Efficient integration of spatial omics data for joint domain detection, matching, and alignment with stMSA

Han Shu[1,2,†], Jing Chen[3,†], Chang Xu[4,†], Jialu Hu[1,2], Yongtian Wang[1,2], Jiajie Peng[1,2], Qinghua Jiang[5,6,*], Xuequn Shang[1,2,*], and Tao Wang[1,2,*]

[1]School of Computer Science, Northwestern Polytechnical University, 1 Dongxiang Rd., 710072, Shaanxi, China

[2]Key Laboratory of Big Data Storage and Management, Ministry of Industry and Information Technology, Northwestern Polytechnical University, 1 Dongxiang Rd., 710072, Shaanxi, China

[3]School of Computer Science and Engineering, Xi'an University of Technology, No.5 South Jinhua rd., 710048, Shaanxi, China

[4]Department of Pharmacy and Pharmaceutical Sciences, National University of Singapore, Block S4A, 117543, Singapore

[5]Center for Bioinformatics, School of Life Science and Technology, Harbin Institute of Technology, No.92 West Dazhi rd., 150000, Heilongjiang, China

[6]School of Interdisciplinary Medicine and Engineering, Harbin Medical University, No.157 Baojian rd., 150081, Heilongjiang, China

[*]Address correspondence to: qhjiang@hit.edu.cn, Shang@nwpu.edu.cn, twang@nwpu.edu.cn
[†]These authors contributed equally to this work.

Abstract

Spatial omics (SO) is a powerful methodology that enables the study of genes, proteins, and other molecular features within the spatial context of tissue architecture. With the growing availability of SO datasets, researchers are eager to extract biological insights from larger datasets for a more comprehensive understanding. However, existing approaches focus on batch effect correction, often neglecting complex biological patterns in tissue slices, complicating feature integration and posing challenges when combining transcriptomics with other omics layers. Here, we introduce stMSA (SpaTial Multi-Slice/omics Analysis), a deep graph contrastive learning model that incorporates graph auto-encoder techniques. stMSA is specifically designed to produce batch-corrected representations while retaining the distinct spatial patterns within each slice, considering both intra- and inter-batch relationships during integration. Extensive evaluations show that stMSA outperforms state-of-the-art methods in distinguishing tissue structures across diverse slices, even when faced with varying experimental protocols and sequencing technologies. Furthermore, stMSA effectively deciphers complex developmental trajectories by integrating spatial proteomics and transcriptomics data and excels in cross-slice matching and alignment for 3D tissue reconstruction.

Keywords: Spatial omics, joint domain detection, 3D reconstruction, cross-slice matching, contrastive learning

INTRODUCTION

Spatial omics (SO) technologies have greatly enhanced our understanding of molecular expression and interactions within tissue architectures (Williams et al. 2022). These methodologies allow for the simultaneous analysis of thousands of genes, proteins, or metabolites, revealing their precise spatial distribution in tissues. Spatial transcriptomics (ST) is a leading technique in this field, providing detailed views of the transcriptome in spatial perspective (Asp et al. 2020). Recently, Spatial CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) (Y Liu et al. 2023) has emerged as a significant advancement in spatial omics. This technique integrates RNA sequencing with the measurement of surface protein expression using specific antibodies, thereby enriching transcriptomic data with valuable spatially resolved proteomic information. This integration allows for a comprehensive understanding of the spatial distribution of proteins, offering invaluable insights into the proteomic landscape within tissues. These advancements have paved the way for novel applications in understanding cellular interactions, identifying spatial biomarkers, and unraveling complex developmental and disease processes (Cheng et al. 2023).

In tandem with the evolution of SO technologies, numerous computational methods have been proposed to analyze individual slices. A critical step in these analyses is the identification of spatial domains, which involves clustering spots with similar molecular expression profiles and spatial locations in an unsupervised manner. Accurate integration of molecular expression and spatial information is essential for this process. For instance, STAGATE (K Dong and Zhang 2022) utilizes self-attention mechanisms and cell type-aware spatial graphs to learn latent representations of spots, while CCST (J Li et al. 2022) employs a Deep Graph Infomax (DGI) model, providing a general framework for multiple ST analysis tasks, including spatial domain detection and cell cycle identification. Although these methods have successfully analyzed individual slices, they remain limited to two-dimensional (2D) analyses due to their single-slice focus.

As the availability of SO data derived from the same tissues, organs, or embryos increases, there is a growing need to integrate multiple slices (Schott et al. 2024). When slices are oriented horizontally, this integration facilitates extended spatial analyses; when vertical, it enables the reconstruction of three-dimensional (3D) models that more accurately reflect true spatial structures and cellular functions. While some methods, such as Scanorama (Hie et al. 2019) and Harmony (Korsunsky et al. 2019), have been adapted for integrating SO data, their performance is often limited due to inadequate utilization of spatial information (Yue et al. 2023). Methods like SEDR (Xu, Fu, et al. 2024) address batch effects by learning latent representations for each slice using deep autoencoders; however, they overlook cross-slice relationships. Recently proposed methods, such as STAligner (Zhou et al. 2023) and STitch3D (G Wang et al. 2023), specifically aim to integrate multiple SO slices but still encounter limitations. For instance, STAligner employs mutual nearest neighbors (MNNs) for integration, while STitch3D requires additional reference data that can impact performance. Moreover, there are currently no effective methods for integrating spatial multi-omics data, such as combining transcriptomics with proteomics, to reveal intricate biological patterns from both perspectives.

To address these challenges, we present stMSA, a novel graph contrastive learning-based method with a core aim: integrating spatial omics (SO) data from multiple sources. stMSA employs a contrastive learning strategy, optimizing two key principles to simultaneously learn the embeddings of cross-batch spots. First, within each slice, the embeddings of spots in connected adjacent spatial niches are made similar, while those of randomly sampled spots remain dissimilar. Second, embeddings of cross-batch spots with similar gene

expression profiles are optimized to be similar, while those of randomly sampled cross-batch spots diverge. By adhering to these principles, stMSA enables joint spatial domain identification across slices, cross-slice spatial matching for correspondence inference, and multi-slice alignment for 3D tissue reconstruction. By unifying these capabilities, stMSA facilitates integrative analysis of spatially resolved multi-omics data, advancing insights into complex tissue architectures.

RESULTS

Overview of stMSA

This study introduces stMSA, a novel deep graph representation learning model designed for integrating multiple spatial omics data (Fig. 1A). The primary objective of stMSA is to generate consistent cross-batch spot embeddings, essential for tasks such as joint domain detection, cross-slice matching, and 3D reconstruction alignment (Fig. 1B).

The model processes preprocessed omics expression data and spatial information from multiple slices as input. For each slice, stMSA constructs a spatial graph using the coordinates of each spot. It then utilizes both the omics expression profiles and the spatial graphs to derive latent representations. The core of stMSA is a contrastive learning strategy that simultaneously considers the inner-batch patterns and the cross-batch patterns to integrate data across multiple slices. This method aims to maximize the similarity of spatially adjacent spots within the same slice and across different slices that exhibit similar gene expression patterns, while minimizing the similarity between dissimilar spot pairs, both within and between slices.

In detail, stMSA enhances the similarity between centroid spots and their associated spatial niche embedding in the latent space, enabling the model to capture detailed local features from neighboring spots. To preserve and highlight the unique expression patterns of each spot, the model also optimizes gene expression reconstruction through an attention-based graph auto-encoder, a process we term the gene expression reconstruction (GER) task. Furthermore, we refine the model's performance by dynamically optimizing the predicted spatial domains using the deep embedding clustering (DEC) method (Xie et al. 2016).

stMSA enhances joint domain identification for human dorsolateral prefrontal cortex slices

We applied stMSA to the human dorsolateral prefrontal cortex (DLPFC) dataset from the 10x Visium platform (Maynard et al. 2021), which includes 12 slices from three donors with detailed spatial annotations (Fig. 2A). To assess its performance in domain identification, we compared stMSA against eight state-of-the-art methods: Scanorama (Hie et al. 2019), Harmony (Korsunsky et al. 2019), SEDR (Xu, Fu, et al. 2024), STAligner (Zhou et al. 2023), Stitch3D (G Wang et al. 2023), SPACEL (Xu, S Wang, et al. 2023), CAST (Tang et al. 2024) and STAMP (Zhong et al. 2024) (detailed in Supplemental Note S1). Domain detection performance (Full domain identification results are visualized in Supplemental Fig. S1) was evaluated using the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), Completeness score (COM), and Homogeneity score (HOM) for each method (Fig. 2B, Supplemental Fig. S2A).

As shown in Fig. 2B, stMSA achieved an average ARI score of 0.544 across 12 DLPFC slices, surpassing both STitch3D (0.535) and SPACEL (0.508). Visual comparison of spatial domain distributions for one

3

donor (slices 151507-151610) further demonstrated stMSA's superior performance, achieving an joint ARI score of 0.508 (Fig. 2D). Other top methods, such as SEDR (0.488) and Stitch3D (0.504), although achieves well ARI score, these methods showed inaccuracies in layers 2-6, whereas stMSA produced more coherent boundaries and shapes. STAligner (0.467) was the closest competitor in spatial domain distribution but was still outperformed by stMSA in accuracy and boundary clarity.

Moreover, we assessed the sensitivity to single-cell reference data quality we cross-applying reference datasets between STitch3D and SPACEL on the DLPFC donor3 slices (151673-151676). Both methods exhibited reduced performance (Supplemental Fig. S2B) when using swapped references, highlighting their dependency on curated SC data for optimal accuracy.

To assess the scalability of stMSA, we applied it to two DLPFC slices from the Stereo-seq platform (S Wei et al. 2024), each containing over 30,000 spots. Comparative analysis of domain identification metrics revealed that stMSA outperformed existing state-of-the-art methods on this benchmark dataset evaluated by the domain prediction accuracy metrics (Fig. 2C, Supplemental Fig. S2). Moreover, stMSA, STAligner, SPACEL, and STAMP well represented the layer structures, while other methods failed to capture the zonal domains (Fig. 2E, Supplemental Fig. S3). Trajectory analyses (Wolf, Hamey, et al. 2019) confirmed that stMSA effectively traced the path from white matter to the arachnoid membrane (Fig. 2F), unlike Scanorama, Harmony, and SEDR, which conflated multiple layers. Notably, stMSA distinguished the thin Layer 1 (domain 3) from the adjacent layers, achieving the highest joint ARI score of 0.665 across two slices.

We assessed stMSA's effectiveness in removing slice-wise batch effects using UMAP visualizations (Supplemental Fig. S3 and S4). The embeddings from three donors were well mixed, with no clear batch effects observed in the 10x Visium dataset. This trend continued in the Stereo-seq dataset, confirming successful integration across donors. Additionally, we evaluated computational efficiency. stMSA processed up to 20 slices with less than 12 GB of GPU memory, while Stitch3D, SPACEL, and SEDR struggled to handle larger datasets (Supplemental Fig. S5).

Identification of cross-batch structures using stMSA in multi-slice mouse brain

We applied stMSA to a sagittal section dataset of the mouse brain obtained via 10x Visium sequencing, which includes multiple slices due to the large tissue volume (Fig.3A). Each slice, covering up to 4,992 spots, was divided into anterior (Fig.3B, left) and posterior (Fig. 3B, right) sections for analysis. Our objective was to identify common spatial domains across slices and assign consistent domain labels.

Fig.3B illustrates the domain detection performance of stMSA relative to the manually annotated structures from the Allen Brain Atlas (Fig.3A). Moreover, we find that stMSA and STAMP can well depict the cross-sections of the hippocampus and its inner subdomains, accurately delineating the "hook"-shaped outline of Ammon's horn. Notably, stMSA is the only method that can accurately represent cross-slice structures, including the six cortical layers in the isocortex and thalamus, while SEDR identified only five layers within the isocortex (Supplemental Fig. S6).

Upon further analysis of fine-grained structures, stMSA demonstrated the ability to accurately recognize subdomains within significant brain structures, such as those in the isocortex and hippocampus. Comparison of gene expression distributions with known markers confirmed the accuracy of these subdomain identifications (Fig. 3C). For example, genes like *Rprm* and *Mobp* exhibited consistent expression patterns corresponding to cortical layers (Sullivan et al. 2023; Grimstvedt et al. 2023; Hoerder-Suabedissen et al. 2022; Fulcher et al. 2019), while *Ckb* showed a "curved tube" structure aligning with our hippocampal

4

domain results (Sandebring-Matton et al. 2021). The spatial distribution of the marker gene *Prkcd* for the thalamus also significantly overlapped with structures identified by stMSA. Additionally, UMAP visualizations and trajectory inference from embeddings learned by stMSA confirmed the correspondence between the relative positions of isocortex layers and their actual physical locations (Fig. 3D, Supplemental Fig. S7).

Next, we assessed stMSA's capability to integrate coronal mouse brain data from various experimental protocols (Supplemental Note S2), including fresh frozen tissue with H&E staining, denoted as FF (H&E), fresh frozen tissue with Immunofluorescence (DAPI) staining, denoted as FF (IF), and formalin-fixed paraffin-embedded (FFPE) tissue with H&E staining, denoted as FFPE (Supplemental Fig. S8A). stMSA standardizes inputs to preprocessed spatial omics expression profiles across all datasets. While auxiliary information (such as H&E/DAPI images) is available in some cases, they are intentionally excluded to ensure model generalizability and focus on core spatial transcriptomic features. We benchmarked stMSA against state-of-the-art methods by analyzing these slices and generating embeddings in a shared feature space. Spatial domain prediction results are separately visualized for each slice to highlight method-specific performance (Fig. 3F, Supplemental Fig. S8B). Results indicated that both stMSA, STAligner, and CAST consistently identified the same tissue structures across different protocols, whereas other methods struggled with consistency. Notably, stMSA and STAligner assigned the same domain labels to the Dentate gyrus across all slices, with stMSA labeling it as domain 1 (Supplemental Fig. S8C) and STAligner as domain 8 (Supplemental Fig. S8B), aligning with its anatomical boundaries in the Allen Brain Atlas (Fig. 3E), where other methods showed inconsistencies in tissue identification across protocols. Additionally, stMSA effectively distinguished between the hypothalamus and white matter, accurately labeling these structures in domain 7 and domain 15, respectively, where STAligner misclassified them into incorrect domains. Moreover, stMSA achieved the highest domain detection ARI score across all slices (Fig. 3G) and other clustering accuracy metrics (Supplemental Fig. S8D), demonstrating its superior ability to learn high-quality latent representations. We further evaluated stMSA against competing methods by integrating two additional mouse brain datasets from FF (fresh frozen) and FFPE (formalin-fixed paraffin-embedded) samples. stMSA demonstrated robust accuracy in identifying spatial domains across both preservation protocols (Supplemental Fig. S9) and achieved the highest overall integration scores, highlighting its enhanced ability to align data across slices and mitigate batch effects in latent representations (Supplemental Fig. S10A).

Overall, our findings highlight stMSA's effectiveness in identifying cross-batch tissue structures, accurately delineating subdomains, and integrating multi-slice datasets. This establishes stMSA as a powerful tool for analyzing complex brain architectures in spatial omics.

Robust integration of spatial omics data across sequencing platforms and dataset sizes

To evaluate the efficacy of stMSA in integrating spatial omics data from various sequencing platforms, we applied it to mouse olfactory bulb data acquired through the Stereo-seq (X Wei et al. 2022) and Slide-seq V2 (Stickels et al. 2021) methods. Similar to the previous experimental settings, we jointly analyzed the two SO slices using each method, with domain prediction taking place in the same feature space. As depicted in Fig. 4A and Supplemental Fig. S11, stMSA successfully recognized consistent structures in both slices, aligning well with the Allen Brain Atlas annotations. Supplemental Fig. S11 demonstrates that STAligner achieved similar results, while other methods encountered difficulties in generating clear domain structures.

Next, we assessed stMSA's performance with smaller datasets. Specifically, we applied stMSA to two mouse olfactory bulb slices obtained via the 10x Visium and Spatial Transcriptomics (Sta˚hl et al. 2016) platforms, which contained 918 and 267 spots, respectively. The domain identification results (Fig. 4A, Supplemental Fig. S12A) demonstrated that stMSA was the only method capable of accurately identifying the joint domain structure. In contrast, other methods incorrectly assigned the same domain in different slices while using distinct cluster labels. Additionally, UMAP visualizations (Supplemental Fig. S12B) indicated that stMSA effectively mitigated batch effects across the two sequencing platforms, validating its capability to learn a batch-effect-free latent representation even in small-scale datasets. Furthermore, the quantitative integration scores demonstrate stMSA's effectiveness in integrating cross-sequencing techniques and eliminating batch effects (Supplemental Fig. S10B).

Furthermore, we examined stMSA's ability to integrate data from imbalanced dataset sizes, as different sequencing platforms may capture data at varying resolutions. To prevent the larger dataset from dominating the integration, stMSA randomly assigns cells from the larger dataset into several new, independent groups for representation learning, thereby balancing its contribution against the smaller dataset. For this evaluation, we utilized E9.5 mouse embryo data from the Stereo-seq platform (4,356 spots) and the Slide-seq V2 platform (14,758 spots) (Fig. 4B). The results revealed that stMSA effectively integrated the imbalanced slices, producing accurate joint domain identification results (Fig. 4B, C). In contrast, Scanorama, Harmony, SEDR, and CAST exhibited significant batch effects in its latent representation, as shown in the batch UMAP visualization, and STAligner and STAMP struggled to identify joint domains (Supplemental Fig. S13).

To further validate the reliability of stMSA's domain identification results, we compared the spatial distribution of identified domains with the expression patterns of corresponding marker genes, which have been confirmed by previous biological experiments. Notable tissue structures, including the brain, heart, spinal cord, lung, and sclerotome, demonstrated a strong correlation with their respective marker genes: *Crabp2* for the brain (Pietila¨ et al. 2023), *Actc1* for the heart (Frank et al. 2019), *Map1b* for the spinal cord (Tortosa et al. 2011), *Ptn* for the lung (Weng and L Liu 2010), and *Pax1* for the sclerotome (J Dong et al. 2018) (Fig. 4D, Supplemental Fig. S14). These findings further corroborate the accuracy of stMSA in domain identification, demonstrating its applicability across various sequencing platforms and dataset sizes.

In conclusion, evaluation through UMAP visualization and quantitative integration scores confirms that stMSA achieves effective elimination of batch effects originating from both experimental protocols and sequencing techniques (Supplemental Fig. S10C).

Deciphering B cell development trajectories through spatial multi-omics integration in human tonsils

To explore the integration of spatial proteomics and transcriptomics data, we applied stMSA to a human tonsil dataset, utilizing Spatial CITE-seq for proteomics and the 10x Visium platform for transcriptomics (Y Liu et al. 2023). This integration enables a more comprehensive understanding of cellular functions (Lundberg and Borner 2019), particularly for B cell development in the tonsil. Notably, the datasets exhibited significant differences in total counts distributions across spots for RNA and protein dataset, posing challenges for effective data harmonization (Fig. 4E). We first assessed domain identification accuracy on multi-omics spatial slices using our input settings (see Methods) and compared performance

6

with other state-of-the-art methods. STAMP, however, could not be applied to integrate RNA and protein data under our preprocessing pipeline because the input datasets violated its foundational assumption of a Gamma-Poisson distribution, which is critical to its statistical model. stMSA successfully identified joint RNA-protein spatial domains (Fig. 4F), demonstrating robust cross-omics integration capabilities. In contrast, CAST failed to harmonize transcriptomic and proteomic slices effectively (Supplemental Fig. S15).

We further evaluated stMSA with methods capable of robust RNA-protein integration (focusing on those demonstrating successful cross-omics alignment). stMSA generated the most expressive latent representations, achieving the lowest Geary's C (minimizing spatial randomness) and highest Moran's I (maximizing spatial dependency) (Fig. 4G). This underscores stMSA's ability to harmonize data patterns from heterogeneous sources in the human tonsil dataset.

To enhance domain resolution, we adjusted the Louvain clustering algorithm's resolution parameter to 1.0, enabling finer domain delineation. This adjustment allowed us to clearly identify the lymph follicular region, with the follicular center designated as domain 6 and the extra-follicular plasma cell region as domain 3, aligning well with histological observations (Fig. 4H, Supplemental Fig. S16). Subsequent analysis of spatially variable genes (SVGs) within the follicular region revealed associations with lymph follicular functions (Hwang et al. 2013; Boyd et al. 2013).

Utilizing the latent representation generated by stMSA for trajectory inference, we found that identified domains aligned well with the developmental stages of B cells. Specific marker proteins and genes, such as IgD and *IGHD* for naive B cells (Espinoza et al. 2023), IgG and *IGHG1* for mature B cells (King et al. 2021), and IgM and *CXCL12* for germinal center (GC) B cells (Barinov et al. 2017), corroborated this alignment (Fig. 4I). The spatial distribution of these markers closely corresponded with the domains identified by stMSA, showcasing a coherent relationship with B cell development. We also utilized CellChat (Jin et al. 2021) to analyze cell-cell communication based on domain identification results. The analysis revealed significant interactions among domains 2, 3, and 4, while domain 0 primarily interacted with domain 1. Notably, domain 1 engaged with all other domains, supporting the inferred B cell development trajectory. Pathway analysis indicated that these interactions were pathway-specific: domain 0 interacted with domain 1 via the EGF pathway, while domains 2, 3, and 4 primarily connected through the PDGF and CXCL pathways (Fig. 4J).

We further validated stMSA by integrating mouse spleen transcriptomics (SPOTS) (Ben-Chetrit et al. 2023) and proteomics (Spatial CITE-seq) (Y Liu et al. 2023) data. stMSA successfully harmonized RNA and protein spots in a unified embedding space (Supplemental Fig. S17B), aligning with tissue histology (Supplemental Fig. S17A) to resolve distinct splenic architectures. In summary, the experimental results demonstrate that stMSA is capable of integrating multi-omics data effectively. It successfully captures the developmental patterns within the multi-omics datasets, accurately identifies domain results, and constructs a correct developmental trajectory of B cells in the human tonsil dataset.

Elucidating tissue identity through cross-slice matching in embryonic development

Matching spots across adjacent spatial omics slices is critical for understanding the developmental trajectories of various cell types, tissues, and the process of organogenesis. This matching involves establishing correspondences between spot pairs from different slices taken at varying developmental stages, ensuring that matched spots exhibit similar gene expression profiles while preserving the spatial

7

relationships inherent in each SO slice. To evaluate the cross-slice matching capabilities of stMSA, we applied the model to the Mouse Organogenesis Spatiotemporal Transcriptomic Atlas (MOSTA) dataset (Chen et al. 2022).

The results of applying stMSA to match tissue types between two mouse embryo slices collected on embryonic days 14.5 (E14.5) and 15.5 (E15.5) are illustrated in Fig. 5A. The Sankey plot in Fig. 5B demonstrates effective matching between these slices, revealing that most tissue types in mature organs, including the brain, heart, muscle, liver, and lung, have been accurately aligned. Furthermore, this matching process enabled us to identify specific organogenesis stages, such as the finding that cartilage at stage E15.5 develops from cartilage and cartilage primordium at stage E14.5. This developmental relationship is supported by the matching results shown in Fig. 5C, which highlights only the matched spot pairs between the two tissues. To further validate this relationship, we visualized the spatial distribution of *Col2a1*, a marker gene for cartilage (Hissnauer et al. 2010), alongside the domain distribution of cartilage and cartilage primordium at stage E14.5. Notably, the expression pattern of *Col2a1* aligned closely with the corresponding domain distributions (Fig. 5D).

To quantitatively assess the overall matching performance of stMSA, we calculated a confidence score that evaluates the similarity between the embedded spots that were matched. This score, defined as the cosine distance between the embeddings of two matched spots, is presented in Fig. 5E for the E14.5 slice, clearly illustrating a high degree of similarity between the matched spots. Additionally, we extended our analysis to examine cross-slice matching between pairs at other stages, specifically 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 results with those obtained using the Spatial-linked Analysis of Transcriptomics (SLAT), which has four different initialization settings based on input types: using all genes (Raw init), only highly variable genes (HVG init), dimension-reduced gene expression profiles based on principal component analysis (PCA init), and dual PCA (DPCA init) (Xia et al. 2023). Fig. 5F summarizes the matching scores obtained by stMSA and the four SLAT versions (detailed in Supplemental Note S1), demonstrating that stMSA outperforms SLAT across various settings.

stMSA enables the multi-slice spatial alignment

Current sequencing technologies limit researchers to capturing two-dimensional (2D) slices of biological tissues, posing challenges in fully understanding three-dimensional (3D) biological structures. To address this, stMSA proposed an automated spatial omics alignment through a landmark domain-guided enhanced ICP algorithm, which optimizes a rotation matrix for cross-slice coordinate harmonization (see Methods).

We first benchmarked stMSA against leading alignment tools PASTE (Zeira et al. 2022), STAligner (Zhou et al. 2023), and SANTO (H Li et al. 2024) using a two-slice mouse spleen ST dataset (Ben-Chetrit et al. 2023). The initial positioning of these two slices exhibits a substantial spatial misalignment, necessitating further refinement to achieve precise alignment (Fig. 6A). stMSA leveraged the macrophage-enriched (MP enriched) region as the landmark domain to guide alignment, validated by the spatial distribution of the macrophage-associated gene marker *Rsad2* (Ding et al. 2025) (Supplemental Fig. S18A). Visual assessment confirmed stMSA's accuracy in reconstructing coherent 3D tissue structures from 2D slices (Fig. 6B), whereas PASTE achieved only partial alignment, and STAligner/SANTO failed to establish spatial correspondence. Quantitatively, stMSA outperformed all competing methods, achieving the highest Pearson's Correlation Coefficient (PCC) and Cell Type Matching Index (CI) scores (Fig. 6C).

To assess scalability, we applied stMSA to align seven coronal sections from mouse brains (Ballester

8

Roig et al. 2023). This dataset contained two flipped orientations, including "Orient 1" (4 slices) and "Orient 2" (3 slices), with significant initial misalignment (Fig. 6E, left). stMSA first identified spatial domains consistent with Allen Brain Atlas annotations (Fig. 3F). For instance, the pyramidal layer (domain 7) in the hippocampus aligned precisely with both the histology-derived hook-like structure and the expression pattern of the marker gene *Hpca* (Tzingounis et al. 2007), prompting its selection as the landmark domain for alignment.

Using a modified ICP algorithm, stMSA successfully reconciled rotations and flips between "Orient 1" and "Orient 2", integrating all slices into a unified coordinate system (Fig. 6E, right). Quantitative evaluation confirmed stMSA's superior alignment scores for both orientations (Fig. 6D), with 2D visualizations further validating coherent integration (Supplemental Fig. S18B). Together, these results demonstrate stMSA's robustness in aligning multiple spatial omics slices, enabling accurate 3D tissue reconstruction.

DISCUSSION

The rapid advancement of sequencing technologies has resulted in the generation of an increasing volume of high-quality spatial omics data. This growth has created a pressing need to integrate spatial omics data from diverse sources, which may vary in experimental protocols, sequencing platforms, size, shape, and orientation. In this work, we present a novel contrastive learning-based method, termed stMSA, designed for the integration of multiple spatial omics datasets. stMSA learns cross-batch spot embeddings by optimizing two key principles: (1) enhancing the similarity of embeddings within a slice for spots situated in connected spatial niches and (2) maximizing the similarity of embeddings between cross-batch spots with analogous gene expression profiles. By adhering to these principles, stMSA effectively captures the distribution patterns of spots both within and across slices. Additionally, stMSA employs graph attention-based autoencoder techniques and deep embedding clustering to optimize model parameters and reconstruct gene expression profiles.

stMSA preserves slice-specific biological signals and prevents over-mixing through two complementary approaches. The framework first employs gene expression reconstruction to constrain the model to reproduce original transcriptional patterns, ensuring conservation of unique biological signatures. Simultaneously, it only identifies high-confidence cross-slice similarity pairs (mutual nearest neighbors) using dynamic stringent thresholding, thereby mitigating over-mixing between different slices. These two strategies prioritize biologically conserved features while maintaining distinct slice-specific profiles, effectively balancing data harmonization with biological fidelity.

A key advantage of stMSA lies in its ability to integrate multiple spatial omics slices, thereby overcoming the limitations of single-slice analysis. Traditional single-slice approaches often produce fragmented or misaligned spatial domains, particularly in tissues with structural continuity. In contrast, stMSA jointly analyzes adjacent sections to yield consistent and biologically meaningful domain identification across slices, eliminating the need for manual correction. Moreover, stMSA projects multi-slice data into a unified latent space, enabling downstream analyses in a pseudo-3D context. This approach facilitates tasks such as reconstructing cross-slice developmental trajectories and aligning multiple unaligned spatial slices. These capabilities allow stMSA to uncover spatially conserved biological processes and microenvironments that are typically inaccessible through isolated single-slice analyses.

While stMSA demonstrates robust performance in spatial omics integration, we acknowledge that its

9

current implementation has higher computational demands compared to state-of-the-art tools. For example, when aligning 20 consecutive slices (4,226 spots per slice), stMSA requires approximately 15 minutes and 12 GB of GPU memory. Under identical hardware conditions, CAST completes the alignment in approximately 8 minutes using 14 GB GPU memory, while STAMP requires approximately 32 minutes with only 1 GB GPU memory (Supplemental Fig. S4). This mainly stems from stMSA's multi-task optimization architecture, while it enhances model performance, they result in higher computational costs. In future work, we will prioritize algorithmic optimizations to reduce the model's runtime and memory consumption while retaining its performance.

In conclusion, our results demonstrate that stMSA offers a comprehensive solution for integrating multiple spatial omics slices from different sources, facilitating integrative analyses of spatial omics data. We strongly believe that as spatial omics sequencing technologies continue to evolve, stMSA will further empower researchers to uncover novel biological insights from an increasingly diverse array of SO datasets.

METHODS

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 (Wolf, Angerer, et al. 2018). To overcome the zero inflation in spatial transcriptomics data, stMSA selects the top k highly variable genes (HVGs) as input features. For SO slices with common gene backgrounds, such as the DLPFC dataset (Maynard et al. 2021), we set k as 3,000 by default. For SO slices with varying genes, usually derived from different platforms, it is suggested to set k higher and use the intersecting genes as input features. For proteomics and transcriptomics integration, we use the Principal Component Analysis (PCA) (Jolliffe and Cadima 2016) to decompose the protein and RNA data and use the top 300 PCs as the input features.

Spatial graph construction for SO 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. We fine-tuned the radius threshold r to achieve spatial graphs with an approximate connectivity of 6 neighbors for each spot by default. The adjacency matrices for spatial graphs are denoted as $\mathbf{A} = \{\mathbf{A}^{(s_1)}, \mathbf{A}^{(s_2)}, ..., \mathbf{A}^{(s_k)}\}$, where $\mathbf{A}_{ij}^{(s_k)}$ is set as 1 only if spot i and spot j within slice s_t are connected, and it is assigned 0 otherwise. For the imbalanced datasets, we randomly sampled the large slice into two small datasets and treated them as two individual slices.

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) (Velicˇkovicˊ et al. 2018) and a fully connected network (FCN) (Yosin- ski et al. 2015) 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.

10

Encoder

stMSA adopts the GAT network to aggregate spot features within a local spatial niche 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 $\mathbf{X}^{(s_m)} = \{\mathbf{x}_1^{(s_m)}, \mathbf{x}_2^{(s_m)}, \ldots, \mathbf{x}_{n^{(s_m)}}^{(s_m)}\}$, 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:

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

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. $\mathbf{W}^{(\text{GAT})}$ 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:

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

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:

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

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 (Clevert et al. 2016).

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:

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

11

$$\mathbf{out}_i^{(s_m)} = \mathbf{W}^{(GAT)^T} \cdot \text{ELU}\left(\hat{\mathbf{h}}_i^{(s_m)}\right)$$

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.

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 SO slices. It consists of two components: inner-batch contrastive learning and cross-batch contrastive learning.

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 spatial niche (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:

$$\mathcal{L}_{\text{inner}} = \frac{1}{N}\sum_{s_m}\sum_i \max\left\{\left\|\mathbf{z}_i^{(s_m)} - \mathbf{env}_i^{(s_m)}\right\|_2 - \left\|\mathbf{z}_i^{(s_m)} - \mathbf{z}_k^{(s_m)}\right\|_2 + \gamma, 0\right\} \tag{5}$$

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 spatial niche embedding constructed from the embeddings of neighboring spots of spot i, which is calculated as follows:

$$\mathbf{env}_i^{(s_m)} = \sum_{j \in N_i^{(s_m)}} \mathbf{z}_j^{(s_m)} \tag{6}$$

Cross-batch pairs identification

stMSA initiates cross-batch contrastive learning by dynamically selecting and refining cross-slice similar pairs. For each two distinct slices from all input slices, stMSA identifies candidate similar pairs by computing the L2 distance between spot embeddings, selecting the k-nearest neighbors (k-NNs) for each spot. To ensure high confidence matches, pairs with L2 distances exceeding 75% of the maximum distance are filtered out, while only pairs with distances below this threshold are retained. Retained pairs are further refined by retaining only mutually nearest neighbors (MNNs) where spots reciprocally rank in each other's k-NN lists to minimize spurious associations. These curated MNNs serve as positive pairs to maximize similarity during cross-batch contrastive learning.

To adapt to evolving embeddings, cross-batch pairs are regenerated every 100 epochs. During initial training (epoch 0), embeddings are derived from the top 30 principal components (PCs) of the input data, enabling stable cold-start optimization.

12

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:

$$\mathcal{L}_{\text{cross}} = \frac{1}{N}\sum_{s_m}\sum_{s_n > s_m}\sum_{i}\max\left\{\left\|\mathbf{z}_i^{(s_m)} - \mathbf{z}_j^{(s_n)}\right\|_2 - \left\|\mathbf{z}_i^{(s_m)} - \mathbf{z}_k^{(s_p)}\right\|_2 + \gamma, 0\right\} \tag{7}$$

where $\mathbf{z}^{(s_m)}$, $\mathbf{z}^{(s_n)}$, and $\mathbf{z}^{(s_p)}$ 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:

$$\mathcal{L}_{CL} = \mathcal{L}_{\text{inner}} + \mathcal{L}_{\text{cross}} \tag{8}$$

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.

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:

$$\mathcal{L}_{\text{GER}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{out}_i - \mathbf{x}_i)^2 \tag{9}$$

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.

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 (Xie et al. 2016). The primary goal is to minimize the distance between the soft clustering distribution (Krizhevsky et al. 2017) and the auxiliary target distribution, as depicted below:

$$\mathcal{L}_{DDT} = KL(P||Q) = \sum_{i}\sum_{j}p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{10}$$

13

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:

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

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{N} q_{ij}}{\sum_{k=1}^{K}(q_{ik}^2 / \sum_{i=1}^{N} q_{ik})} \tag{11}$$

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 (Blondel et al. 2008), 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:

$$\mathcal{L} = \mathcal{L}_{CL} + \mathcal{L}_{\text{GER}} + 0.01 \cdot \mathcal{L}_{\text{DDT}} \tag{12}$$

In all experiments conducted in this study, we set the loss weights to 1.0 for both the contrastive learning and gene expression reconstruction loss, and to 0.01 for the domain distribution tuning loss. This weighting scheme emerged from empirical validation across diverse datasets, demonstrating effective balance between objectives while maintaining robust task performance in our evaluations. While these fixed parameters delivered consistent results across all tested conditions, users working with their own data types may optionally adjust these weights to prioritize specific tasks.

Moreover, we conducted an ablation study to assess the individual contributions of each loss function (Supplemental Note S3.1), quantifying their impact on spatial integration performance.

Joint domain detection across SO slices

After encoding the spot features across different SO slices into a coherent embedding space, we use the unsupervised clustering technique to identify the spatial domains in the joint embedding space. 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 (Scrucca et al. 2016) to identify spatial domains. In cases where the cluster number is unknown, stMSA utilizes the Louvain clustering algorithm for community detection (Blondel et al. 2008) which has been implemented in the Python package SCANPY (Wolf, Angerer, et al. 2018).

Cross-slice matching

The objective of matching spots across two adjacent SO 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 SO 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.

14

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 SO alignment methods, we have enhanced the Iterative Closest Point (ICP) algorithm (Besl and McKay 1992) to align slices using the domain detection output from stMSA.

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:

$$R, t = \arg \min_{R,t} \frac{1}{n} \sum_{i=1}^{n} \left\| p_i^{(t)} - \left(R \cdot p_i^{(s)} + t \right) \right\|^2 \tag{13}$$

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 (Yang et al. 2016). 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 (Supplemental Note S3.2).

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 (Kingma and Ba 2017) with a learning rate of 0.001. We implement the stMSA model using the popular graph neural network framework PyTorch Geometric (PyG) (Fey and Lenssen 2019). We set the number of training epochs to 500. In the cross-batch contrastive learning, we update the cross-batch pairs, and for the DDT task, we update the DEC p distribution every 100 epochs.

Evaluation Metrics

The descriptions of all evaluation metrics applied in this study are detailed in Supplemental Note S4.

Data availability

Multiple SO 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 Stereo-seq DLPFC dataset can be found at NGDC database BioProject PRJCA009779. The mouse brain serial dataset can be accessed at STOmicsDB (Anterior: https://db.cngb.org/stomics/datasets/STDS0000018, Posterior:

https://db.cngb.org/stomics/datasets/STDS0000021). The mouse brain coronal fresh frozen H&E and DAPI (IF) stained dataset can be found at squidpy dataset (H&E: https://squidpy.readthedocs.io/en/stable/api/squidpy.datasets.visium_hne_adata.html, DAPI: https://squidpy.readthedocs.io/en/stable/api/squidpy.datasets.visium_fluo_adata.html). The mouse brain coronal FFPE dataset can be found at STOmicsDB (https://db.cngb.org/stomics/datasets/STDS0000052). The FF and FFPE mouse brain dataset can be found at 10X Genomics database (FF: https://www.10xgenomics.com/datasets/adult-mouse-brain-coronal-section-visium-ff-1-standard, FFPE: https://www.10xgenomics.com/datasets/mouse-brain-coronal-section-1-ffpe-2-standard). 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 Visium mouse olfactory bulb can be found at GEO GSM4656181 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4656181). The Spatial Transcriptomics mouse olfactory bulb can be found at NCBI database BioProject ID PRJNA316587. The mouse organogenesis spatiotemporal transcriptomic atlas dataset can be found at STOmicsDB (https://db.cngb.org/stomics/mosta/spatial/). The Slide-seq V2 mouse embryo dataset can be found at cellxgene database (https://datasets.cellxgene.cziscience.com/acc80ff4-5dee-46cc-bf22-84a9a83c9c38.h5ad). The proteomics human tonsil and mouse spleen dataset can be found at GEO GSE213264 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213264). The 10X Visium human tonsil dataset can be found at 10X Genomics database (https://www.10xgenomics.com/dataset s/visium-cytassist-gene-and-protein-expression-library-of-human-tonsil-with-add-on-antibodies-h-e-6-5-mm-ffpe-2-standard). The mouse spleen spatial transcriptomics dataset can be found at GSE198353 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198353). The mouse brain coronal dataset can be found at STOmicsDB (https://db.cngb.org/stomics/datasets/STDS0000218).

A summary of the dataset information can be found in Supplemental Table S1, 2.

Software availability

The implemented code is available at GitHub (https://github.com/hannshu/stMSA) and as Supplemental Code.

Competing interest statement

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

Acknowledgements

Author contributions:

T.W., J.C., Q.J., and X.S. conceived the study and experiments, H.S. implemented the software and conducted the analyses, H.S., C.X., J.H., Y.W., and J.P. analyzed the results, T.W., J.C., H.S, and C.X. wrote and reviewed the manuscript and X.S., T.W., and J.P. supervised the research and provided funding support.

References

Asp M, Bergenstråhle J, and Lundeberg J. 2020. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays*. **42**: 1900221.

Ballester Roig MN, Leduc T, Dufort-Gervais J, Maghmoul Y, Tastet O, and Mongrain V. 2023. Probing pathways by which rhynchophylline modifies sleep using spatial transcriptomics. *Biol Direct*. **18**: 21.

Barinov A, Luo L, Gasse P, Meas-Yedid V, Donnadieu E, Arenzana-Seisdedos F, and Vieira P. 2017. Essential role of immobilized chemokine CXCL12 in the regulation of the humoral immune response. *Proc Natl Acad Sci*. **114**: 2319–2324.

Ben-Chetrit N et al. 2023. Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol*. **41**: 788–793.

Besl P and McKay ND. 1992. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell.* **14**: 239–256.

Blondel VD, Guillaume JL, Lambiotte R, and Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech*. **2008**: P10008.

Boyd SD, Natkunam Y, Allen JR, and Warnke RA. 2013. Selective Immunophenotyping for Diagnosis of B-cell Neoplasms: Immunohistochemistry and Flow Cytometry Strategies and Results. *Appl Immunohistochem Mol Morphol*. **21**: 116–131.

Chen A et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*. **185**: 1777–1792.e21.

Cheng M, Jiang Y, Xu J, Mentis AFA, Wang S, Zheng H, Sahu SK, Liu L, and Xu X. 2023. Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. *J Genet Genomics*. **50**: 625–640.

Clevert DA, Unterthiner T, and Hochreiter S 2016. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv. doi: 10.48550/arxiv.1511.07289.

Ding X et al. 2025. *RSAD2*: A pathogenic interferon-stimulated gene at the maternal-fetal interface of patients with systemic lupus erythematosus. *Cell Rep Med*. **6**: 101974.

Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, Guo H, Wen L, and Tang F. 2018. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol*. **19**: 31.

Dong K and Zhang S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun*. **13**: 1739.

Espinoza DA, Coz CL, Cabrera EC, Romberg N, Bar-Or A, and Li R. 2023. Distinct stage-specific

17

transcriptional states of B cells derived from human tonsillar tissue. *JCI Insight*. **8**: e155199.

Fey M and Lenssen JE 2019. Fast Graph Representation Learning with PyTorch Geometric. arXiv. doi: 10.48550/arxiv.1903.02428.

Frank D et al. 2019. Cardiac α-Actin (ACTC1) Gene Mutation Causes Atrial-Septal Defects Associated With Late-Onset Dilated Cardiomyopathy. *Circ Genom Precis Med*. **12**: e002491.

Fulcher BD, Murray JD, Zerbi V, and Wang XJ. 2019. Multimodal gradients across mouse cortex. *Proc Natl Acad Sci*. **116**: 4689–4695.

Grimstvedt JS, Shelton AM, Hoerder-Suabedissen A, Oliver DK, Berndtsson CH, Blankvoort S, Nair RR, Packer AM, Witter MP, and Kentros CG. 2023. A multifaceted architectural framework of the mouse claustrum complex. *J Comp Neurol*. **531**: 1772–1795.

Hie B, Bryson B, and Berger B. 2019. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol*. **37**: 685–691.

Hissnauer TN et al. 2010. Identification of molecular markers for articular cartilage. *Osteoarthritis Cartilage*. **18**: 1630–1638.

Hoerder-Suabedissen A, Ocana-Santero G, Draper TH, Scott SA, Kimani JG, Shelton AM, Butt SJB, Molna´r Z, and Packer AM. 2022. Temporal origin of mouse claustrum and development of its cortical projections. *Cereb Cortex*. **33**: 3944–3959.

Hwang IY, Hwang KS, Park C, Harrison KA, and Kehrl JH. 2013. Rgs13 Constrains Early B Cell Responses and Limits Germinal Center Sizes. *PLoS One*. **8**: e60139.

Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, and Nie Q. 2021. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. **12**: 1088.

Jolliffe IT and Cadima J. 2016. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. **374**: 20150202.

King HW, Orban N, Riches JC, Clear AJ, Warnes G, Teichmann SA, and James LK. 2021. Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol*. **6**: eabe6291.

Kingma DP and Ba J 2017. Adam: A Method for Stochastic Optimization. arXiv. doi: 10.48550/arxiv.1412.6980.

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh Pr, and Raychaudhuri S. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. **16**: 1289–1296.

Krizhevsky A, Sutskever I, and Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. *Commun ACM*. **60**: 84–90.

Li H, Lin Y, He W, Han W, Xu X, Xu C, Gao E, Zhao H, and Gao X. 2024. SANTO: a coarse-to-fine alignment and stitching method for spatial omics. *Nat Commun*. **15**: 6048.

Li J, Chen S, Pan X, Yuan Y, and Shen HB. 2022. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat Comput Sci*. **2**: 399–408.

Liu Y et al. 2023. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol*. **41**: 1405–1409.

Lundberg E and Borner GHH. 2019. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol*. **20**: 285–302.

18

Maynard KR et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci*. **24**: 425–436.

Pietila¨ R et al. 2023. Molecular anatomy of adult mouse leptomeninges. Neuron. 111: 3745–3764.e7. Sandebring-Matton A, Axenhus M, Bogdanovic N, Winblad B, Schedin-Weiss S, Nilsson P, and Tjernberg LO. 2021. Microdissected Pyramidal Cell Proteomics of Alzheimer Brain Reveals Alterations in Creatine Kinase B-Type, 14-3-3-gamma, and Heat Shock Cognate 71. *Front Aging Neurosci*. **13**: 735334.

Schott M et al. 2024. Open-ST: High-resolution spatial transcriptomics in 3D. *Cell*. 3953–3972.e26.

Scrucca L, Fop M, Murphy TB, and Raftery AE. 2016. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. **8**: 289–317. PMID: 27818791.

Sta˚hl PL et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. **353**: 78–82.

Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, Arlotta P, Macosko EZ, and Chen F. 2021. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol*. **39**: 313–319.

Sullivan KE, Kraus L, Kapustina M, Wang L, Stach TR, Lemire AL, Clements J, and Cembrowski MS. 2023. Sharp cell-type-identity changes differentiate the retrosplenial cortex from the neocortex. *Cell Rep*. **42**: 112206.

Tang Z, Luo S, Zeng H, Huang J, Sui X, Wu M, and Wang X. 2024. Search and match across spatial omics samples at single-cell resolution. *Nat Methods*. **21**: 1818–1829.

Tortosa E, Montenegro-Venegas C, Benoist M, Ha¨rtel S, Gonza´lez-Billault C, Esteban JA, and Avila J. 2011. Microtubule-associated Protein 1B (MAP1B) Is Required for Dendritic Spine Development and Synaptic Maturation. *J Biol Chem*. **286**: 40638–40648.

Tzingounis AV, Kobayashi M, Takamatsu K, and Nicoll RA. 2007. Hippocalcin Gates the Calcium Activation of the Slow Afterhyperpolarization in Hippocampal Pyramidal Cells. *Neuron*. **53**: 487–493.

Velicˇkovic´ P, Cucurull G, Casanova A, Romero A, Lio` P, and Bengio Y 2018. Graph Attention Networks. arXiv. doi: 10.48550/arxiv.1710.10903.

Wang G, Zhao J, Yan Y, Wang Y, Wu AR, and Yang C. 2023. Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nat Mach Intell*. **5**: 1200–1213.

Wei S et al. 2024. Charting the Spatial Transcriptome of the Human Cerebral Cortex at Single-Cell Resolution. bioRxiv. doi: 10.1101/2024.01.31.576150.

Wei X et al. 2022. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*. **377**: eabp9444.

Weng T and Liu L. 2010. The role of pleiotrophin and β -catenin in fetal lung development. *Respir Res*. **11**: 80.

Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, and Haque A. 2022. An introduction to spatial transcriptomics for biomedical research. *Genome Med*. **14**: 68.

Wolf FA, Angerer P, and Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. **19**: 15.

Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Go¨ttgens B, Rajewsky N, Simon L, and Theis FJ. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol*. **20**: 59.

Xia CR, Cao ZJ, Tu XM, and Gao G. 2023. Spatial-linked alignment tool (SLAT) for aligning heterogenous

slices. *Nat Commun*. **14**: 7236.

Xie J, Girshick R, and Farhadi A 2016. Unsupervised Deep Embedding for Clustering Analysis. arXiv. doi: 10.48550/arxiv.1511.06335.

Xu H, Fu H, et al. 2024. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med*. **16**: 12.

Xu H, Wang S, et al. 2023. SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nat Commun*. **14**: 7603.

Yang J, Li H, Campbell D, and Jia Y. 2016. Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration. *IEEE Trans Pattern Anal Mach Intell*. **38**: 2241–2254.

Yosinski J, Clune J, Nguyen A, Fuchs T, and Lipson H 2015. Understanding Neural Networks Through Deep Visualization. arXiv. doi: 10.48550/arxiv.1506.06579.

Yue L, Liu F, Hu J, Yang P, Wang Y, Dong J, Shu W, Huang X, and Wang S. 2023. A guidebook of spatial transcriptomic technologies, data resources and analysis approaches. *Comput Struct Biotechnol J*. **21**: 940–955.

Zeira R, Land M, Strzalkowski A, and Raphael BJ. 2022. Alignment and integration of spatial transcriptomics data. *Nat Methods*. **19**: 567–575.

Zhong C, Ang KS, and Chen J. 2024. Interpretable spatially aware dimension reduction of spatial transcriptomics with STAMP. *Nat Methods*. **21**: 2072–2083.

Zhou X, Dong K, and Zhang S. 2023. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nat Comput Sci*. **3**: 894–906.
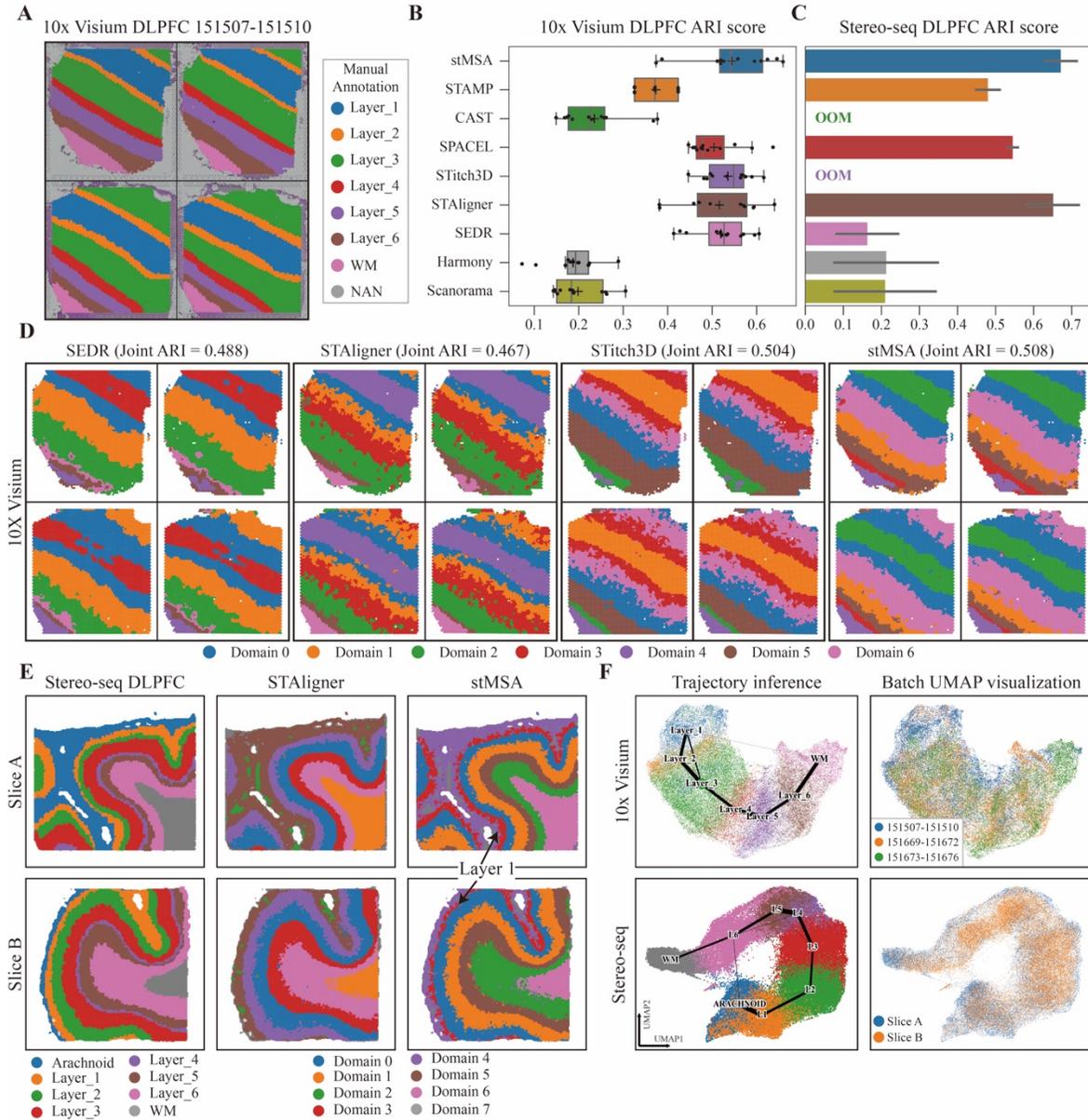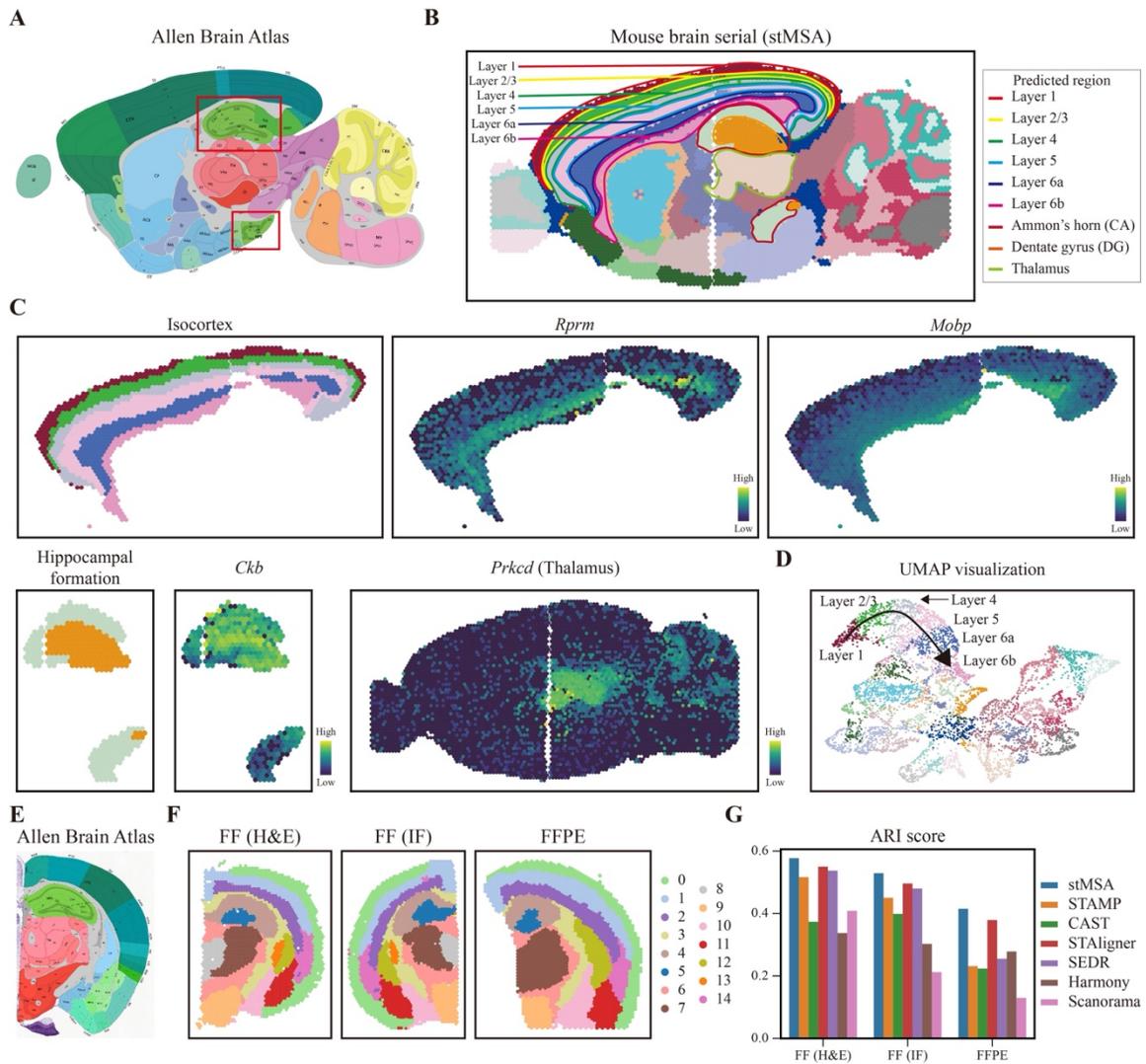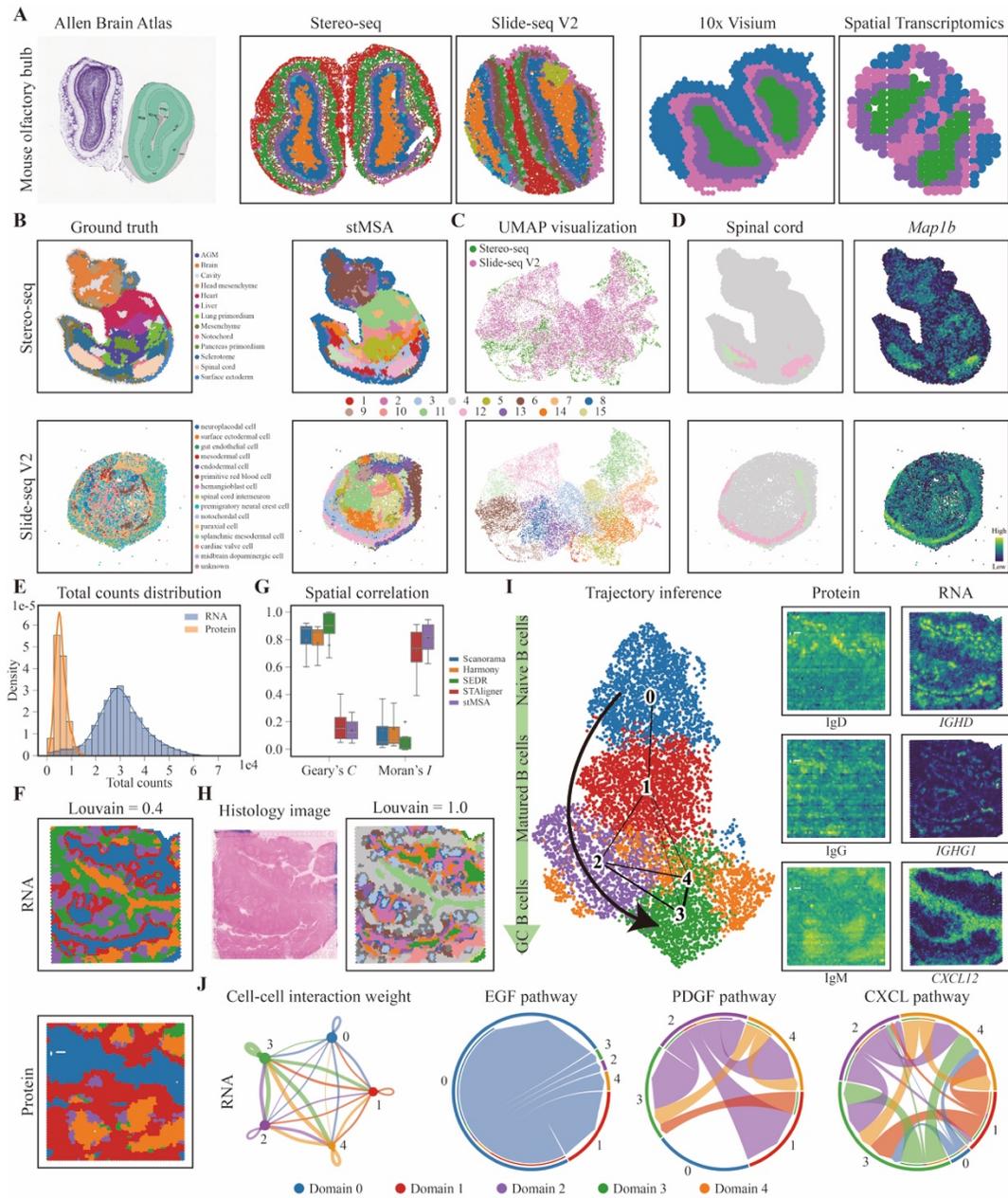
Figures
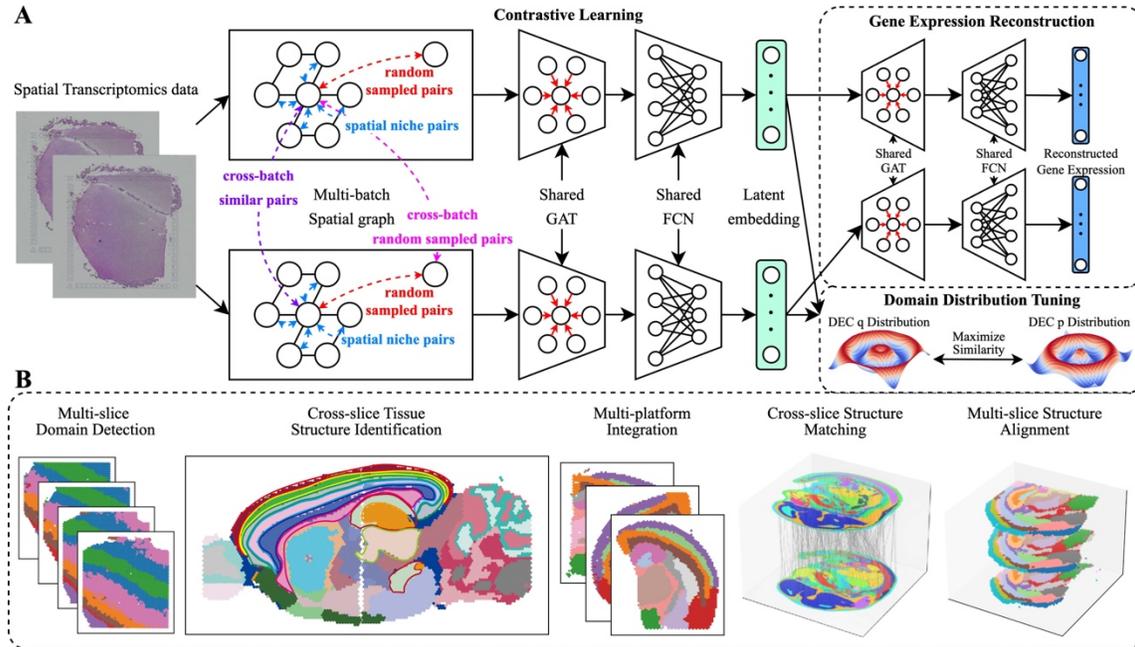
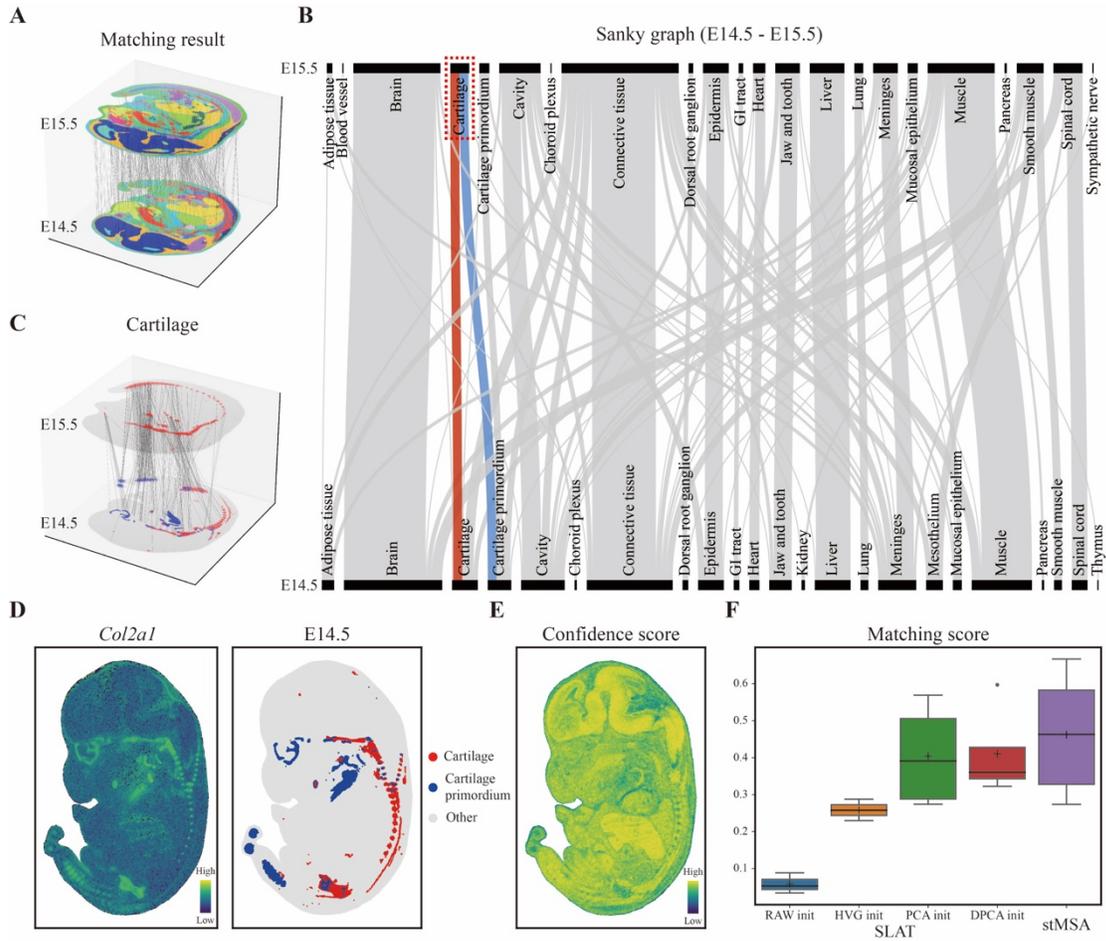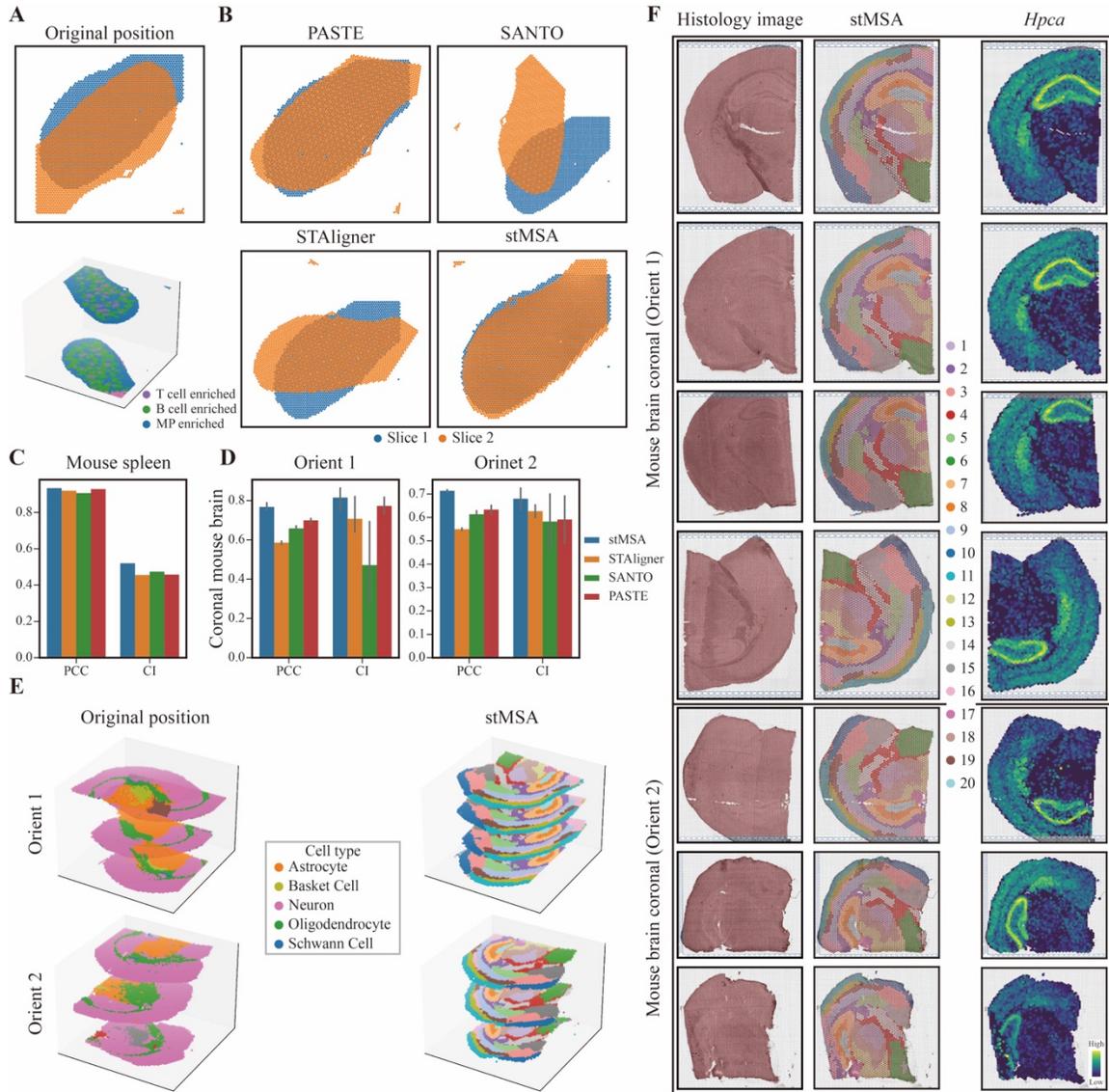Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure Legends

Figure 1: Overview of the stMSA framework. (A) stMSA leverages multi-slice omics 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 joint domain identification for human dorsolateral prefrontal cortex slices.
(A) Manual annotations, ranging from the cortex layer 1 to the white matter, are provided with the histology image of a donor (slice identifiers: 151507-151510). **(B)** The box plot illustrates the ARI scores across 12 10x Visium DLPFC slices, generated by training separately for each donor. **(C)** The bar plot illustrates the average ARI scores for the two Stereo-seq DLPFC slices. Error bars represent the 95% confidence intervals of the ARI scores. Notably, STitch3D and CAST were unable to produce results for the Stereo-seq DLPFC dataset on our server with 24 GB GPU memory and 128 GB RAM. **(D)** Visualization of domain detection results for SEDR, STAligner, Stitch3D, and stMSA. The domain detection processes are conducted using all four slices as inputs. **(E)** The ground truth, STAligner, and stMSA domain detection result of the two Stereo-seq obtained DLPFC slices. **(F)** The trajectory inference plot generated based on the embeddings learned by stMSA using all 12 DLPFC slices for the 10x Visium and the two Stereo-seq DLPFC slices (The color reflects the ground truth domains). And the UMAP visualization for different donors/slices (The color reflects three donors/slices).

Figure 3: stMSA can identify cross-batch tissue structures in the 10x Visium-obtained mouse brain dataset. (A) The anatomical structure of the 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)** Subdomains of the isocortex, hippocampus, and thalamus, along with the expression levels of their corresponding marker genes. **(D)** 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. **(E)** The anatomical structure of the coronal mouse brain provided by the Allen Brain Atlas. **(F)** The clustering result of stMSA in fresh frozen with H&E stained, DAPI stained, and FFPE preserved coronal mouse brain slices. **(G)** The ARI clustering score for stMSA and other state-of-the-art methods in the three coronal mouse brain slices.

Figure 4: stMSA achieves robust performance across diverse sequencing platforms, different dataset sizes, and multi-omics data. (A) The Allen Brain Atlas, the domain identification results for Stereo-seq and Slide-seq V2 platform and 10x Visium and Spatial Transcriptomics platform for mouse olfactory bulb slices. **(B)** The manual annotations and the domain identification result of stMSA in the Stereo-seq and Slide-seq V2 obtained mouse embryo dataset. **(C)** The UMAP visualization of the distribution of different sequencing platforms and the clustering result. **(D)** The spatial distribution of spinal cord region stMSA predicted and its corresponding marker gene. **(E)** The total RNA and protein counts distribution in the human tonsil dataset. **(F)** The domain

27

identification results for the human tonsil protein and RNA dataset, with Louvain resolution set to 0.4. **(G)** The spatial correlation score Geary's C and Moran's I for the latent representation generated by Scanorama, Harmony, SEDR, STAligner, and stMSA. For Geary's C, a low C score indicated high correlation, and for Moran's I, a high I score indicated high correlation. **(H)** The histology image and the domain identification result for the human tonsil RNA dataset, with the Louvain resolution set to 1.0. **(I)** The trajectory inference for the human tonsil dataset and the corresponding spatial distribution of the protein marker and gene marker for the naive B cells (IgD/*IGHD*), Matured B cells (IgG/*IGHG1*), and GC B cells (IgM/*CXCL12*). **(J)** The cell-cell communication analyses for the five domains stMSA identified.

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. The confidence scores quantify the accuracy of cross-slice cell type matches, with higher scores reflecting stronger alignment confidence. For each spot, the score is computed as the proportion of matched cell type pairs relative to the total identified pairs (cell type matched pairs / total pairs). Thus, a higher score indicates greater precision in the biological correspondence for that spot. **(F)** Comparison of the matching scores between stMSA and SLAT under different settings (RAW, HVG, PCA, and DPCA init).

Figure 6: The alignment performance of stMSA on multi-slice SO datasets. (A) Original unaligned 2D/3D spatial coordinates from the mouse spleen dataset. **(B)** Comparative visualization of alignment results from stMSA and three benchmark methods on the mouse spleen dataset. **(C)** Quantitative evaluation of alignment performance using PCC and CI scores. Error bars represent the 95% confidence intervals of the PCC and CI scores. **(D)** Alignment robustness assessment across different anatomical orientations (coronal sections) of the mouse brain, measured by PCC and CI. **(E)** 3D reconstruction of original (left) and stMSA-aligned (right) spatial coordinates, demonstrating accurate spatial preservation. **(F)** Original coronal mouse brain orientation, stMSA-derived spatial domains, and pyramidal cell layer marker gene spatial distribution.

Efficient integration of spatial omics data for joint domain detection, matching, and alignment with stMSA

Han Shu, Jing Chen, Chang Xu, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2025/09/12/gr.280584.125.DC1
P<P	Published online August 14, 2025 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see https://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**