JADE: Joint Alignment and Deep Embedding for Multi-Slice Spatial Transcriptomics

Yuanchuan Guo
Department of Statistics
Harvard University
yguo1@g.harvard.edu

Jun S. Liu*
Department of Statistics
Tsinghua University
junsliu@tsinghua.edu.cn

Huimin Cheng*
Department of Biostatistics
Boston University
huimin23@bu.edu

Ying Ma*
Department of Biostatistics
Center for Computational Molecular Biology
Brown University
ying_ma@brown.edu

Abstract

As spatially resolved transcriptomics (SRT) datasets increasingly span multiple adjacent or replicated slices, effective joint analysis across slices is needed to reconstruct tissue structures and identify consistent spatial gene expression patterns. This requires resolving spatial correspondences between slices while capturing shared transcriptomic features, two tasks that are typically addressed in isolation. Multi-slice analysis remains challenging due to physical distortions, technical variability, and batch effects. To address these challenges, we introduce Joint Alignment and Deep Embedding for multi-slice SRT (JADE), a unified computational framework that simultaneously learns spatial location-wise alignments and shared low-dimensional embeddings across tissue slices. Unlike existing methods, JADE adopts a roundtrip framework in which each iteration alternates between alignment and embedding refinement. To infer alignment, we employ attention mechanisms that dynamically assess and weight the importance of different embedding dimensions, allowing the model to focus on the most alignment-relevant features while suppressing noise. To the best of our knowledge, JADE is the first method that jointly optimizes alignment and representation learning in a shared latent space, enabling robust multi-slice integration. We demonstrate that JADE outperforms existing alignment and embedding methods across multiple evaluation metrics in the 10x Visium human dorsolateral prefrontal cortex (DLPFC) and Stereo-seq axolotl brain datasets. By bridging spatial alignment and feature integration, JADE provides a scalable and accurate solution for cross-slice analysis of SRT data.

1 Introduction

Spatially resolved transcriptomics (SRT) technologies provide high-throughput measurements of gene expression within tissue sections while preserving spatial context [6, 43, 47, 60, 28]. By enabling the joint study of molecular states and tissue architecture, SRT has transformed our understanding of developmental processes [9, 70], disease microenvironments [46, 81], and spatial cellular organization across diverse biological systems, from cancer [55, 5, 59, 26, 46, 1, 16] to neuroscience [42, 44, 52, 75]. As the resolution and scale of SRT continue to improve, there is growing interest in multi-slice SRT datasets, where multiple adjacent or replicated tissue sections are profiled to reconstruct 3D structures, map spatial trajectories, or assess reproducibility across individuals and conditions [19, 56, 21, 41, 14, 30, 40, 34]. However, analyzing multi-slice SRT data introduces a set of unique

*Co-corresponding authors

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

challenges. Each tissue slice may undergo non-linear spatial deformation during sectioning and mounting, while also exhibiting substantial variation in transcript capture efficiency and local tissue composition [77, 36]. These factors confound the alignment of corresponding regions across slices and obscure shared biological signals. Effective multi-slice integration must therefore address two tightly coupled tasks: spatial alignment and representation learning [38, 66, 35]. Alignment resolves spatial location -level correspondences across slices, while representation learning compresses high-dimensional gene expression data into a shared latent space that supports robust downstream analysis. Despite their interconnected nature, existing methods typically address these tasks in isolation, limiting their ability to perform coherent, biologically meaningful integration.

Previous approaches fall into two broad categories. Alignment-based methods [36, 73, 12, 27, 20], such as PASTE [77], estimate spatial location - level mappings across adjacent slices by jointly considering spatial coordinates and gene expression similarity, enabling reconstruction of 3D tissue volumes. However, these methods operate directly on raw expression profiles, which are often sparse, noisy, and affected by batch effects. These batch effects refer to systematic technical variations introduced during sample processing or sequencing, which can obscure true biological signals. As a result, they do not provide low-dimensional representations that are essential to downstream tasks such as spatial domain detection or trajectory inference. Conversely, representation learning–based methods [22, 33, 80, 78, 51, 23, 7, 72, 76], such as GraphST [37] and STAGATE [15], leverage graph neural networks to extract informative low-dimensional embeddings. While effective for extracting latent embeddings and identifying spatial domains, these methods typically process slices independently or jointly without explicitly resolving anatomical correspondences. As a result, homologous tissue regions may be represented inconsistently across slices, undermining interpretability and cross-slice comparison. To address the need for joint analysis across samples, integration methods originally developed for single-cell RNA sequencing data [71, 18], including Harmony [32], Seurat [56], and scVI [39], do not account for spatial context and assume shared coordinate systems across samples, assumptions that rarely hold in spatial data. Recently, approaches such as STAligner [79] and PRECAST [35] have extended integration techniques to spatial transcriptomics, but they either require additional input (e.g., batch id or histology image) or do not explicitly model spatial location-level alignment. These methods focus on harmonizing latent features across slices but cannot account for physical tissue distortions that are critical for spatial reconstruction.

Together, these limitations highlight the need for a unified framework that can resolve spatial correspondences and learn biologically meaningful representations across multiple slices, allowing alignment to guide representation learning, and vice versa. To address this need, we introduce JADE (Joint Alignment and Deep Embedding), a computational framework that integrates multi-slice SRT data by simultaneously learning (1) a probabilistic alignment between spatial locations and (2) a shared low-dimensional embedding space. JADE performs alignment in the latent space via attention-based optimal transport and enforces spatial and transcriptomic consistency through graph-based contrastive learning. This coupling ensures that learned embeddings are mutually aligned and biologically coherent, while correspondences between spatial locations across tissue slices respect both spatial geometry and gene expression structure.

To the best of our knowledge, JADE is the first method to jointly optimize spatial location-level alignment and representation learning within a unified, spatially informed model. We benchmark JADE on multi-slice SRT datasets from the human dorsolateral prefrontal cortex (DLPFC) [44] and the regenerating axolotl brain [67], and show that it consistently outperforms state-of-the-art alignment and embedding methods in spatial clustering accuracy, alignment fidelity, and biological interpretability. JADE offers a scalable and robust solution for integrative spatial analysis, particularly in settings that require the joint resolution of anatomical correspondence and functional representation across complex tissue landscapes.

2 Methods

Problem Definition. Before we present our method JADE, we first formally introduce the problem setup. Consider a pair of SRT slices (S_1, X_1) and (S_2, X_2), where $S_1 \in \mathbb{R}^{n_1 \times 2}$ and $S_2 \in \mathbb{R}^{n_2 \times 2}$ represent the spatial coordinates of n_1 and n_2 spatial locations in the two tissue slices, and $X_1 \in \mathbb{R}^{n_1 \times p}$ and $X_2 \in \mathbb{R}^{n_2 \times p}$ correspond to their respective gene expression matrices, where p represents the same set of genes measured across tissue slices. Given both gene expression and spatial coordinates for the two slices, our objective is to jointly learn: (1) an alignment matrix $\Pi \in \mathbb{R}^{n_1 \times n_2}$ between

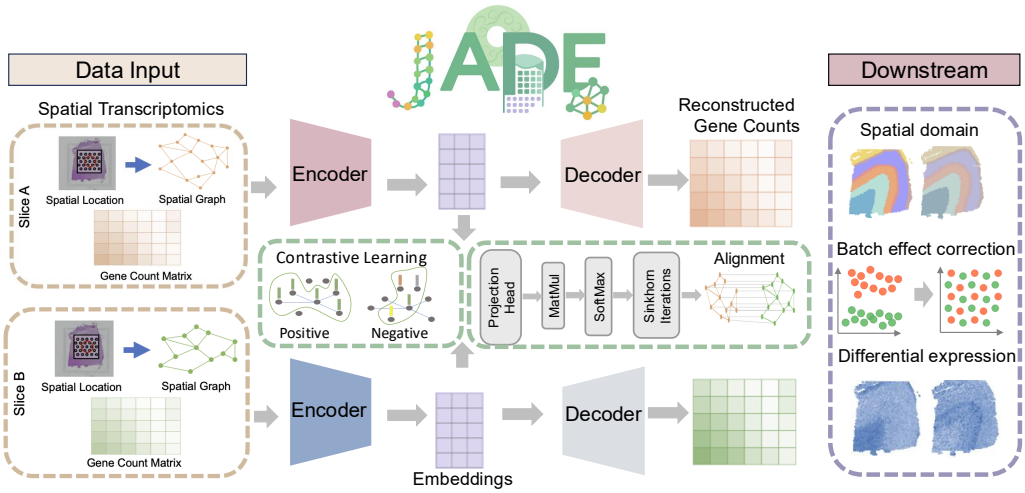

Figure 1: Workflow of JADE. Given two slices of SRT data (left box), JADE learns a probabilistic alignment and low-dimensional embeddings simultaneously. One unique feature of JADE is that embedding and alignment are updated in a roundtrip way in which each iteration alternates between alignment and embedding refinement. Downstream applications (right box) include spatial domain detection, batch effect correction, and differential expression analysis.

these spatial locations, where Π_{ij} encodes the correspondence between spatial location i in slice 1 and spatial location j in slice 2, and (2) the low-dimensional embedding representations H_1 and H_2.

Overview of JADE. Figure 1 illustrates the workflow of JADE. The pipeline begins with the input of SRT data from two slices (left panel). These inputs are passed through encoders to extract low-dimensional embeddings. Simultaneously, JADE infers cross-slice correspondences via a graph attention module to obtain embedding-space alignment. This module consists of a projection head, matrix multiplication, softmax, and iterative sinkhorn operations. To enhance generalizability, we employ a contrastive learning module that maximizes agreement between spatially neighboring locations while distinguishing dissimilar ones. A key innovation of JADE is its roundtrip learning scheme, in which embeddings and alignments are refined alternately within each training iteration. The outputs of JADE include a probabilistic alignment matrix between spatial locations across two tissue slices and embedding representation of each spot. These outputs support downstream analyses such as 3D tissue reconstruction (via the alignment matrix), spatial domain detection (via the learned embeddings), batch effect correction, and differential expression analysis. Detailed descriptions of each module are provided in the following sections and the pseudo code is shown in Appendix A. The implementation of JADE, along with data preprocessing scripts and pretrained models, is publicly available at https://github.com/YMa-lab/JADE.

2.1 Graph Autoencoder

Let $A_1 \in \mathbb{R}^{n_1 \times n_1}$, and $A_2 \in \mathbb{R}^{n_2 \times n_2}$ denote the two adjacency matrices of the spatial graph in two slices obtained using the method described in Appendix B. Given two SRT slices (A_1, X_1) and (A_2, X_2), we first encode them independently using graph convolutional networks (GCNs) [31] to obtain initial embeddings. Each encoder aggregates information from neighboring spatial locations to learn low-dimensional embeddings using a one-layer GCN: $H_1 = \text{Relu}(\tilde{A}_1 X_1 W_{1e} + b_{1e})$, $H_2 = \text{Relu}(\tilde{A}_2 X_2 W_{2e} + b_{2e})$, where $\tilde{A}_i = D_i^{-\frac{1}{2}} A_i D_i^{-\frac{1}{2}}$ is the degree normalized adjacency matrix, ensuring proper feature scaling across neighboring locations. Without normalization, a node with more neighbors would accumulate disproportionately large feature values, leading to biased representations. W_{ie} and b_{ie} are trainable weight and bias parameters of the encoder for slice $i \in \{1, 2\}$. ReLU is applied as a nonlinear activation function to enhance feature expressiveness. $H_1 \in \mathbb{R}^{n_1 \times d}$ and $H_2 \in \mathbb{R}^{n_2 \times d}$ are the latent embeddings of two slices, where d is a hyperparameter that defines the latent dimension.

To ensure that the learned embeddings retain key biological information, we decode them back into the original feature space using a GCN-based decoder: $\hat{X}_1 = \text{Relu}(\tilde{A}_1 H_1 W_{1d} + b_{1d})$, $\hat{X}_2 = \text{Relu}(\tilde{A}_2 H_2 W_{1d} + b_{1d})$, where W_{id} and b_{id} represent trainable parameters of the decoder for slice

$i \in \{1, 2\}$. The reconstruction loss ensures that the decoded outputs $\hat{X}_1, \hat{X}_2$ accurately recover the original gene expression profiles $\mathcal{L}_{\text{recon}} = \frac{1}{n_1}||X_1 - \hat{X}_1||_F^2 + \frac{1}{n_2}||X_2 - \hat{X}_2||_F^2$, where $||\cdot||_F$ denotes the Frobenius norm, measuring the difference between the reconstructed and original feature matrices.

2.2 Graph Attention for Embedding-Space Alignment

Given embeddings H_1, H_2, we obtain the alignment matrix Π through the following steps.

Step 1: Compute cross-attention weights. We first compute an attention-based similarity matrix $C \in [0, 1]^{n_1 \times n_2}$ to measure correspondences between embeddings of two slices:

$$C = \text{Softmax}\left(\frac{-H_1^T M M^T H_2}{\sqrt{d}}\right). \tag{1}$$

Here $M \in \mathbb{R}^{d \times d}$ is a learnable linear projection implementing an attention mechanism [63], enabling the model to focus on key cross-slice features.

Step 2: Obtain the alignment matrix from cross-attention weights. The cross-attention weights C encode pairwise embedding similarities, but it is only row-stochastic: each row sums to 1, yet the column totals are unrestricted. This imbalance can lead to biased mappings, where some target locations accumulate disproportionately high mass, while others receive very little.

To achieve a balanced alignment, we reinterpret the alignment problem within an *optimal transport* (OT) [61] framework, enforcing both rows and columns to satisfy uniform marginal constraints. In this framework, the alignment matrix satisfy $\Pi \in [0, 1]^{n_1 \times n_2}$, $\sum_{i,j} \Pi_{ij} = 1$. Each element Π_{ij} denotes the amount of probability mass transported from source point i to target point j, with the constraint that the entire matrix satisfies specific marginal distributions. Specifically, we apply the Sinkhorn-Knopp algorithm [54, 13] to transform the raw similarity scores in C into a doubly-stochastic alignment matrix Π. This iterative procedure alternately normalizes rows and columns, enforcing each row sum to $1/n_1$ and each column sum to n_2, ensuring that every spatial location contributes and receives an equal amount of alignment mass. Thus, the alignment becomes fair and balanced, avoiding situations where a few locations dominate the mappings. Additionally, we introduce a marginal regularization term into our objective function: $\mathcal{L}_{\text{marginal}} = \text{KL}\left(\frac{\mathbf{1}_n}{n_1} \cdot \Pi \,||\, \frac{\mathbf{1}_{n_2}}{n_2}\right)$. This penalty is defined as the KL divergence between the column sums of Π and the desired uniform marginal distribution. In doing so, we explicitly discourage deviations from uniformity, further reinforcing balanced alignments.

Step 3: Spatial and embedding aware alignment losses. To refine the alignment and maintain spatial structure, we define two primary losses motivated by the fused Gromov-Wasserstein distance [45, 64]: (1) Spatial structure preservation loss (Mis-Maintain Loss) ensures that spatial relationships in both slices remain consistent after alignment $\mathcal{L}_{\text{maintain}} = \frac{1}{n_1}||D_1 - n_2^2 \Pi D_2 \Pi^T||_F + \frac{1}{n_2}||D_2 - n_1^2 \Pi^T D_1 \Pi||_F$, where D_1 and D_2 are pairwise spatial distance matrices. Each entry d_{1ij} (for slice 1) and d_{2ij} (for slice 2) represents the squared Euclidean distance between spatial locations i and j. The scaling factors n_2^2 and n_1^2 ensure that the transported distance matrices $\Pi D_2 \Pi^T$ and $\Pi^T D_1 \Pi$ remain on the same scale as the original D_1 and D_2, compensating for the normalization inherent in Π. Minimizing $\mathcal{L}_{\text{maintain}}$ encourages spatial locations that were close to each other before alignment to remain close after alignment, while spatial locations that were far apart should continue to be far apart. (2) Embedding alignment loss (Mis-Alignment Loss) ensures that the aligned embeddings are close to each other, maintaining meaningful biological correspondence $\mathcal{L}_{\text{align}} = \frac{1}{n_1}||H_1 - n_2 \Pi H_2||_F + \frac{1}{n_2}||H_2 - n_1 \Pi^T H_1||_F$. Similarly, the scaling factors n_2 and n_1 ensure that the transported embedding matrices $n_2 \Pi H_2$ and $n_1 \Pi^T H_1$ remain on the same scale as the original H_1 and H_2. The first term in the above equation minimizes the discrepancy between slice 1 embeddings and their aligned counterparts from slice 2. The second term enforces the same alignment constraint in the reverse direction.

2.3 Self-supervised Graph Contrastive Learning

We refine our embeddings via self-supervised contrastive learning similar to GraphST [37], using the Graph Infomax objective [65] to maximize mutual information between each spot and its local neighborhood. This drives adjacent, biologically similar spots closer in latent space while pushing structurally distinct or distant spots apart.

For each spatial location i, we define its local spatial proxy r_i as: $r_i = \frac{1}{|N(i)|}\sum_{j\in N(i)} h_j$, where $N(i)$ denotes the set of immediate neighbors of spatial location i, h_j is the embedding of neighboring spatial location j. The pair (h_i, r_i) forms a positive pair, reflecting spatial proximity and biological similarity. To create negative examples, we generate a shuffled embedding matrix H' by randomly permuting the rows of H, destroying spatial coherence. For each i, let h'_i be the shuffled embedding and r'_i its corresponding shuffled proxy, forming a negative pair (h'_i, r'_i). Finally, we train a discriminator Φ via binary cross-entropy loss to distinguish these positive from negative pairs. The slice-specific contrastive loss is: $\mathcal{L}^k_{\text{SCL}} = -\frac{1}{n_k}\sum_{i=1}^{n_k}\Big[\log\Phi(h_{ki}, r_{ki}) + \log\big(1-\Phi(h'_{ki}, r'_{ki})\big)\Big]$ for $k = 1, 2$. The overall contrastive loss is $\mathcal{L}_{\text{SCL}} = \mathcal{L}^1_{\text{SCL}} + \mathcal{L}^2_{\text{SCL}}$.

2.4 Final Loss Function

In summary, the training objective of JADE comprises three components:

$$\mathcal{L}_{\text{SCL}} + \lambda_2\mathcal{L}_{\text{recon}} + \lambda_3\mathcal{L}_{\text{maintain}} + \lambda_4\mathcal{L}_{\text{align}} + \lambda_5\mathcal{L}_{\text{marginal}}.$$

Throughout the paper, we set $\lambda_2 = 10$, $\lambda_4 = 0.1$, $\lambda_5 = 1$. We use a data-driven method to select λ_3 from 0.2 to 2.0 depending on a calculated similarity score between slices. If two slices are similar, we use a larger λ_3 encourage information sharing; otherwise, we use a smaller λ_3 to prevent negative transfer. A comprehensive description of hyperparameter tuning procedures is provided in Appendix E. Furthermore, the contribution of each individual component in JADE is evaluated through systematic ablation studies, as detailed in Appendix F. Results show that performance deteriorates markedly when any loss term is omitted, demonstrating that both the misalignment and mismaintain losses are crucial for optimal embedding and alignment quality.

Downstream analysis. After training, we obtained two sets of low-dimensional embeddings. We normalized the length of each embedding vector and applied the `mclust` algorithm independently, specifying the cluster number based on prior knowledge. Following [79, 37], we set the cluster number from 5 through 7 for the DLPFC dataset, and following [20], from 16 through 17 for the axolotl brain dataset, with each cluster corresponding to a distinct cell type. We then leveraged the inferred domains to conduct differential expression analysis. For the batch-effect correction evaluation, we projected the embeddings into a 2-dimensional space using UMAP, visualized their spatial distribution, and computed the local Simpson diversity index to quantify how well the embeddings from different slices are intermingled. To assess alignment quality, we visualized and quantified the highest-probability correspondences in the alignment matrix within each ground-truth domain, explicitly marking correct and incorrect matches, and computed accuracy scores to measure alignment performance.

3 Fast Computation for JADE

A major challenge for SRT alignment methods is their high computational cost, scaling quadratically with the number of spatial locations. Methods such as PASTE [77] and our JADE algorithm require computing similarity matrices between all location pairs, resulting in prohibitive runtime and memory usage for high-resolution datasets. The primary computational bottleneck in JADE is the cross-attention step, which calculates pairwise similarities and scales as $O(n_1 n_2 d)$. To accelerate this, we introduce an approach that aligns at a coarser resolution using aggregated spatial units ("hyperspots"), significantly reducing computational complexity. Despite this approximation, Fast-JADE achieves performance comparable to JADE on the DLPFC dataset (see Appendix G for detailed comparison).

Hyperspot embedding construction. For each tissue slice, we apply K-means clustering to group spatial locations into fewer, coarser spatial units, resulting $m_1 \ll n_1$ and $m_2 \ll n_2$ hyperspots respectively. The number of hyperspots is set to approximately 10-20% of the original number of spatial locations for each slice. Each hyperspot embedding is computed by averaging embeddings of the spatial locations within the cluster, resulting in compact representations.

Compute cross-attention between hyperspots. We compute the cross-attention weights and alignment matrix at the hyperspot level. Specifically, $C^{\text{hyper}} = \text{Softmax}\left(-H_1^{\text{hyper},T} M M^T H_2^{\text{hyper}}/\sqrt{d}\right)$ is similar to (1). Using these hyperspot-level attention weights, we then compute the alignment and maintenance losses, $\mathcal{L}_{\text{maintain}}$ and $\mathcal{L}_{\text{align}}$ at the hyperspot level.

From hyperspot-level to spatial location-level Alignment. After training, we transfer the learned alignment back to the full-resolution space. Specifically, we reuse the same projection head M to com-

pute the fine-grained spatial location-wise attention weights: $C = \text{Softmax}\left(-H_1^T M M^T H_2/\sqrt{d}\right)$, where $H_1 \in \mathbb{R}^{n_1 \times d}$ and $H_2 \in \mathbb{R}^{n_2 \times d}$ are the original spatial location embeddings. This ensures consistency between coarse and fine levels and avoids retraining at full resolution. The resulting matrix C is then passed through the Sinkhorn-Knopp algorithm to obtain an alignment matrix Π with uniform marginals.

4 Experiments

We benchmarked JADE against six representative baseline models, selected for their relevance to either spatial alignment or embedding tasks in spatial transcriptomics. For alignment accuracy, we compared with PASTE [77], Seurat [56], and STAligner [79], three widely used methods that align SRT slices based on transcriptomic similarity and spatial proximity. For evaluating the learned embeddings through spatial domain detection, we included GraphST [74], STAGATE [15], and STAligner, which incorporate spatial or graph-based information to enhance domain identification.

To provide a unified assessment, we used a comprehensive alignment metric that accounts for both correctly aligned and unaligned locations (Appendix B). Embedding quality was quantified using the Adjusted Rand Index (ARI) for biological domain recovery, the integrated local inverse Simpson's index (iLISI) [53] for cross-slice integration, and UMAP visualizations for qualitative evaluation of coherence and batch correction. For biological interpretability, we conducted differential expression analysis to identify domain-specific marker genes and validated them against known anatomical annotations and literature, confirming that the learned embeddings capture biologically meaningful spatial structures and gene associations.

We also tested JADE on two additional datasets in Appendix H: the MERFISH dataset [10] and the Breast-Cancer Visium/Xenium dataset [25], previously used in SLAT [69]. As summarized in Table 11, JADE consistently outperforms or matches state-of-the-art baselines, including SLAT, STAligner and SpotScape[49], in both domain-detection accuracy (ARI) and alignment accuracy (ACC). These results demonstrate JADE's robustness and adaptability across distinct spatial transcriptomics platforms (MERFISH, Visium, Xenium) and tissue types (brain and breast).

4.1 Joint Spatial Alignment and Representation Learning Across Human DLPFC Slices

Dataset description. We evaluated our proposed method on the Human Dorsolateral Prefrontal Cortex (DLPFC) dataset generated with the 10x Visium platform [55], a standard benchmark in spatial transcriptomics. The dataset comprises 12 serial tissue sections, including four sequential slices (A–D) from each of three donors (I–III), with expression profiles of 33,538 genes measured at 47,681 spatial spots. Within each individual, slices A-B and C-D are adjacent pairs separated by 10 μm, while slices B and C have a larger separation of 300 μm. Differences across individuals are primarily attributed to batch effects arising from technical variations in sample processing, sequencing, or tissue handling. Each slice was annotated with seven spatial domains, corresponding to six cortical layers and the white matter [44]. These expert-provided annotations served as the ground truth for spatial domain detection analysis. Following preprocessing in Appendix B, the two selected slices retained about 1500 shared highly variable genes for each pair.

Improved clustering accuracy. Figure 2(A) provides a visual comparison of the predicted spatial domains from each method against the ground truth for both Slice A and Slice B of Sample III. Visually, JADE demonstrates the highest accuracy in recovering the annotated cortical structure, across slices A and B with clear, smooth boundaries. In contrast, other methods yield noisier and less coherent spatial domains. Quantitatively, JADE achieves median ARI values of approximately 0.61 and 0.65 for Slices A and B, for Sample III as shown in Figure 2(B). This represents a statistically significant improvement over GraphST (0.59 and 0.58), STAligner (0.56 and 0.59), and STAGATE (0.52 and 0.53). The top row of Figure 2(C) further visualizes the learned low-dimensional embeddings produced by each method using UMAP, colored by their true biological cluster annotation. JADE consistently demonstrates superior performance, yielding highly distinct and well-separated biological layers that form cohesive clusters with minimal overlap. Appendix C shows additional results for other samples and other adjacent slices.

JADE effectively mitigates batch effects in multi-slice integration. The bottom row of Figure 2(C) illustrates the effectiveness of batch effect removal across different methods by visualizing UMAP embeddings colored by slice identity (slice A: blue, slice B: red). Effective integration is reflected by

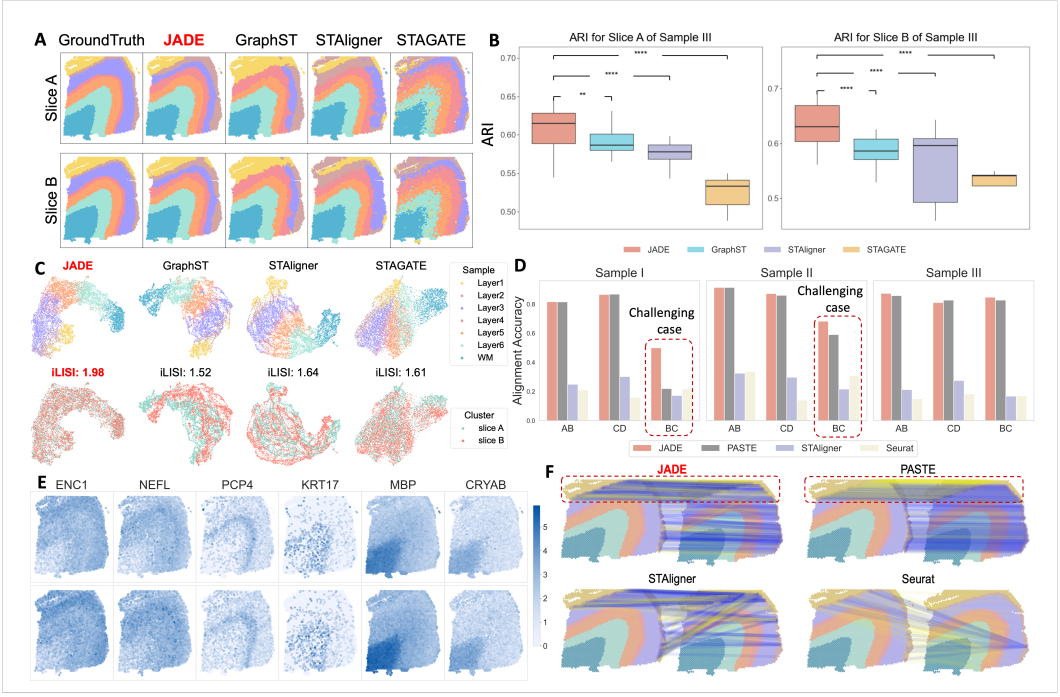

Figure 2: (A) Predicted spatial domains for slices A and B (Sample III) using five methods. (B) ARI for slices A and B of Sample III in DLPFC. p-values are calculated by Wilcoxon-rank sum test, where ** $p < 1\%$, ***$p < 0.1\%$ and **** $p < 0.01\%$ (C) UMAP of embeddings colored by predicted clusters (top) and slice (bottom). (D) Alignment accuracy across adjacent slice pairs (AB, CD, BC). (E) Spatial expression of marker genes confirms biologically relevant domain structures. (F) Visual comparison of alignment accuracy for Layer 1 in Sample III (Slice A vs. Slice B). True alignments are represented in blue; wrong ones are shown in yellow.

the intermixing of spatial locations from different slices, indicating successful alignment in a shared latent space. JADE achieves the highest inter-slice mixing, suggesting excellent batch correction performance. In contrast, GraphST, STAligner and STAGATE display limited alignment: the two slices remain largely separated within the UMAP space, suggesting poor batch removal efficacy. JADE also attained the highest iLISI score of 1.98, outperforming GraphST (1.52), STAligner (1.64), and STAGATE (1.61), confirming its superior performance in integrating SRT data across slices while correcting for batch effects.

Superior alignment performance. Figure 2(D) presents a comprehensive quantitative assessment of alignment accuracy across adjacent slice pairs (AB, BC, CD) for three distinct samples, comparing JADE against existing methods including PASTE, STAligner, and Seurat. Among these, the B–C pairs represent the most challenging alignment scenario due to their larger inter-slice distance ($300\mu m$), compared to the $10\mu m$ separation in A–B and C–D pairs. Across all samples and slice combinations, JADE consistently demonstrates superior alignment accuracy, particularly in challenging scenarios. While PASTE achieves comparable performance to JADE in easier settings (AB and CD pairs), it exhibits substantial performance degradation when confronted with the more challenging BC alignments. Notably, for BC pairs in samples I and II, JADE surpasses PASTE by margins exceeding 50%. STAligner shows moderate accuracy but is consistently lower than JADE and PASTE across all slice pairs. Seurat performs the poorest in all cases, with particularly low alignment accuracy in the BC pairs. These results highlight JADE's superior alignment performance, particularly in challenging integration settings. In Figure 2(F), we visualize Layer 1 alignment between Slice A and Slice B in Sample III across four methods. Blue lines indicate correct matches within the same annotated layer, while yellow lines indicate errors. JADE achieves the best outcome, with mostly correct (blue) alignments and very few incorrect (yellow) ones.

Domain-specific differential expression gene analysis. We performed domain-specific differential expression (DE) analysis based on the spatial domains detected by JADE and identified six representative marker genes (*ENC1*, *NEFL*, *PCP4*, *KRT17*, *MBP*, and *CRYAB*). These genes were selected for their strong domain-level enrichment and subsequently validated through literature for their well-established region-specific expression patterns in the human cortex. Figure 2(E) displays the

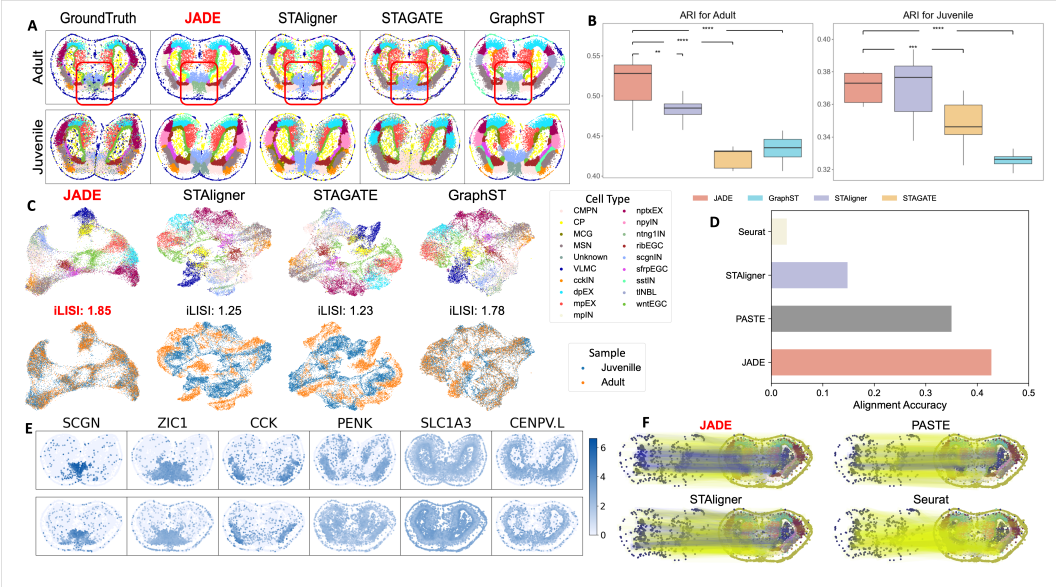

Figure 3: (A) Domain segmentation results for Juvenile and Adult slices comparing ground truth against four methods. (B) ARI scores for four methods. p-values are calculated the same as in Figure 2. (C) UMAP of embeddings colored by predicted clusters (top) and slice (bottom). (D) Alignment accuracy for Juvenile and Adult slices. (E) Spatial expression of marker genes confirms biologically relevant domain structures. (F) Visual comparison of alignment accuracy for CP cell type between the juvenile and the adult. True alignments are represented by blue lines, while wrong alignments are shown in yellow.

spatial expression patterns of these genes across two adjacent DLPFC tissue slices (top and bottom rows). *ENC1*, *NEFL*, and *PCP4* exhibit clear laminar structures consistent with neuronal populations localized to middle and deep cortical layers [44, 68, 17]. Their spatial localization is preserved across slices, demonstrating coherent biological structure. *KRT17*, typically associated with epithelial-like or glial cells, shows more punctate and scattered expression, particularly in the upper portions of the slices [11, 44]. *MBP*, a myelination-related gene, is strongly expressed in the lower regions of both slices, likely corresponding to white matter (WM), and the spatial coherence across slices further supports successful alignment [48, 44, 24]. *CRYAB*, a stress-response and astrocyte-associated marker, exhibits a broader expression domain with intermediate intensity [4].

Beyond expression-based baselines, we further evaluated JADE against GPSA [27], an image-informed alignment model that integrates histological features in Appendix H. As summarized in Table 10, JADE achieves higher alignment accuracy across all DLPFC slice pairs while maintaining comparable clustering quality, indicating that reliable alignment can be achieved without reliance on histology images.

4.2 Joint Learning of Axolotl Brain Dataset Across Different Developmental Stages

We applied our method to the task of jointly learning alignments and embeddings of SRT data across distinct developmental stages. This integration is crucial for gaining insights into the dynamic processes of cell proliferation and differentiation. However, achieving such comprehensive integration is inherently challenging due to technical batch effects and fundamental biological shifts, both of which significantly complicate the establishment of accurate underlying alignment.

Dataset description. We evaluated JADE on a SRT dataset of the axolotl telencephalon (a region of the brain) generated using the Stereo-seq platform [9]. This dataset captures gene expression across five developmental stages: three embryonic stages, the juvenile stage, and the adult stage. In this study, we focused our analysis on the last two, which together span a broader range of mature brain architecture. The juvenile stage and adult stage have 17 and 16 distinct cell types respectively. Among them, fourteen are shared between both stages, while the rests from the juvenile stage differentiate or transition into new cell types in the adult stage. Following preprocessing, the two selected slices retained 1,000 shared highly variable genes, with 11,698 and 8,243 spatial locations, respectively. To reduce computational time, we applied the accelerated JADE method introduced in Section 3, using 1,000 hyperspots per slice, corresponding to about 10% of the spatial locations.

Improved clustering accuracy. Figure 3(A) provides a visual comparison of the domains detection results of JADE and baseline methods in Adult (top row) and Juvenile (bottom row) slices. Compared to the ground truth, JADE consistently recovers structurally coherent and anatomically accurate domains in both stages. Notably, in the adult brain, JADE is the only method that successfully reconstructs the blue peripheral ring cluster, representing vascular leptomeningeal cells (VLMC) [62]. This region is either partially fragmented (in GraphST), blurred and mixed with neighboring domains (in STAGATE, STAligner). In addition, JADE accurately produces a coherent, bilaterally symmetric red region in the telencephalon core, while STAGATE yields considerable color mixing within this region, STAligner only partially recovers it, and GraphST produces an overexpanded red region, suggestive of excessive smoothing. These qualitative patterns are supported by the quantitative results in Figure 3(B). For the Adult stage, JADE achieves the highest median ARI score (approximately 0.53), significantly outperforming STAligner (0.48), STAGATE (0.42), and GraphST (0.43). For the Juvenile stage, JADE also obtains a high median ARI (approximately 0.37), demonstrating statistically significant improvement over STAGATE (0.35) and GraphST (0.32). While JADE's performance for the Juvenile stage is quantitatively comparable to STAligner's by ARI, its qualitative domain recovery remains notably more coherent and anatomically accurate as shown in Figure 3(A). The top row of Figure 3(C) further visualizes the UMAP for the embedding features, colored by the true cluster annotation. JADE produces well-separated and compact clusters, indicating a clear differentiation of spatial domains in the latent space, while other methods exhibit considerable mixing of domain colors, reflecting ambiguity in domain boundaries.

JADE effectively mitigates batch effects in multi-slice integration. The bottom row of Figure 3(C) compares methods via UMAP embeddings colored by slice (Adult: red, Juvenile: blue). JADE shows near-perfect mixing within structured clusters, indicating superior batch-effect removal. In contrast, STAligner, STAGATE, and GraphST display notable batch-induced separation. Quantitatively, JADE achieves the highest LISI score (1.85), surpassing GraphST (1.78), STAligner (1.25), and STAGATE (1.23), highlighting its effectiveness in integrating data across developmental stages.

Superior alignment performance. As shown in the bar plot in Figure 3(D), JADE achieves the highest alignment accuracy among all methods, substantially outperforming PASTE, STAligner, and Seurat. This result underscores JADE's ability to establish precise correspondences between spatial locations across slices, even under the challenging setting of cross-developmental stage alignment, where substantial morphological and transcriptional variation exists. Figure 3(F) further visualizes the alignment accuracy for cortical plate cell type between the Juvenile slice and the Adult slice, to support this finding. JADE yields the highest proportion of correct (blue) alignments, demonstrating its superior ability to preserve anatomical consistency during cross-slice registration.

Domain-specific gene. Figure 3(E) displays the spatial expression patterns of six canonical marker genes, *SCGN*, *ZIC1*, *CCK*, *PENK*, *SLC1A3*, *CENPV.L*, across Juvenile and Adult slices (top and bottom rows). These genes are known to exhibit region-specific expression within the axolotl brain and thus serve as internal benchmarks for spatial alignment and biological interpretability. Specifically, *SCGN* and *ZIC1*, for example, are associated with distinct neuronal populations and developmental patterning, and their laminar or domain-restricted expression patterns are preserved across developmental stages, reflecting coherent spatial organization [2, 50, 58]. *CCK* and *PENK*, which are neuropeptide-related genes, show distinct regional enrichment, highlighting the emergence of functional specialization in the maturing brain [29, 8]. *SLC1A3*, a glutamate transporter gene involved in astrocytic function, exhibits broad but domain-enriched expression, consistent with known glial distribution [3]. *CENPV.L*, associated with nuclear and centrosomal processes, displays a sharply localized expression pattern, providing a clear contrast across domains [57]. Together, these genes illustrate the biological relevance of the spatial domains identified by JADE, demonstrating its ability to uncover consistent, developmentally regulated gene expression patterns across stages of brain maturation.

5 Conclusions

In this paper, we address the critical yet understudied challenge of joint spatial alignment and representation learning across multi-slice SRT data. We propose JADE, a unified framework that simultaneously infers spatial correspondences and learns biologically meaningful low-dimensional embeddings. Through comprehensive evaluations on human DLPFC and axolotl brain datasets, JADE demonstrates superior performance in spatial domain detection, alignment accuracy, and batch effect correction compared to state-of-the-art methods. This study provides a robust computational tool

for integrating multi-slice SRT data, with the potential to advance 3D tissue reconstruction, cross-condition comparison, and spatially informed transcriptomic discovery. However, several limitations remain. First, while JADE performs well across a range of spatial transcriptomics platforms and tissue types, its performance may be affected by extreme sparsity, highly unbalanced slice resolutions, or limited overlap in tissue regions across slices. Second, JADE currently models pairwise slice alignment, which may limit its ability to fully reconstruct larger tissue volumes or resolve global correspondences across long serial sections. Third, the current framework assumes a static snapshot of spatial organization and does not account for temporal variation, which is increasingly relevant in developmental or regenerative contexts. Future work will extend JADE to handle fully joint alignment across multiple slices, incorporate spatiotemporal transcriptomics data, and integrate additional data modalities such as histological imaging or spatial epigenomics. We also aim to improve scalability for ultra-high-resolution datasets and explore methods to increase robustness under technical noise or sample heterogeneity. To the best of our knowledge, no potential negative impacts resulting from our work have been identified.

6 Acknowledgment

This work was supported by the National Science Foundation (NSF) and the National Institute of General Medical Sciences (NIGMS) under award number R01GM152814, by the National Institutes of Health (NIH) under award number R35GM160372, and by the NSF under award numbers DBI-2526948 and IIS-2500960.

References

[1] Rohit Arora, Christian Cao, Mehul Kumar, Sarthak Sinha, Ayan Chanda, Reid McNeil, Divya Samuel, Rahul K Arora, T Wayne Matthews, Shamir Chandarana, et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nature communications*, 14(1):5029, 2023.

[2] Jun Aruga, Takashi Inoue, Jun Hoshino, and Katsuhiko Mikoshiba. Zic2 controls cerebellar development in cooperation with zic1. *Journal of Neuroscience*, 22(1):218–225, 2002.

[3] Mykhailo Y Batiuk, Araks Martirosyan, Jérôme Wahis, Filip de Vin, Catherine Marneffe, Carola Kusserow, Jordan Koeppen, João Filipe Viana, João Filipe Oliveira, Thierry Voet, et al. Identification of region-specific astrocyte subtypes at single cell resolution. *Nature communications*, 11(1):1220, 2020.

[4] Lina Vanessa Becerra-Hernández, Martha Isabel Escobar-Betancourt, Hernán José Pimienta-Jiménez, and Efraín Buriticá. Crystallin alpha-b overexpression as a possible marker of reactive astrogliosis in human cerebral contusions. *Frontiers in Cellular Neuroscience*, 16:838551, 2022.

[5] Emelie Berglund, Jonas Maaskola, Niklas Schultz, Stefanie Friedrich, Maja Marklund, Joseph Bergenstråhle, Firas Tarish, Anna Tanoglidi, Sanja Vickovic, Ludvig Larsson, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature communications*, 9(1):2419, 2018.

[6] Darren J Burgess. Spatial transcriptomics coming of age. *Nature Reviews Genetics*, 20(6): 317–317, 2019.

[7] Jiazhang Cai, Huimin Cheng, Shushan Wu, Wenxuan Zhong, Guo-Cheng Yuan, and Ping Ma. West is an ensemble method for spatial transcriptomics analysis. *Cell Reports Methods*, 4(11), 2024.

[8] Daniela Calvigioni, Zoltán Máté, János Fuzik, Fatima Girach, Ming-Dong Zhang, Andrea Varro, Johannes Beiersdorf, Christian Schwindling, Yuchio Yanagawa, Graham J Dockray, et al. Functional differentiation of cholecystokinin-containing interneurons destined for the cerebral cortex. *Cerebral cortex*, 27(4):2453–2468, 2017.

[9] Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

[10] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[11] Shuo Chen, Yuzhou Chang, Liangping Li, Diana Acosta, Yang Li, Qi Guo, Cankun Wang, Emir Turkes, Cody Morrison, Dominic Julian, et al. Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in alzheimer's disease. *Acta neuropathologica communications*, 10(1):188, 2022.

[12] K Clifton, M Anant, G Aihara, L Atta, OK Aimiuwu, JM Kebschull, MI Miller, D Tward, and J Fan. Stalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. nat. commun. 14, 8123, 2023.

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[14] Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Gromov-wasserstein optimal transport to align single-cell multi-omics data. *BioRxiv*, pages 2020–04, 2020.

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

[16] Yanhua Du, Jintong Shi, Jiaxin Wang, Zhenzhen Xun, Zhuo Yu, Hongxiang Sun, Rujuan Bao, Junke Zheng, Zhigang Li, and Youqiong Ye. Integration of pan-cancer single-cell and spatial transcriptomics reveals stromal cell features and therapeutic targets in tumor microenvironment. *Cancer Research*, 84(2):192–210, 2024.

[17] Yun Gong, Mohammad Haeri, Xiao Zhang, Yisu Li, Anqi Liu, Di Wu, Qilei Zhang, S Michal Jazwinski, Xiang Zhou, Xiaoying Wang, et al. Stereo-seq of the prefrontal cortex in aging and alzheimer's disease. *Nature Communications*, 16(1):482, 2025.

[18] Tiantian Guo, Zhiyuan Yuan, Yan Pan, Jiakang Wang, Fengling Chen, Michael Q Zhang, and Xiangyu Li. Spiral: integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biology*, 24(1):241, 2023.

[19] Laleh Haghverdi, Aaron TL Lun, Michael D Morgan, and John C Marioni. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nature biotechnology*, 36(5):421–427, 2018.

[20] Peter Halmos, Xinhao Liu, Julian Gold, Feng Chen, Li Ding, and Benjamin J Raphael. Dest-ot: Alignment of spatiotemporal transcriptomics data. *Cell Systems*, 16(2), 2025.

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

[22] Jian Hu, Xiangjie Li, Kyle Coleman, Amelia Schroeder, Nan Ma, David J Irwin, Edward B Lee, Russell T Shinohara, and Mingyao Li. Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods*, 18(11):1342–1351, 2021.

[23] Yunfei Hu, Yuying Zhao, Curtis T Schunk, Yingxiang Ma, Tyler Derr, and Xin Maizie Zhou. Adept: Autoencoder with differentially expressed genes and imputation for robust spatial transcriptomics clustering. *Iscience*, 26(6), 2023.

[24] Louise Huuki-Myers, Abby Spangler, Nick Eagles, Kelsey D Montgomery, Sang Ho Kwon, Boyi Guo, Melissa Grant-Peters, Heena R Divecha, Madhavi Tippani, Chaichontat Sriworarat, et al. Integrated single cell and unsupervised spatial transcriptomic analysis defines molecular anatomy of the human dorsolateral prefrontal cortex. *BioRxiv*, 2023.

[25] Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature communications*, 14(1):8353, 2023.

[26] 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, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *cell*, 182 (2):497–514, 2020.

[27] Andrew Jones, F William Townes, Didong Li, and Barbara E Engelhardt. Alignment of spatial genomics data using deep gaussian processes. *Nature methods*, 20(9):1379–1387, 2023.

[28] Namyoung Jung and Tae-Kyung Kim. Spatial transcriptomics in neuroscience. *Experimental & molecular medicine*, 55(10):2105–2115, 2023.

[29] L Just, C Olenik, B Heimrich, and DK Meyer. Glutamatergic control of the expression of the proenkephalin gene in rat frontoparietal cortical slice cultures. *Cerebral cortex (New York, NY: 1991)*, 8(8):702–709, 1998.

[30] Joyce B Kang, Aparna Nathan, Kathryn Weinand, Fan Zhang, Nghia Millard, Laurie Rumker, D Branch Moody, Ilya Korsunsky, and Soumya Raychaudhuri. Efficient and precise single-cell reference atlas mapping with symphony. *Nature communications*, 12(1):5890, 2021.

[31] TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[32] 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, 2019.

[33] 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, 2022.

[34] Senlin Lin, Zhikang Wang, Yan Cui, Qi Zou, Chuangyi Han, Rui Yan, Zhidong Yang, Wei Zhang, Rui Gao, Jiangning Song, et al. Bridging the dimensional gap from planar spatial transcriptomics to 3d cell atlases. *Nature Methods*, pages 1–13, 2025.

[35] Wei Liu, Xu Liao, Ziye Luo, Yi Yang, Mai Chan Lau, Yuling Jiao, Xingjie Shi, Weiwei Zhai, Hongkai Ji, Joe Yeong, et al. Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with precast. *Nature communications*, 14(1):296, 2023.

[36] Xinhao Liu, Ron Zeira, and Benjamin J Raphael. Partial alignment of multislice spatially resolved transcriptomics data. *Genome Research*, 33(7):1124–1132, 2023.

[37] Yahui Long, Kok Siong Ang, Mengwei Li, Kian Long Kelvin Chong, Raman Sethi, Chengwei Zhong, Hang Xu, Zhiwei Ong, Karishma Sachaphibulkij, Ao Chen, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications*, 14(1):1155, 2023.

[38] Sophia K Longo, Margaret G Guo, Andrew L Ji, and Paul A Khavari. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics*, 22(10):627–644, 2021.

[39] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

[40] Mohammad Lotfollahi, Mohsen Naghipourfar, Malte D Luecken, Matin Khajavi, Maren Büttner, Marco Wagenstetter, Žiga Avsec, Adam Gayoso, Nir Yosef, Marta Interlandi, et al. Mapping single-cell data to reference atlases by transfer learning. *Nature biotechnology*, 40(1):121–130, 2022.

[41] Igor Mandric, Brian L Hill, Malika K Freund, Michael Thompson, and Eran Halperin. Batman: fast and accurate integration of single-cell rna-seq datasets via minimum-weight matching. *IScience*, 23(6), 2020.

[42] Silas Maniatis, Tarmo Äijö, Sanja Vickovic, Catherine Braine, Kristy Kang, Annelie Mollbrink, Delphine Fagegaltier, Žaneta Andrusivová, Sami Saarenpää, Gonzalo Saiz-Castro, et al. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science*, 364 (6435):89–93, 2019.

[43] Vivien Marx. Method of the year: spatially resolved transcriptomics. *Nature methods*, 18(1): 9–14, 2021.

[44] 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, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24(3):425–436, 2021.

[45] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.

[46] Reuben Moncada, Dalia Barkley, Florian Wagner, Marta Chiodin, Joseph C Devlin, Maayan Baron, Cristina H Hajdu, Diane M Simeone, and Itai Yanai. Integrating microarray-based spatial transcriptomics and single-cell rna-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature biotechnology*, 38(3):333–342, 2020.

[47] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5): 534–546, 2022.

[48] William T Norton. Isolation and characterization of myelin. In *Myelin*, pages 161–199. Springer, 1984.

[49] Yunhak Oh, Junseok Lee, Yeongmin Kim, Sangwoo Seo, Namkyeong Lee, and Chanyoung Park. Global context-aware representation learning for spatially resolved transcriptomics. *arXiv preprint arXiv:2506.15698*, 2025.

[50] Chandrasekhar S Raju, Julien Spatazza, Amelia Stanco, Phillip Larimer, Shawn F Sorrells, Kevin W Kelley, Cory R Nicholas, Mercedes F Paredes, Jan H Lui, Andrea R Hasenstaub, et al. Secretagogin is expressed by developing neocortical gabaergic neurons in humans but not mice and increases neurite arbor size and complexity. *Cerebral Cortex*, 28(6):1946–1958, 2018.

[51] Honglei Ren, Benjamin L Walker, Zixuan Cang, and Qing Nie. Identifying multicellular spatiotemporal organization of cells with spaceflow. *Nature communications*, 13(1):4076, 2022.

[52] Hailing Shi, Yichun He, Yiming Zhou, Jiahao Huang, Kamal Maher, Brandon Wang, Zefang Tang, Shuchen Luo, Peng Tan, Morgan Wu, et al. Spatial atlas of the mouse central nervous system at molecular resolution. *Nature*, 622(7983):552–561, 2023.

[53] Edward H Simpson. Measurement of diversity. *nature*, 163(4148):688–688, 1949.

[54] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[55] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[56] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *cell*, 177(7):1888–1902, 2019.

[57] Ana Mafalda Baptista Tadeu, Susana Ribeiro, Josiah Johnston, Ilya Goldberg, Dietlind Gerloff, and William C Earnshaw. Cenp-v is required for centromere organization, chromosome alignment and cytokinesis. *The EMBO journal*, 27(19):2510–2522, 2008.

[58] Silvia Tapia-González and Javier DeFelipe. Secretagogin as a marker to distinguish between different neuron types in human frontal and temporal cortex. *Frontiers in Neuroanatomy*, 17: 1210502, 2023.

[59] Kim Thrane, Hanna Eriksson, Jonas Maaskola, Johan Hansson, and Joakim Lundeberg. Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage iii cutaneous malignant melanoma. *Cancer research*, 78(20):5970–5979, 2018.

[60] Luyi Tian, Fei Chen, and Evan Z Macosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782, 2023.

[61] Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

[62] Michael Vanlandewijck, Liqun He, Maarja Andaloussi Mäe, Johanna Andrae, Koji Ando, Francesca Del Gaudio, Khayrun Nahar, Thibaud Lebouvier, Bàrbara Laviña, Leonor Gouveia, et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature*, 554(7693): 475–480, 2018.

[63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[64] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.

[65] Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. *arXiv preprint arXiv:1809.10341*, 2018.

[66] Britta Velten, Jana M Braunger, Ricard Argelaguet, Damien Arnol, Jakob Wirbel, Danila Bredikhin, Georg Zeller, and Oliver Stegle. Identifying temporal and spatial patterns of variation from multimodal data using mefisto. *Nature methods*, 19(2):179–186, 2022.

[67] Xiaoyu Wei, Sulei Fu, Hanbo Li, Yang Liu, Shuai Wang, Weimin Feng, Yunzhi Yang, Xiawei Liu, Yan-Yun Zeng, Mengnan Cheng, et al. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022.

[68] Maggie MK Wong, Zhiqiang Sha, Lukas Lütje, Xiang-Zhen Kong, Sabrina Van Heukelum, Wilma DJ Van de Berg, Laura E Jonkman, Simon E Fisher, and Clyde Francks. The neocortical infrastructure for language involves region-specific patterns of laminar gene expression. *Proceedings of the National Academy of Sciences*, 121(34):e2401687121, 2024.

[69] 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, 2023.

[70] Keke Xia, Hai-Xi Sun, Jie Li, Jiming Li, Yu Zhao, Lichuan Chen, Chao Qin, Ruiying Chen, Zhiyong Chen, Guangyu Liu, et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in arabidopsis leaves. *Developmental cell*, 57(10):1299–1310, 2022.

[71] Chang Xu, Xiyun Jin, Songren Wei, Pingping Wang, Meng Luo, Zhaochun Xu, Wenyi Yang, Yideng Cai, Lixing Xiao, Xiaoyu Lin, et al. Deepst: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, 50(22):e131–e131, 2022.

[72] Hang Xu, Huazhu Fu, Yahui Long, Kok Siong Ang, Raman Sethi, Kelvin Chong, Mengwei Li, Rom Uddamvathanak, Hong Kai Lee, Jingjing Ling, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1):12, 2024.

[73] Hao Xu, Shuyan Wang, Minghao Fang, Songwen Luo, Chunpeng Chen, Siyuan Wan, Rirui Wang, Meifang Tang, Tian Xue, Bin Li, et al. Spacel: deep learning-based characterization of spatial transcriptome architectures. *Nature Communications*, 14(1):7603, 2023.

[74] Junlin Xu, Jielin Xu, Yajie Meng, Changcheng Lu, Lijun Cai, Xiangxiang Zeng, Ruth Nussinov, and Feixiong Cheng. Graph embedding and gaussian mixture variational autoencoder network for end-to-end analysis of single-cell rna sequencing data. *Cell Reports Methods*, 3(1), 2023.

[75] Zizhen Yao, Cindy TJ Van Velthoven, Michael Kunst, Meng Zhang, Delissa McMillen, Changkyu Lee, Won Jung, Jeff Goldy, Aliya Abdelhak, Matthew Aitken, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, 624(7991): 317–332, 2023.

[76] Zhiyuan Yuan, Fangyuan Zhao, Senlin Lin, Yu Zhao, Jianhua Yao, Yan Cui, Xiao-Yong Zhang, and Yi Zhao. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods*, 21(4):712–722, 2024.

[77] Ron Zeira, Max Land, Alexander Strzalkowski, and Benjamin J Raphael. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19(5):567–575, 2022.

[78] Yuansong Zeng, Rui Yin, Mai Luo, Jianing Chen, Zixiang Pan, Yutong Lu, Weijiang Yu, and Yuedong Yang. Deciphering spatial domains by integrating histopathological image and tran-scriptomics via contrastive learning. *biorxiv*, pages 2022–09, 2022.

[79] Xiang Zhou, Kangning Dong, and Shihua Zhang. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science*, 3 (10):894–906, 2023.

[80] Yongshuo Zong, Tingyang Yu, Xuesong Wang, Yixuan Wang, Zhihang Hu, and Yu Li. const: an interpretable multi-modal contrastive learning framework for spatial transcriptomics. *BioRxiv*, pages 2022–01, 2022.

[81] Chunman Zuo, Junjie Xia, and Luonan Chen. Dissecting tumor microenvironment from spatially resolved transcriptomics data by heterogeneous graph learning. *Nature Communications*, 15(1): 5057, 2024.