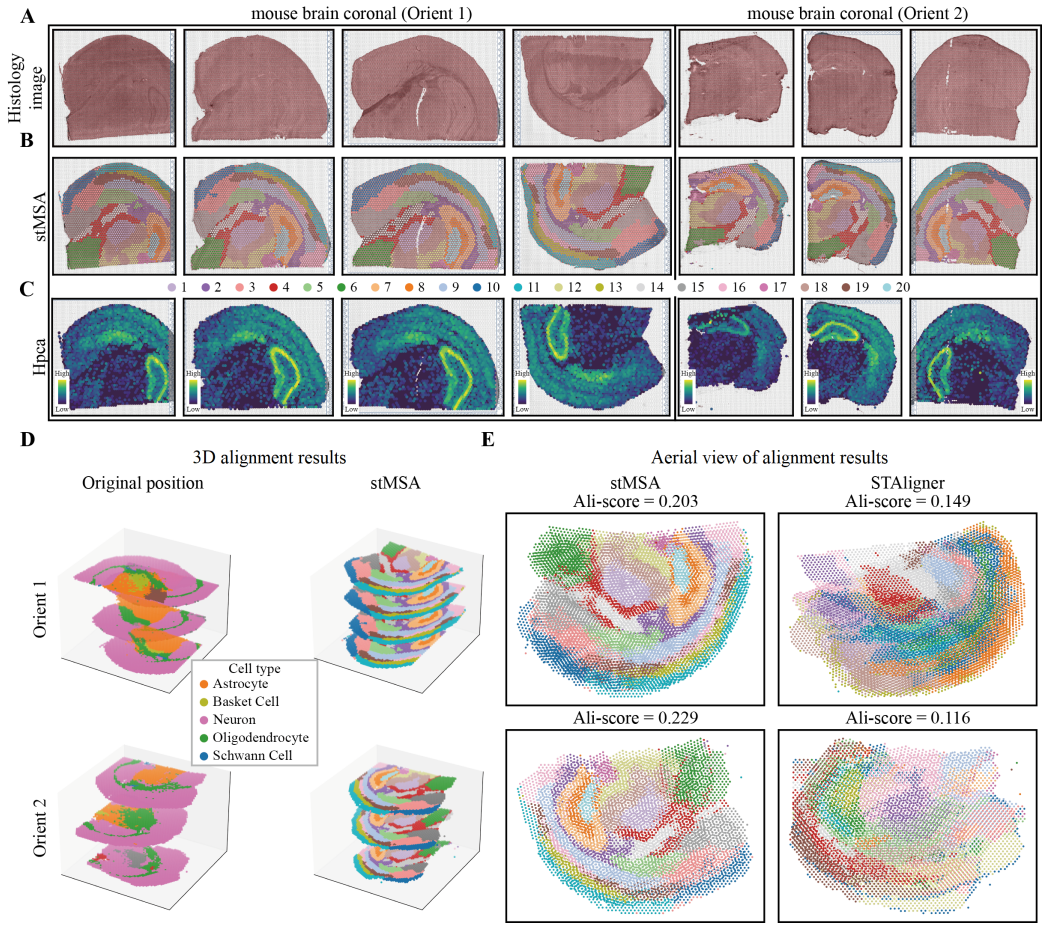

Figure 6: **The performance of stMSA on a mouse brain coronal section ST dataset.** **(A)** Original orientation of the histology image of the mouse brain coronal dataset. **(B)** Visualization of spatial domains detected by stMSA. **(C)** Spatial distribution of marker genes for the pyramidal layer in the hippocampus. **(D)** 3D visualization showing the original position of the dataset and the alignment results achieved by stMSA. **(E)** Aerial view of alignment results comparing stMSA with STAligner.