A Correspondence-Driven Framework for Un-paired Spatial Multi-Omics Integrative Analysis

Wenhao Cai[1,2,3], Weizhong Li[1,2,*]

[1] School of Medicine, Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University, Guangming District, Shenzhen, Guangdong 518107, China

[2] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong 510080 China

[3] Department of Computer Science, University of Manchester, Manchester, United Kingdom

* Corresponding author: Weizhong Li (liweizhong@mail.sysu.edu.cn)

ORCID: Weizhong Li (0000-0002-9003-7733)

Abstract

Recent advances in spatial multi-omics technologies provide unprecedented opportunities to interpret molecular features in tissue micro-environments but remain challenging in integrative analysis across heterogeneous datasets. Here we present SpatialFuser, a correspondence-driven deep learning framework for integrative analysis across un-paired spatial epigenomics, transcriptomics, proteomics, and metabolomics. SpatialFuser introduces a Multi-head Collaborative Graph Attention auToEncoder (MCGATE) to infer multi-scale cellular correspondences for fine-grained characterization of spatial heterogeneity beyond predefined spatial neighbourhoods. By incorporating flexible geometric pre-matching for coarse initialization and inferring adaptive cross-slice correspondences via iteratively refined optimal transport, SpatialFuser enables robust integration across heterogeneous datasets with varying geometries, spatial resolutions, developmental stages, and molecular modalities. Benchmarking demonstrates superior performance and reliability against existing state-of-the-art methods in spatial domain identification, cross-slice alignment, and multi-omics integration. Applications to real datasets demonstrate that SpatialFuser resolves

precise molecular patterns, reveals developmental dynamics, and recovery of complementary signals across modalities. Cross-resolution integration of weakly correlated modalities by our method further uncovers previously obscured biological variation. Our framework is generalizable and versatile, enabling customized analytical scenarios and potential extension for emerging omics.

Keywords: SpatialFuser, spatial multi-omics, cross-modality integration, graph neural networks, contrastive learning

Highlights:

- A unified deep learning framework for spatial multi-omics integrative data analysis
- Superior performance against state-of-the-art methods in spatial identification，alignment, and multi-omics integration
- Unprecedented cross-modality analysis scenarios to offer a holistic view of spatial multi-omics
- Comprehensive framework design with generalizability and versatility for customized scenarios and potential extension

Introduction

Spatial technologies enable researchers to measure expression levels and reveal cellular heterogeneity across multiple molecular layers such as transcriptomics, epigenomics, proteomics, and metabolomics at specific spatial locations, and have been widely applied in studies of tissue organization and molecular dynamics[1]. These technologies can be roughly categorized into imaging-based approaches including osmFISH[2], MERFISH[3], CODEX[4], MIBI[5], CosMx SMI[6], and MALDI-MSI[7], and sequencing-based techniques such as 10X Visium[8], BaristaSeq[9], Stereo-seq[10], MAGIC-seq[11], Stereo-CITE-seq[12], Spatial ATAC-seq[13], and

Spatial ATAC–RNA-seq[14]. Recent advances also cover AI-powered techniques such as PLATO[15], which leverage computational approaches to overcome current limitations in biotechnology, such as low sequencing depth. With different data modalities conveying unique biological insights from divers e complementary perspectives, we now have an unprecedented opportunity to achieve a comprehensive spatial landscape of cellular profile, thereby enhancing our understanding of key processes such as cell function[16], tissue development[17], and disease progression[18].

Computational methods have been developed to decipher intra-slice heterogeneity from typically spatial transcriptomics data with high-dimension and high-sparsity through spatial domain detection. SpaGCN[19] employs a graph convolutional network to integrate spatial gene expression with coordinate information for spatial domain identification. STAGATE[20] adopts a spatially-aware graph attention autoencoder to capture spatial structures, while SEDR[21] uses a deep variational autoencoder to jointly learn gene expression and spatial relationships. These unimodal methods model spatial relationships through predefined neighbourhoods, implicitly assuming that spatial proximity reflects biological similarity, an assumption that often breaks down in contexts such as tumour microenvironments, immune infiltration, and developmental migration. Meanwhile, the growing expansion of spatial technologies into multi-omics indicates that increasing numbers of tissue slices will be profiled from proteomic, epigenomic, and other molecular perspectives, underlying this is an urgent need for unified computational tools capable of decoding spatial data from diverse omics modalities.

To achieve a deeper understanding of complex biological systems, integrative analysis of cross-modality spatial data has become increasingly important. However, heterogeneous datasets often suffer from batch effects or modality biases, resulting in discrepancies in feature structures, expression distributions, and biological semantics, posing challenges for effective data integration and slice alignment. Existing

computational tools attempt to address these issues but remain limited, particularly for heterogeneous multi-omics scenarios. For example, SIMO[22] employs probabilistic alignment to map single-cell multi-omics data onto homologous spatial transcriptomics slices for detecting spatial cellular topological patterns, but it requires paired datasets and is restricted to within-slice analysis. STAligner[23] employs a graph attention autoencoder and triplet adversarial learning to align spatial transcriptomic slices, but it does not extend to multi-omics integration. SLAT[24] uses graph convolutional networks and adversarial learning to reconstruct 3D tissues while correcting batch effects across slices, yet it fails to integrate modalities with weakly correlated features, such as transcriptomics and proteomics.

In the presence of substantial feature semantic discrepancies across modalities, explicitly modeling feature dependencies becomes challenging. MISO[25] addresses this by learning modality-specific embeddings and their interactions to identify biologically relevant spatial domains, but it cannot perform cross-slice integration. Similarly, SpatialGlue[26] and MultiGATE[27] apply dual-attention graph models to jointly analyse spatial multi-omics from the same tissue, yet they heavily rely on shared spatial coordinate systems or hard-coded biological priors to link weakly correlated features. These integrative methods are effective for unimodal or same-slice multi-omics data but lack generalizability in heterogeneous multi-omics scenarios. Methodologically, many integrative tools assume consistent feature structures across datasets, requiring dimensionality matching through preprocessing. This can lead to information loss when the intrinsic complexity of raw modalities differs. Some also assume similar cell compositions across slices and adopt rigid integration strategies, increasing the risk of over-alignment. Therefore, new computational tools are urgently needed for spatially aware cross-omics integration that can offer better biological interpretability and generalizability through more flexible and robust analysis across diverse and emerging spatial omics modalities.

Here, we introduce SpatialFuser, a unified correspondence-centric deep learning framework for un-paired spatial multi-omics data integrative analysis, enabling accurate spatial interpretation, effective cross-slice alignment, and robust cross-modality integration. Instead of relying on predefined similarity metrics or spatial neighbourhood structures, SpatialFuser treats correspondences as guidance for model learning. Specifically, it employs graph neural networks to learn intra-slice multi-scale cell–cell correspondences in a latent space, capturing fine-grained spatial heterogeneity, and incorporates flexible geometric pre-matching for coarse initialization under geometric mismatch. By iteratively inferring cross-slice correspondences via soft optimal transport and coupling them with contrastive representation learning, SpatialFuser jointly refines cross-slice correspondence estimation and modality fusion, enabling robust integration across heterogeneous and cross-resolution datasets.

We benchmarked SpatialFuser on classical 10X Visium datasets to evaluate its quantitative accuracy and robustness in spatial domain identification, and then validated its cross-platform and cross-modality generalizability on osmFISH and CODEX datasets. In unimodal multi-slice joint analysis, SpatialFuser better conducted adjacent BaristaSeq slices integration and alignment than state-of-the-art methods reported in previous benchmarks[28], and effectively captured fine-grained developmental dynamics through integrative analysis of multi-developmental-stage spatial Stereo-seq samples. In spatial multi-omics scenarios, SpatialFuser outperforms existing methods in resolving subtle spatial patterns of cellular states and enhances the detection of lowly expressed but epigenetically primed marker genes by integrating complementary information from spatial ATAC–RNA-seq data, thereby revealing lineage-associated regulatory signals. In integrative analysis across multi-omics samples with weakly correlated features, SpatialFuser effectively integrates transcriptomic (MAGIC-seq), proteomic (PLATO), and metabolomic (MALDI-MSI) data across varying spatial resolutions to improve the fine-grained characterization of spatial topological patterns. These results highlight the capability of SpatialFuser to

infer cell state transitions and transcriptional readiness within spatial tissue architectures, and to identify fine-grained functional regions that are challenging to resolve with low-resolution technologies alone. SpatialFuser demonstrates versatility and generalizability in decoding complex biological systems through comprehensive spatial multi-omics integration.

Results

The overview architecture of SpatialFuser

SpatialFuser takes molecular features and spatial coordinates from spatial omics data as input, enabling both fine-grained single-slice analysis and cross-sample integrative analysis across modalities including epigenomics, transcriptomics, proteomics, and metabolomics (Fig. 1a). SpatialFuser is a unified framework designed to support diverse tasks, such as accurate spatial interpretation, robust cross-slice alignment, and effective cross-modality integration (Fig. 1b). It first represents the input as graph-structured data by constructing a spatial adjacency network based on spatial coordinates, and employs a Multi-head Collaborative Graph ATtention autoEncoder (MCGATE) to learn modality-specific embeddings through a reconstruction task (Fig. 1b Step 2 & Fig. 1c). In the encoder, MCGATE uses the predefined correspondence graph to guide local attention aggregation, while also constructing a feature similarity graph in the low-dimensional latent space to capture long-range feature correspondences. This similarity graph is fed back into the encoder to dynamically refine local attention weights and establish long-range attention. The collaborative attention mechanism enables the integration of multi-scale spatial context and semantic information, allowing SpatialFuser to more effectively learn spatial molecular patterns. As a general embedding model for spatial omics data, MCGATE is designed to be modular and scalable, enabling SpatialFuser to be broadly applicable across diverse modalities and experimental platforms.

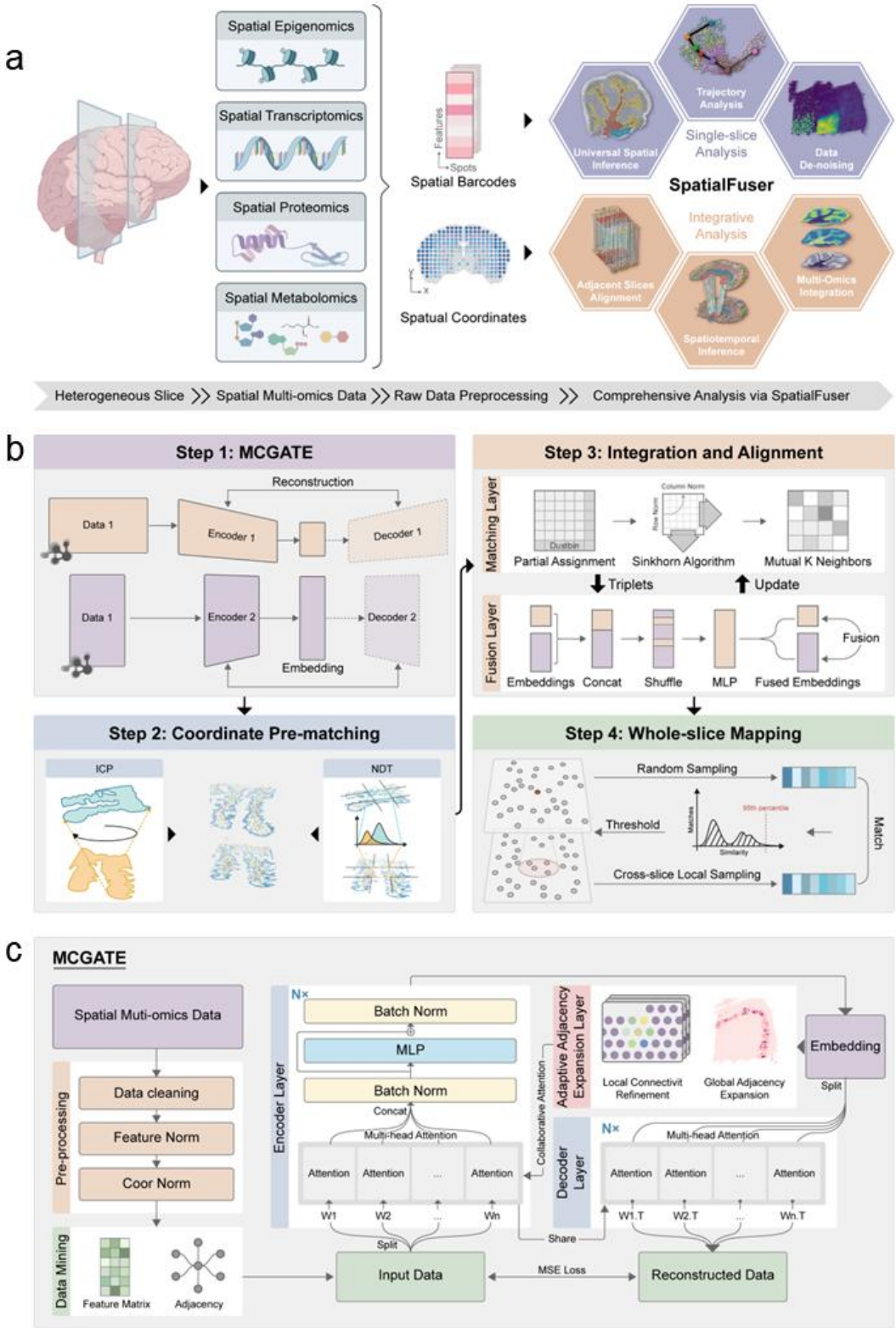

Fig. 1. A unified deep learning framework for single-sample and cross-slice spatial multi-omics analysis. a, Overview of SpatialFuser. SpatialFuser integrates molecular features and spatial coordinates from spatial omics data, supporting both fine-grained single-slice profiling and cross-sample analysis across modalities such as epigenomics, transcriptomics, proteomics, and metabolomics. **b,** Schematic

representation of the SpatialFuser framework. The analysis workflow includes: MCGATE-based embedding learning, coordinate pre-matching for rigid slice registration, dual matching and fusion layers for multimodal integration, and whole-slice spot mapping for downstream integrative analysis. **c,** Network architecture of MCGATE. MCGATE represents spatial omics data as a graph by constructing a spatial adjacency network from coordinates. It applies multi-head collaborative graph attention to learn modality-specific embeddings via a reconstruction task, guided by a spatial graph for local attention and a dynamically updated feature similarity graph for long-range attention. Designed to be modular and scalable, MCGATE can be applied across diverse spatial omics modalities and platforms.

The initial pre-matching of coordinates across tissue slices is an essential step for the alignment of spatial omics data[29]. To accommodate diverse spatial distributions, SpatialFuser models spatial expression data as 2D point clouds and offers two geometric registration algorithms: Iterative Closest Point (ICP) and Normal Distributions Transform (NDT) (Fig. 1b Step 2). ICP is well suited for datasets with regularly spaced spots and clear structural grids (e.g., 10X Visium), while NDT performs better on datasets with densely packed, irregularly distributed spots and indistinct boundaries (e.g., CODEX and Stereo-seq). This flexible pre-matching strategy provides robust initialization for subsequent fine-grained alignment.

Alignment and integration are not strictly independent tasks[28]. To support cross-slice analysis, SpatialFuser jointly addresses these two tasks through an iteratively trained dual-layer architecture consisting of a matching layer based on the Sinkhorn algorithm and a fusion layer grounded in contrastive learning (Fig. 1b Step 3). In the matching layer, the alignment problem is formulated as a graph matching task and solved as an optimal transport problem via the Sinkhorn algorithm. To better support cross-resolution alignment and prevent over-alignment, a dustbin channel mechanism[30] is incorporated to filter out unmatched or ambiguous points. In the fusion layer, SpatialFuser employs a contrastive learning strategy, where high-confidence cell-cell correspondences from the matching layer serve as anchor-positive pairs to guide the correction of batch effects and modality biases, and randomly

sampled anchor-negative pairs from non-neighbouring regions within the same slice to preserve the model's sensitivity to true biological variation. SpatialFuser introduces an absolute distance constraint into the triplet loss to enhance model stability and convergence[31], ensuring accurate and consistent mapping of spatial data into a shared embedding space. Finally, SpatialFuser conducts rigorous quality control to achieve whole-slice spot mapping (Fig. 1b Step 4), providing a reliable reference for downstream integrative analyses.

In brief, SpatialFuser consists of four key components: MCGATE, a coordinate pre-matching module, a matching layer based on the Sinkhorn algorithm, and a fusion layer driven by contrastive learning. Each component can operate independently or seamlessly integrate into a unified workflow, providing a comprehensive and streamlined solution for downstream analysis of spatial multi-omics data across diverse experimental platforms and resolutions.

Accurate spatial inference across diverse experimental platforms and modalities

SpatialFuser proposes a unified tool for spatial multi-omics inference, enabling accurate detection of spatial distributions of tissue domains or cell types. To evaluate the performance of SpatialFuser against existing spatial domain detection algorithms, we conducted a benchmarking using the classic 10X Visium[8] human dorsolateral prefrontal cortex (DLPFC) dataset[32], which has well-defined morphological boundaries and reliable manual annotations (Fig. 2a & Supplementary Fig. 1). The original labels were used as ground truth, and five evaluation metrics were employed: Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI), Homogeneity, Completeness, and V-Measure. We first compared SpatialFuser with other state-of-the-art spatial domain detection methods based on the Mclust[33] clustering strategy, including SpaGCN[19], STAGATE[20], SEDR[21], and STAligner[23] (Fig. 2b). On average across the 12 DLPFC slices, SpatialFuser achieved the highest accuracy (mean

ARI = 0.625; max = 0.809; min = 0.533), followed by STAligner (mean ARI = 0.556; max = 0.675; min = 0.469). Notably, although SpaGCN integrates high-resolution histological images during training, its performance was substantially lower than that of other methods relying solely on molecular and spatial information (mean ARI = 0.422; max = 0.541; min = 0.225), suggesting that the utility of histological information is highly dependent on the modeling and integration strategy, rather than on its inclusion alone. To further evaluate SpatialFuser's advantages in downstream analysis, we also benchmarked it using commonly used clustering methods in omics data analysis, including Leiden[34] (Supplementary Fig. 2a) and Louvain[35] (Supplementary Fig. 2b). These results demonstrate that SpatialFuser enables accurate spatial domain identification and consistently outperforms existing state-of-the-art tools across diverse clustering contexts.

We conducted an ablation study to evaluate the effectiveness of the collaborative attention mechanism. The results demonstrate that the long-range attention significantly enhances the SpatialFuser's ability to capture spatial distribution patterns (Fig. 2b, Supplementary Fig. 2a-b). Using slice-151676 from the DLPFC dataset as an example, we randomly selected 10 spatial spots from cortical layers 3 and 6 and tracked their long-range correspondence sampling during training. The distribution of sampled spots closely aligned to the tissue region patterns (Fig. 2c), further validating the effectiveness and biological relevance of long-range information propagation. We also tested the robustness of SpatialFuser by comparing clustering accuracy across different hyperparameters (see Methods), demonstrating that the model is insensitive to network structure and random seeds (Supplementary Fig. 3a-b).

Spatial omics data generated from different experimental platforms and modalities exhibits significant differences in feature structures, distributions, and biological semantics. We further evaluated SpatialFuser's ability to identify spatial domains across multiple platforms and modalities using the osmFISH somatosensory cortex dataset[2] and the CODEX[4] human muscle-invasive bladder cancer tumor dataset[36].

On the image-based osmFISH dataset, SpatialFuser accurately identified spatial domains in the somatosensory cortex with an ARI of 0.61 (Fig. 2d). For the CODEX spatial proteomics dataset, we re-labelled the three main cell types, epithelial cells, stromal cells, and immune cells, mentioned in the original paper based on the annotations provided by the authors for downstream spatial inference. SpatialFuser successfully captured the spatial distribution patterns of all the three key cell types (Fig. 2e, ARI = 0.45) The relatively lower ARI score may be attributed to the higher spatial heterogeneity of tumour tissues[37]. These results demonstrate the robustness and versatility of SpatialFuser in cross-platform, multi-omics spatial data analysis.

To evaluate SpatialFuser's utility in downstream analysis, we assessed its low-dimensional embeddings for trajectory inference and reconstructed feature for spatial variable gene (SVG) identification. Partition-based graph abstraction[38] (PAGA) was employed to infer trajectories on DLPFC slice-151675 using embeddings learned by SpatialFuser, SEDR, and STAligner. In the UMAP plot (Fig. 2f), SpatialFuser's embedding closely recapitulated the developmental progression from white matter to layer 6 and sequentially to layer 1. Comparatively, embeddings from SEDR and STAligner exhibited less plausible developmental trajectories. We further compared the expression patterns of three region-enriched marker genes between the raw data and SpatialFuser-denoised data in DLPFC slice-151675. After reconstruction, all selected genes showed clear differential expression in their enriched regions, while the raw spatial expression was noisy and inconsistent (Fig. 2g). Collectively, these results demonstrate that SpatialFuser can accurately capture feature expression patterns and reduce noise, thereby better supporting single-sample analysis and spatial heterogeneity studies.

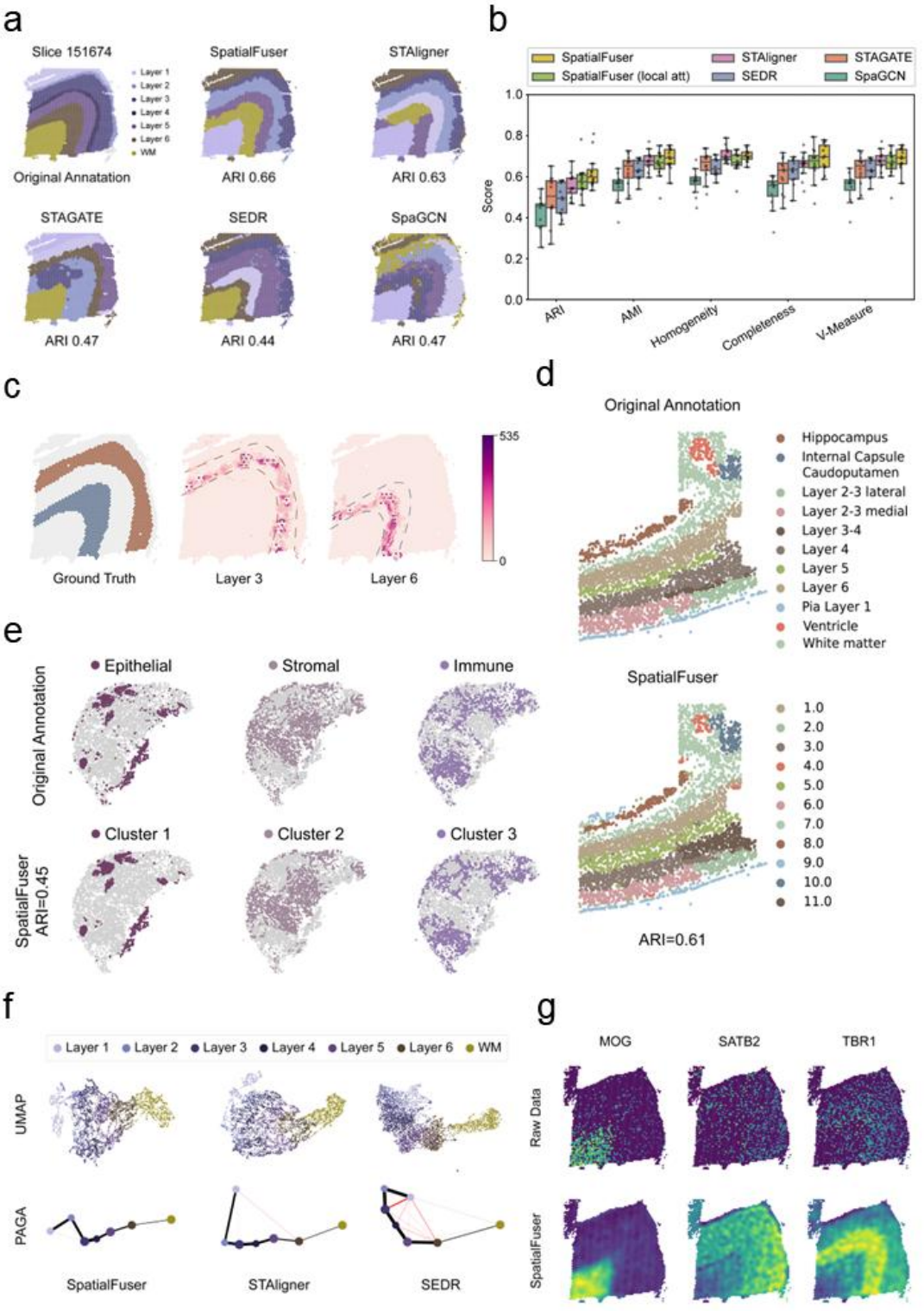

Fig. 2. SpatialFuser accurately resolves tissue domains across diverse experimental platforms and modalities. a, Comparison of spatial domains identified by SpaGCN, STAGATE, SEDR, STAligner, and SpatialFuser for DLPFC slice-151674. **b,** Boxplots of five evaluation metrics (ARI, AMI, Homogeneity, Completeness, V-measure) for Mclust clustering results across 12 DLPFC sections. **c,**

Long-range feature propagation paths of ten randomly selected spots from cortical layers 3 and 6 during MCGATE training on DLPFC slice-151676. **d,** Spatial domains identified by Mclust clustering on low-dimensional embeddings generated by SpatialFuser in the osmFISH somatosensory cortex dataset. **e,** Distributions of epithelial, stromal, and immune cells identified by Mclust clustering on low-dimensional embeddings generated by SpatialFuser for slice-210308_TMA2_reg6 of the CODEX muscle-invasive bladder cancer dataset. **f,** UMAP visualizations and PAGA graphs generated from SEDR, STAligner, and SpatialFuser embeddings for DLPFC section-151675. Red lines highlight incorrect trajectory inferences. **g,** SpatialFuser enhances spatial patterns of layer-enriched and marker genes in the DLPFC dataset. Raw and SpatialFuser-denoised spatial expression patterns are shown for MOG, SATB2, and TBR1 in DLPFC section-151675. ARI, adjusted Rand index; AMI, adjusted mutual information; DLPFC, dorsolateral prefrontal cortex; MCGATE, multi-head collaborative graph attention autoencoder.

Spatiotemporal alignment for revealing tissue heterogeneity and developmental dynamics

Integrative and comparative analysis of spatial omics datasets is critical for unravelling spatial complexity and temporal changes in tissue systems[39]. To assess the performance of SpatialFuser in pairwise slice alignment and integration, we first applied it to two adjacent slices from the BaristaSeq mouse visual cortex dataset[9], comprising 1525 and 2042 spots, respectively (Fig. 3a). We evaluated spot-to-spot alignment accuracy across the entire tissue section (see Methods), comparing SpatialFuser with three state-of-the-art methods reported in previous benchmarks[28], including SPACEL[40], SLAT[24], and STAligner[23] (Fig. 3b). Owing to its use of ground-truth labels during training (see Methods), SPACEL achieved the highest alignment accuracy (0.987, 1519 matches), followed closely by SpatialFuser (0.972, 1516 matches) and STAligner (0.949, 1508 matches), while SLAT showed noticeably lower performance (0.857, 1525 matches).

An effective integration method should not only achieve accurate alignment but also preserve underlying biological patterns in the latent space. To this end, we further evaluated the spatial fidelity of the integrated embeddings. UMAP visualization showed that all four methods effectively removed batch effects (Fig. 3c). With original annotations used as supervision, SPACEL maintained strong region-level separation (Supplementary Fig. 4), but its embedding space appeared fragmented (Supplementary Fig. 5), suggesting limited intra-domain coherence despite minimal cross-domain mixing. STAligner and SLAT preserved the hierarchical structure of the six cortical layers to some extent, but layer boundaries were blurred and showed considerable mixing between adjacent layers. Compared to these methods, SpatialFuser produced embeddings with clearer expression patterns showed in UMAP (Fig. 3c & Supplementary Fig. 5) and more hierarchically organized, anatomically consistent domain structures that closely matched the ground truth (Supplementary Fig. 4).

To quantitatively compare the effectiveness of these methods in supporting integrative spatial heterogeneity analysis, we performed clustering on the joint embeddings from each pair of BaristaSeq slices using the Mclust algorithm, focusing primarily on the ARI before and after alignment and integration. Since SLAT and SPACEL operates only in multi-slice mode, ARI scores before integration were left unpopulated. The bar plot (Fig. 3d) shows that SpatialFuser demonstrated the best performance for post-integration, and the integration process enabled more accurate capture of spatial patterns, as clustering on the integrated embeddings yielded spatial domains more consistent with the ground truth (average ARI=0.86).

We further evaluated SpatialFuser's efficacy in integrating non-consecutive slices from different developmental stages, which exhibit significant differences in cell-type composition and distribution. We employed a Stereo-seq[12] dataset representing axolotl regenerative telencephalon[10], which has manual annotations for each tissue region as reference, and applied SpatialFuser to align and integrate two slices from stage 54 and stage 57 (Fig. 3e). Through alignment, most spots were accurately matched

despite substantial variation in spot numbers and types across slices (Fig. 3f), and dynamic changes in cell identities during development were successfully captured (e.g., mature nptxEX at stage 57 could be traced back to immature nptxEX at stage 54). In parallel, SpatialFuser effectively corrected batch effects while preserving expression pattern differences among distinct cell-types (Fig. 3g).

Interestingly, a large region of development-related ependymoglial cells (dEGCs) was enriched with less-aligned cells during the iterative training process of SpatialFuser (Supplementary Fig. 9). We hypothesize that this may be attributed to intra-regional heterogeneity within the seemingly homogeneous dEGCs domain. For instance, previously unannotated substructures may have emerged during telencephalon development from stage 54 to stage 57, reducing cross-slice similarity within this region and thereby hindering the formation of high-quality correspondences. To support this hypothesis, we performed Mclust clustering on the integrated embeddings and identified a distinct subpopulation within the dEGCs domain (Cluster 17) (Fig. 3h), likely reflecting emerging subregional or functional differentiation. Further spatial differential analysis confirmed that this subregion exhibits a unique spatial expression pattern, with FGF17, STRA6, and SLIT2 significantly enriched (Fig. 3i). These genes are known to play key roles in embryonic development and neurodevelopmental processes[41-44], aligning well with the spatial distribution and potential functions of the subregion. Collectively, these findings demonstrate SpatialFuser's capability to reveal dynamic cellular changes through spatiotemporal integrative analysis.

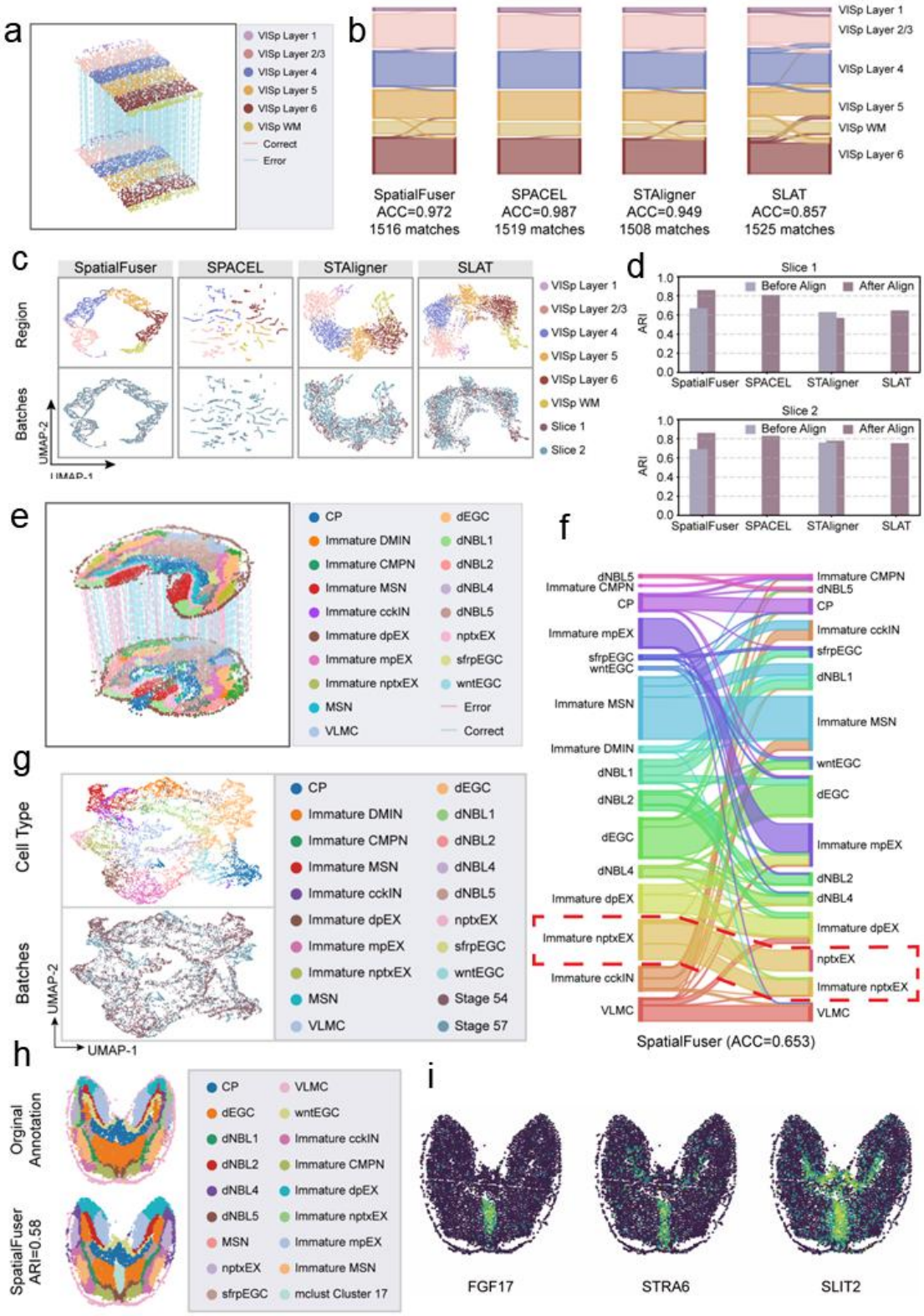

Fig. 3. Spatial alignment across consecutive tissue slices and non-consecutive slices from different developmental stages. a, Alignment of slices 1 and 2 from the BaristaSeq mouse visual cortex dataset, coloured by region labels (300 alignment pairs shown for clarity). **b,** Sankey plot showing region type correspondence based on alignments from SpatialFuser, SPACEL, STAligner, and SLAT for slices 1 and 2 of

the BaristaSeq dataset. **c,** UMAP visualizations of integration results for SpatialFuser, SPACEL, STAligner, and SLAT on slices 1 and 2 of the BaristaSeq dataset, coloured by region labels (top) and batch labels (bottom). **d,** ARI scores before and after integration for domain identification in slices 1 and 2 of the BaristaSeq dataset. **e,** Alignment of Stage 54 and Stage 57 slices from the Stereo-seq axolotl regenerative telencephalon dataset, coloured by region labels (300 alignment pairs shown for clarity). **f,** Sankey plot showing region type correspondence based on alignments from SpatialFuser for Stage 54 and Stage 57 slices of the Stereo-seq dataset. **g,** UMAP visualizations of integration results for SpatialFuser on Stage 54 and Stage 57 slices of the Stereo-seq dataset, coloured by region labels (top) and batch labels (bottom). **h,** Spatial domains in the Stage 57 slice of the Stereo-seq dataset identified by Mclust clustering on SpatialFuser-integrated embeddings. **i,** Expression distributions of FGF17, STRA6, and SLIT2, specifically enriched in a potential subregional structure within the dEGCs region of the Stage 57 slice from the Stereo-seq dataset.

Effective and efficient spatial multi-omics integration enables complementary tissue profiling

Incorporating multiple modalities into the analysis of biological samples holds great potential for deepening our understanding of the mechanisms underlying cellular and tissue organization[45]. To evaluate the multi-omics integration capability of SpatialFuser, we applied it to spatial ATAC–RNA-seq data from an embryonic day 13 (E13) mouse embryo section (Fig. 4a), in which the telencephalon region was annotated based on marker genes for the cortical plate, ventricular zone, and striatal primordium (Fig. 4b), and compared its performance with current state-of-the-art methods. Leiden clustering of the integrated embeddings from all models recovered the Eye region (cluster 9 in both RNA and ATAC Leiden clustering) (Fig. 4c). However, only SpatialFuser and SpatialGlue preserved the layered organization of the hindbrain with high structural fidelity. Focusing on the dorsolateral telencephalon in the forebrain, SpatialFuser and MISO uniquely distinguished the developing pallial cortical plate from the pallial ventricular zone. This neuronal population was previously shown to be difficult to resolve using either modality alone, as demonstrated by joint epigenomic and transcriptomic profiling in a prior study [14]. During embryonic telencephalon

development, key transcription factors exhibit spatial gradients rather than sharply defined boundaries [17, 46]. To quantify the spatial concordance between identified CP clusters and CP marker enrichment, we evaluated cluster–marker agreement using purity and coverage metrics across a range of marker-enriched spot definitions (see Methods). Although the cortical plate cluster identified by MISO shows near maximal marker purity (Fig. 4d, left), it is largely restricted to localized expression hotspots. In contrast, SpatialFuser delineates a broader and spatially continuous region, maintaining a high level of marker purity while achieving substantially higher coverage score (Fig. 4d, right), consistent with the gradual transitions expected across developing cortical territories. By comparison, MultiGATE produces less distinct regional boundaries despite achieving overall modality alignment (Fig. 4c and Supplementary Fig. 10), whereas SIMO fails to clearly resolve the original anatomical domains even when label information is incorporated during training (Fig. 4c). We also observed that increasing the number of mapped cells in SIMO leads to increasing spatial disorganization and reduced domain consistency (Supplementary Fig. 11).

At embryonic day 13, chromatin accessibility may not fully resolve all the cell types defined by transcriptional profiles, potentially due to lineage priming[47, 48], as some genes exhibit differing spatial signal patterns between ATAC and RNA modalities. For instance, Vax2 (ventral retina marker[49]), Kcnj10 (ventricular zone and astrocyte progenitor marker[50]), Pax6 (cortical progenitor and optic vesicle marker[51]), and Myt1l (hindbrain neuron marker[52]) displayed relatively low RNA expression in their respective regions at stage E13 in the mouse embryo, despite marked chromatin accessibility. For methods that reconstruct omics features, we further evaluated their ability to capture complementary expression patterns by integrating chromatin accessibility with transcriptomic signals. Compared with MultiGATE and SIMO, SpatialFuser preferentially enhances the detection of weakly expressed yet epigenetically informative genes (Fig. 4e). However, it is important to note that while SpatialFuser functions as an integrative enhancement tool that leverages

complementary information from different modalities to reconstruct a more comprehensive and biologically meaningful gene expression landscape, it does not explicitly differentiate whether low RNA expression arises from true lineage priming or technical limitations in transcript detection. Disentangling these causes requires further experimental validation.

As spatial multi-omics datasets continue to increase in scale and complexity, computational efficiency becomes increasingly important for integration methods. We therefore compared training runtime across methods using the E13 mouse embryo spatial ATAC RNA-seq dataset (see Methods), and the results show that SpatialFuser required less training time than all other approaches (Fig. 4f). Of note, MultiGATE is implemented in legacy software environments and built on an early version of the TensorFlow framework, which posed practical challenges for deployment on our GPU systems. Therefore, MultiGATE was trained on CPU, with each run requiring approximately 1.5 to 2 hours. Together, our findings indicate that SpatialFuser can effectively capture molecular signals associated with transcriptional readiness within complex tissue contexts that remained difficult to resolve using unimodal or earlier integration approaches. This capability facilitates the identification of emerging cell subtypes or transitional cellular states relevant to developmental processes and early disease progression.

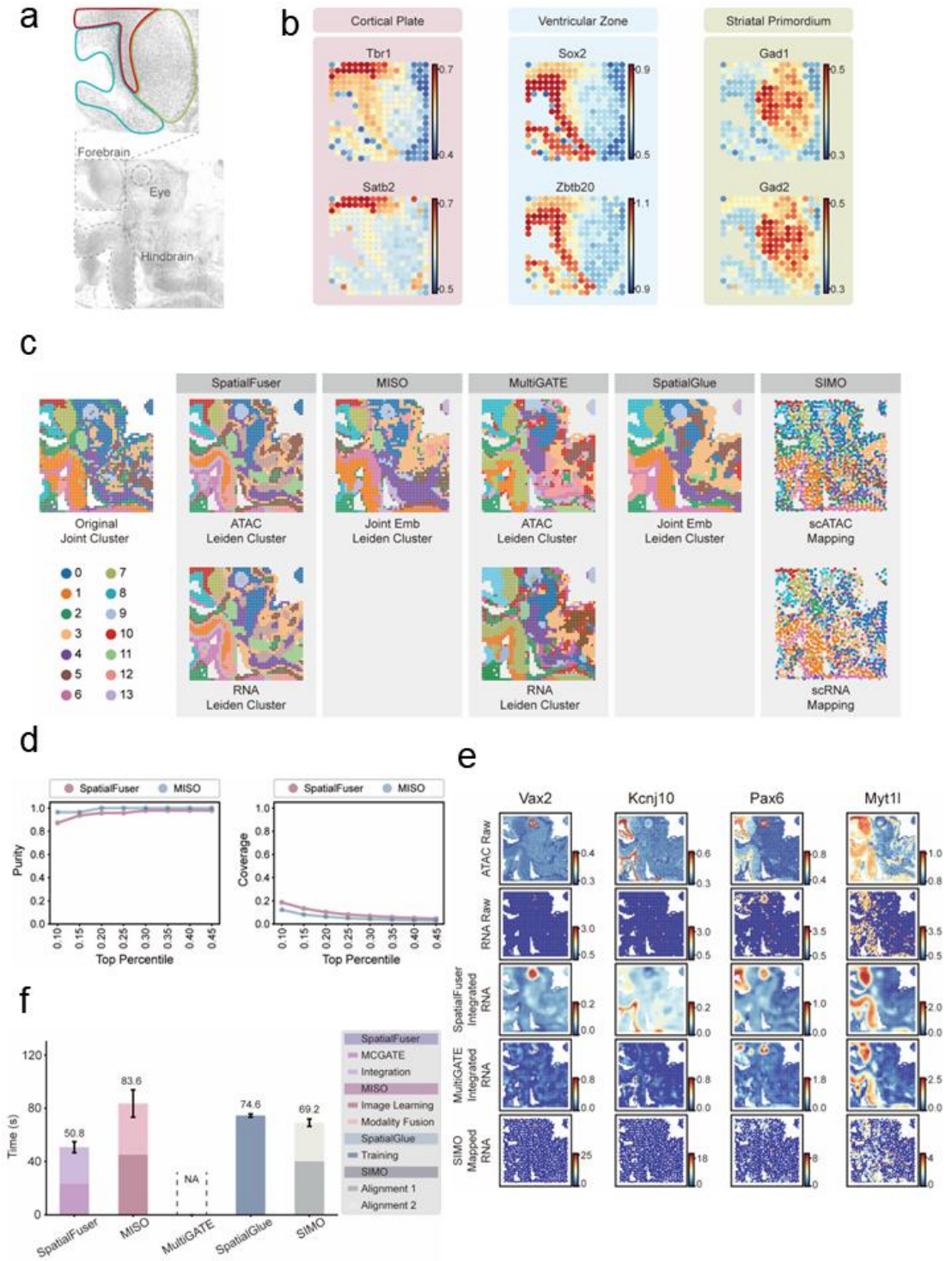

Fig. 4. Fast and precise integrative analysis of epigenome–transcriptome co-profiled mouse embryo slice. a, H&E-stained histological image of spatial ATAC–RNA-seq data from an E13 mouse embryo, with magnified and annotated views of the telencephalon. **b,** Spatial gene expression patterns of marker genes for the cortical plate, ventricular zone, and striatal primordium in the telencephalon. **c,** Spatial domains identified by Leiden clustering on integrated embeddings or mapped features generated by SpatialFuser, MISO, MultiGATE, SpatialGlue, and SIMO using the E13

mouse embryo spatial ATAC–RNA-seq dataset. SpatialFuser, MultiGATE, and MISO produce modality-specific representations, whereas SpatialGlue and SIMO learn joint representations across modalities. **d,** Purity–Coverage analysis of cortical plate marker enrichment under varying percentile thresholds. **e,** Spatial expression patterns of Vax2, Kcnj10, Pax6, and Myt1l in the original ATAC-derived gene activity, original RNA expression, and reconstructed or mapped RNA profiles produced by SpatialFuser, MultiGATE, and SIMO after integration. **f,** Computational efficiency comparison of methods evaluated on the E13 mouse embryo spatial ATAC–RNA-seq dataset.

Cross-resolution integration of weakly correlated spatial modalities advances tissue-level information

Compared to the one-to-one correspondence between the genome and transcriptome, the linkage between the proteome or metabolome and the transcriptome or epigenome is generally weak[53], as protein abundance and metabolite levels are influenced by multiple regulatory processes beyond gene expression, such as translation, post-translational modifications, and cellular environment[54, 55]. To explore cross-modality integration under such weakly correlated conditions, we applied SpatialFuser to integrate and align transcriptomic (MAGIC-seq[11]), proteomic (PLATO[15]), and metabolomic (MALDI-MSI[7]) data under varying resolutions obtained from consecutive mouse cerebellum tissue slices. Result shows SpatialFuser effectively captured shared spatial molecular patterns across different omics layers and successfully corrected modality biases while preserving true biological variation, as evidenced by the spatial domains identified through Louvain clustering of the latent embeddings, which aligned well with the original anatomical annotations (Fig. 5a). Most spots across the tissue were well aligned, with high alignment similarity scores (Supplementary Fig. 13) and matching accuracy (Supplementary Fig. 14, Transcriptome–Proteome: 0.84, Transcriptome–Metabolome: 0.75, Proteome–Metabolome: 0.86). It is noteworthy that manual annotation is affected by the granularity and distribution discrepancies across modalities. For example, although MAGIC-seq transcriptomic data and PLATO proteomic data share the same spatial

coordinates, the matching rate of their original labels is only 83%, which undermines the reliability of using label-matching accuracy as a metric to evaluate spot-to-region mapping performance. By visualizing the embeddings with UMAP, we further demonstrate SpatialFuser's ability to accurately align the same cell types across different omics modalities, as spots of the same type are consistently co-localized in the embedding space, even under weak cross-modality cellular correspondence (Fig. 5b).

Due to differences in spatial resolution and barcode density, the MALDI-MSI slice captured 3908 spots, which is significantly more than the 1677 spots in the transcriptomic and proteomic datasets, thereby enabling finer-grained tissue domain annotations. Through cross-modality and cross-resolution integration and alignment, SpatialFuser is able to leverage the detailed histological distribution patterns in the metabolomic data to enhance the delineation of fiber tracts in the spatial transcriptomic data, which were previously difficult to resolve (Fig. 5c).

To assess whether the reconstructed transcriptomic profiles preserved biologically relevant spatial patterns, we evaluated the ability of known canonical oligodendrocyte marker genes (Cldn11[56], Mog[57], and Cnp[58]) to discriminate the originally annotated fiber tracts before and after integration. ROC curve analysis showed that the reconstructed features achieved higher AUC scores in distinguishing the fiber tracts defined by the original annotation (Fig. 5d), indicating improved concordance with expected spatial expression patterns. Notably, when using Louvain-identified Fiber tracts cluster as the reference, we observed even higher AUC scores (Fig. 5e). The strong spatial concordance between the reconstructed marker gene expression patterns and the refined anatomical delineation supports the biological validity of the finer-grained structure revealed by the SpatialFuser integration.

Together, these results demonstrate the versatility of SpatialFuser in decoding complex biological systems through spatial multi-omics integration. By bridging modalities with different biological meanings and technical characteristics,

SpatialFuser enables deeper and more comprehensive tissue profiling across resolutions, offering a powerful tool for spatial systems biology.

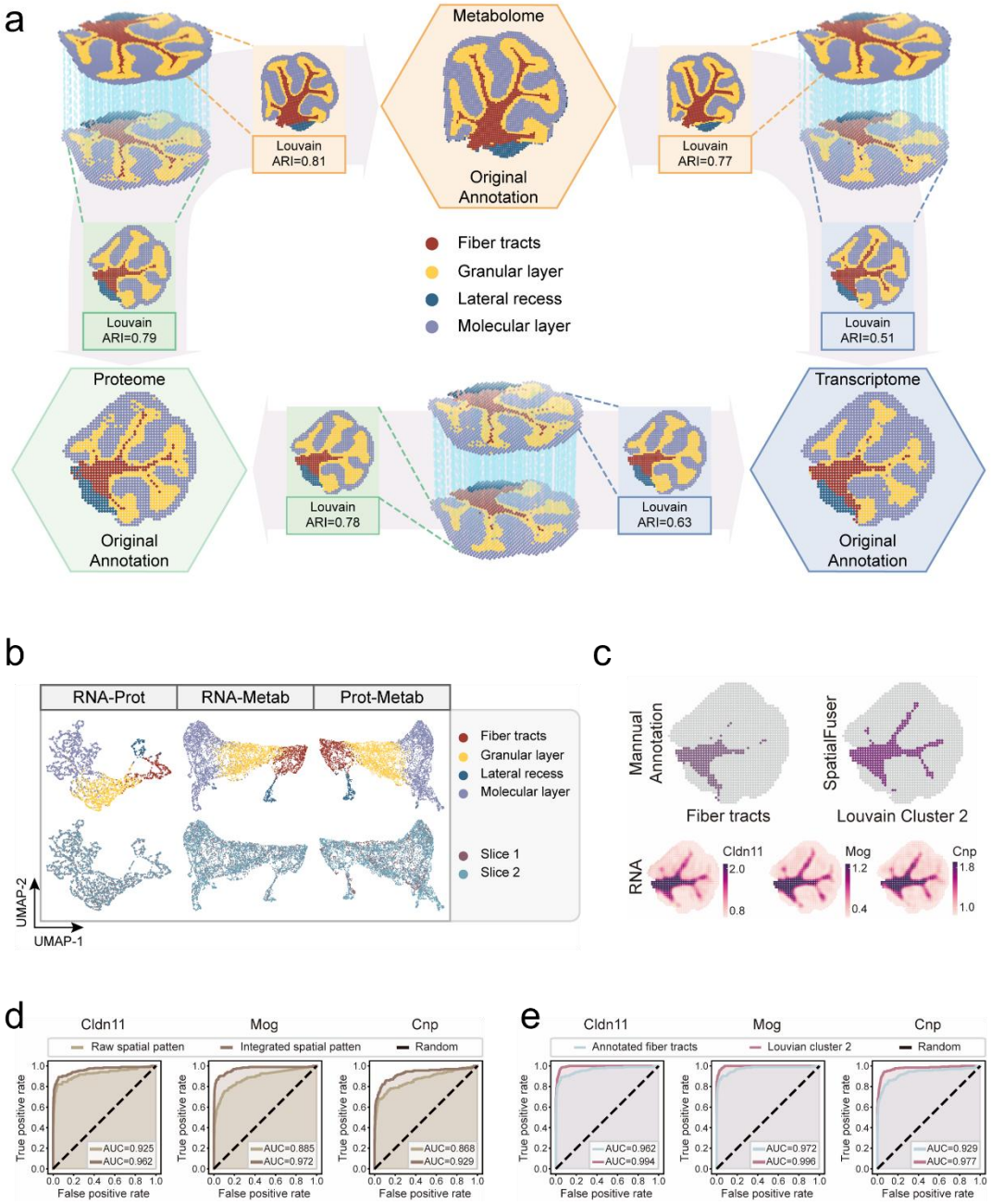

Fig. 5. SpatialFuser effectively integrates and aligns cross-resolution datasets from consecutive mouse cerebellum tissue slices across transcriptomic, proteomic, and metabolomic data. a, Pair-wise integration and alignment (300 alignment pairs shown for clarity) of transcriptomic (MAGIC-seq), proteomic (PLATO), and metabolomic (MALDI-MSI) data from consecutive mouse cerebellum tissue slices, with Louvain clustering for tissue domain detection. **b,** UMAP visualizations of modality bias correction results from SpatialFuser on pairwise integrations of the mouse cerebellum datasets, coloured by region labels (top) and

batch labels (bottom). **c,** Distribution of the original Fiber tract annotation in the MAGIC-seq slice and the corresponding RNA–metabolome integration result (Louvain cluster 2), alongside spatial expression patterns of the Fiber tract marker genes Cldn11, Mog, and Cnp. **d,** ROC curves of marker gene expression before and after integration for distinguishing the original Fiber tract annotation. **e,** ROC curves of reconstructed marker gene expression distinguishing Louvain-detected Fiber tracts (cluster 2) and the original Fiber tract annotation.

Discussion

Spatial technologies are expanding into multi-omics, enabling researchers to measure transcriptomics, epigenomics, proteomics, metabolomics, and other molecular layers at defined tissue locations. Underlying this is an unprecedented opportunity to integrate and interpret molecular features from diverse complementary perspectives for a deeper understanding of biological mechanisms within microenvironment[59-61]. In this study, we propose SpatialFuser, a deep learning-based correspondence-driven framework designed for detailed molecular profiling within individual tissue sections and cross-modality, multi-sample integrative analysis of spatial multi-omics data. Through benchmarking on real spatial multi-omics datasets generated by diverse technologies, we show that SpatialFuser outperforms existing state-of-the-art methods in spatial domain detection, consecutive slice alignment, and spatial multi-omics data integration. Notably, we further demonstrate SpatialFuser's unique capability for cross-resolution integration of weakly correlated modalities, which has not been previously established. While each component of SpatialFuser can function independently, their synergistic combination within the framework provides a holistic and streamlined solution for accurate spatial interpretation, robust cross-modality integration, and effective cross-slice alignment of spatial multi-omics data.

To accommodate diverse input modalities, SpatialFuser introduces MCGATE to learn unified representations from multi-platform spatial omics data. Unlike methods such as STAGATE and SEDR that primarily focus on local information propagation, MCGATE employs a multi-head collaborative attention mechanism that further

captures long-range feature correspondences in the embedding space, enabling more accurate identification of spatially coherent patterns. The effectiveness of this design is supported by ablation and interpretability analyses, and beyond performance gains, the architecture improves model transparency while providing new perspectives for interpreting spatial molecular organization.

SpatialFuser introduces foundationally methodological advances and broader utility in cross-slice alignment and multi-modality integration compared with previous approaches. Unlike alignment methods such as STAligner and SPACEL, SpatialFuser uses slice-specific embeddings without requiring matched feature spaces across technologies and applies a more relaxed alignment strategy to model cross-slice correspondences, thereby better supporting scenarios involving different technologies. Beyond alignment, integration tools such as MISO, SpatialGlue, and MultiGATE are designed for co-profiled data measured on the same tissue section and often assume similar cellular composition or strong biological priors. Rather than relying on one-to-one molecular correspondences across modalities, SpatialFuser captures correlated spatial molecular patterns in the embedding space to correct batch effects and modality biases. It learns latent correspondences directly from data and is built to handle unpaired sections across different samples, resolutions, technologies, and developmental stages, extending multi-omics analysis to a broader range of real-world scenarios.

A current limitation of SpatialFuser lies in memory footprint of multi-head collaborative attention mechanism. The lack of memory-efficient parallel operations for high-dimensional sparse matrices in current deep learning frameworks leads to substantial GPU memory demands for large datasets [62, 63]. We anticipate that ongoing advances in hardware and software will mitigate these bottlenecks. Moreover, we plan to develop more efficient sparse computation strategies in future versions of SpatialFuser to better support analyses of increasingly large tissue sections at higher spatial resolution. In fact, SpatialFuser remains generally fast. With sparse-optimized single-head mode, MCGATE embedding generation requires only ~30 seconds per

DLPFC slice on an NVIDIA GeForce RTX 4090 GPU (24 GB), and integration efficiency is reflected in our benchmarking results.

Although image-derived features can improve tissue region identification and spatial molecular pattern recognition[25, 64], SpatialFuser does not currently include a dedicated histology embedding module, as histological images are not consistently available across spatial omics platforms, limiting the general applicability of image-dependent models. Moreover, effective image representation learning often relies on large neural networks with substantial training costs. In contrast, SpatialFuser is intentionally lightweight. In lower-throughput modality analysis scenarios such as CODEX, SpatialFuser may involve only a few thousand trainable parameters, which supports its deployment in resource-constrained settings.

In summary, SpatialFuser does not merely incrementally improve upon existing methods, but also provides a versatile, robust, and broadly applicable framework that empowers researchers to integrate spatial multi-omics data across a wide spectrum of experimental designs, beyond the limitations of current methods. To facilitate broad adoption by the research community, we have released the SpatialFuser package along with detailed tutorials and demo cases for all presented experiments at our GitHub repository https://github.com/liwz-lab/SpatialFuser.

Methods

Data Preprocessing

Filtering and normalization. SpatialFuser takes cellular information and spatial coordinate information of spatial multi-omics data as input, and we refer to the cells or the capture voxels in omics experiment as spots. We first filtered spots lacking valid annotations for reliable analysis. Spatial coordinates of each spot were normalized to $[0,1]$ to better incorporate spatial information into the model. The normalization of cellular information matrices was conducted using Scanpy[65] package. For each slice, we normalized the raw cellular matrix by total counts across all features, scaling each

spot to 10,000 total counts, and applied a log-transformation to the normalized matrix for better variance stabilization and interpretability of expression differences. For dataset generated from sequencing-based and high-throughput AI-enhanced technology, we select the top 3000 spatial variable features.

Construction of graph inputs. SpatialFuser models given spatial multi-omics data as graph-structured inputs $\mathcal{G} = \{(x_i, s_i), i = 1,2,\cdots,N\}$, where N is the number of spots within slice, $x_i \in \mathbb{R}^d$ is pre-processed feature of spot i, and S_i is adjacency spots of spot i constructed by K-Nearest Neighbours algorithm (KNN) based on spatial coordinates. We adjust the parameter K adaptively based on spatial domain heterogeneity and connectivity patterns across different multi-omics scenarios and platforms.

MCGATE for multi-omics data

Encoder. Given input $\mathcal{G}$, the k-th encoder layer $(k \in \{1,2,\cdots,L\})$ of spot i first adaptively learns the pairwise relationship between its adjacent spots j, represented by an edge weight in the graph structure. The edge weight e_{ij}^k at layer k is computed as:

$$e_{ij}^k = Sigmoid\left(V_1^{k^T}\left(W_1^k x_i^{k-1}\right) + V_2^{k^T}\left(W_1^k x_j^{k-1}\right)\right), \tag{1}$$

where W_1 is a learnable weight matrix, V_1 and V_2 are trainable weight vectors. To obtain the local attention weights for information aggregation, the edge e_{ij}^k weights are normalized using a sparse $Softmax$ function over S_i of spot i:

$$att_{ij}^k = Softmax_{j \in S_i}\left(e_{ij}^k\right). \tag{2}$$

Based on local attention weights, the encoder aggregates information from S_i of spot i to compute the smoothed representation x'^k as:

$$x'^k = BN\left(\sum_{j \in S_i} att_{ij}^k \left(W_1^k x_j^{k-1}\right)\right), \tag{3}$$

where BN denotes batch normalization[66], which stabilizes training and accelerates convergence[67].

Finally, a feedforward transformation with residual connection is applied to introduce nonlinearity and enhance the model's representational capacity:

$$x_i^k = BN\left(\sigma\left(W_2^k x'^k + b^k\right) + x'^k\right), \tag{4}$$

where W_2 is a learnable weight matrix, b is a trainable bias, and σ is a customized activation function. The encoder takes the pre-processed feature vector x_i^0 as input, and the output of the final layer x_i^L serves as the low-dimensional feature embedding learned by MCGATE.

Adaptive adjacency expansion layer. Based on the assumption that spatially adjacent cells exhibit high feature similarity, previous methods typically rely on convolutional or neighbourhood-based attention mechanisms for information aggregation. However, our experiments show that in spatial omics data, the spot pairs with the highest feature similarity are often not spatially adjacent. This suggests that relying solely on local features as regularization may limit the model's ability to capture true spatial expression patterns, highlighting the importance of long-range information propagation for accurate embedding learning. However, computing global attention within the encoder becomes increasingly expensive as the number of spots grows, making it impractical for large-scale slices learning. To address this, we introduce a collaborative attention mechanism that efficiently models spatial expression patterns in the low-dimensional embedding space.

The adaptive adjacency expansion layer first computes cosine similarity between spot embeddings as follow:

$$\begin{aligned} Cos_{ij}^{local} &= Softmax_{S_i}\left(\frac{x_i^L \cdot x_j^L}{\|x_i^L\| \cdot \|x_j^L\|}\right), \\ Cos_{ij}^{long} &= Softmax_{S_i^c}\left(\frac{x_i^L \cdot x_j^L}{\|x_i^L\| \cdot \|x_j^L\|}\right), \end{aligned} \tag{5}$$

where S_i denotes the spatial neighbourhood of spot i, and S_i^c denotes the set of non-spatial neighbors.

The local cosine similarity matrix Cos_{ij}^{local} is fed back to the encoder to enhance the local attention mechanism previously defined in Equation (2):

$$att_{local_{ij}}^{k} = Softmax_{S_i}\left(Cos_{ij}^{local} * e_{ij}^{k}\right), \tag{6}$$

where e_{ij}^{k} is the attention coefficient computed at k-th encoder layer.

For each spot i, the adaptive adjacency expansion layer selects the mutual nearest neighbours[68] (MNNs) spots with highest similarity outside its spatial neighbourhood from Cos_{ij}^{long}, generating a sparse long-range correspondence matrix that defines reliable pathways for distant information propagation. During training process, let G_i denotes the set of long-range neighbours for spot i, the long-range attention weights are computed as:

$$att_{long_{ij}}^{k} = Softmax_{G_i}\left(e_{ij}^{k}\right). \tag{7}$$

The collaborative attention weights are then defined as the linear addition of local and long-range attention:

$$att_{co_{ij}}^{k} = \alpha\, att_{long_{ij}}^{k} + (1-\alpha) att_{local_{ij}}^{k}, \tag{8}$$

where $\alpha \in [0,1]$ (default $\alpha = 0$) is a slice-specific hyperparameter controlling the relative influence of long-range correspondence during information aggregation. By integrating long-range and local attention, the adaptive adjacency expansion layer enables efficient global information propagation and aggregation, thereby enhancing MCGATE's ability to learn meaningful spatial expression patterns in spatial omics data.

Decoder. Symmetric to the encoder, the k-th decoder layer ($k \in \{1,2,\cdots,L\}$) reconstructs the representation of spot i as follows:

$$\hat{x}_i^{k-1} = \sigma\left(\widehat{W}_1^{k}\left(\sum_{j\in S_i} \widehat{att}_{co_{ij}}^{k}\, \hat{x}_j^{k}\right)\right). \tag{9}$$

The decoder takes the low-dimensional feature embedding x^{L} as input and outputs the reconstructed low-noise representation $\hat{x}_i^{0}$, which approximates the original data.

To reduce the risk of overfitting, MCGATE adopts a parameter sharing strategy between the encoder and the decoder. Specifically, for the k-th decoder layer, we set $\widehat{W}_1^k = W_1^{k^T}$ and $\widehat{att}_{co_{ij}}^k = att_{co_{ij}}^k$ to share both transformation matrices and attention weights. However, the decoder does not replicate the feedforward structure of the encoder, in order to prevent overly rigid architectural symmetry, which may constrain the model's expressiveness.

The training of MCGATE is driven by a reconstruction objective, which optimizes the low-dimensional embeddings through minimization of the mean squared error (MSE) loss between input and decoder output:

$$Loss_{rec} = \frac{1}{N}\sum_{i=1}^{N}\left\|x_i^0 - \hat{x}_i^0\right\|_2, \tag{10}$$

where N is the total number of spots.

Multi-head attention. Inspired by previous works[69], MCGATE incorporates multi-head attention mechanism to enable more fine-grained information aggregation by partitioning input features into multiple sub-space. In the k-th encoder layer ($k \in \{1,2,\cdots,L\}$), the h-th collaborative attention weight ($h \in \{1,2,\cdots,H\}$) between spots i and j, denoted $att_{co_{ij}}^{k,h}$, is computed in parallel using independent parameters $W_1^{k,h}$, $V_1^{k,h}$, and $V_2^{k.h}$. Notably, in the adaptive adjacency expansion layer, cosine similarity is also computed separately in each subspace, ensuring that feature aggregation in each head captures the diversity of underlying feature structures.

Under the multi-head attention paradigm, the overall information aggregation process in both encoder and decoder is realized by concatenating the outputs of multiple parallel attention heads. The aggregation formula in the encoder (extending Equation 3) is:

$$x'^k = BN\left(Concat_{h=1}^{H}\left(\sum_{j\in S_i} att_{co_{ij}}^{k,h}\left(W_1^{k,h}x_j^{k-1,h}\right)\right)\right). \tag{11}$$

Similarly, the decoder propagation process (extending Equation 9) becomes:

$$\hat{x}_i^{k-1} = \sigma\left(Concat_{h=1}^{H}\left(\widehat{W}_1^{k,h}\left(\sum_{j\in S_i} \widehat{att}_{co_{ij}}^{k,h}\, \hat{x}_j^{k,h} \right)\right)\right). \tag{12}$$

Due to the lack of efficient sparse multi-channel tensor operations in deep learning frameworks (e.g., PyTorch[69]), the multi-head attention mechanism often leads to high GPU memory usage. To ensure scalability on large spatial multi-omics datasets, MCGATE adopts sparse tensor operations in the single-head mode to effectively reduce memory usage and computational cost during training.

Matching layer

SpatialFuser models the alignment of spatial omics slices as a graph matching problem, which is essentially a classic optimal transport task. The core objective is to find an optimal coupling matrix that defines cross-slice correspondences by minimizing the Wasserstein distance[70] between embedding distributions. Let the low-dimensional embeddings of the two slices be $E_1 \in \mathbb{R}^{n\times d'}$ and $E_2 \in \mathbb{R}^{m\times d'}$, SpatialFuser introduces spatial proximity constraints to define a similarity measure across slices:

$$Simi(i,j) = \begin{cases} \dfrac{\dfrac{E_1[i]\cdot E_2[j]}{\|E_1[i]\|\cdot\|E_2[j]\|}+1}{2\tau+10^{-6}}, & if\ \|Coo_1[i]-Coo_1[j]\|_2 \le \mathrm{r} \\ 0, & otherwise \end{cases} \tag{13}$$

where the indices $i \in \{1,2,\cdots,n\}$ and $j \in \{1,2,\cdots,m\}$ denote arbitrary spots on each slice, τ is a temperature coefficient, $Coo_1[i]$ and $Coo_2[j]$ represent their corresponding spatial coordinate vectors, and r denotes the spatial neighbourhood radius, restricting the cross-sample correspondences to a local spatial window.

Based on the similarity matrix defined in Equation (13), SpatialFuser formulates the alignment task as the following optimal transport problem:

$$\underset{\mathrm{T}\in\mathcal{T}}{\operatorname{argmax}} \sum_{i=0}^{n}\sum_{j=0}^{m} Simi(i,j)T_{ij} + \lambda\cdot Reg(T), \tag{14}$$

where $T \in [0,1]^{n\times m}$ is the soft coupling matrix to be optimized. The set $\mathcal{T} = \{T \in \mathbb{R}^{n\times m} \,|T1_m \leq 1_n, T^T 1_n \leq 1_m\}$ defines the feasible region. $Reg(\cdot)$ denotes the entropy regularizer, and λ is its weighting factor. SpatialFuser employs the Sinkhorn algorithm[71] to solve this linear program via efficient iterative updates.

While Equation (14) implicitly assumes that every spot in the two slices can be reliably matched, heterogeneous spatial omics datasets often differ in cell-type composition, resolution, and batch size. A robust alignment algorithm should therefore tolerate unmatched spots. To address this, SpatialFuser introduces a relaxed mapping strategy by augmenting the original similarity matrix $Simi \in \mathbb{R}^{n\times m}$ defined in Equation (13) with additional dustbin channels, assigning them predefined weights as follows:

$$\overline{Simi}(i,j) = \begin{cases} Simi(i,j), & for\ \mathrm{i} \leq \mathrm{n}, \mathrm{j} \leq \mathrm{m} \\ S_{bin}, & \text{otherwise} \end{cases} \tag{15}$$

where $i \in \mathbb{Z}_{[0,n+1]}$ and $j \in \mathbb{Z}_{[0,m+1]}$, $\overline{Simi}$ is the augmented similarity matrix with the dustbin channels, and S_{bin} is a constant weight ($S_{bin} = 1$ in default) assigned to the dustbin.

The dustbin channel strategy is a form of relaxation, which essentially converts the inequality-constrained problem in Equation (14) into an equality-constrained problem with an expanded feasible region:

$$\bar{\mathcal{T}} = \left\{\bar{T} \in \mathbb{R}^{(n+1)\times(m+1)} \,\middle|\, \bar{T}1_{m+1} = [1_n^T, m]^T \ , T^T 1_n = [1_m^T, n]^T\right\}. \tag{16}$$

By allowing unmatched or low-confidence points to be assigned to the dustbin channels, the model can adaptively distinguish meaningful correspondences from noisy ones, thus relaxing the strong assumption of compositional similarity between slice and avoiding forcible alignment.

To ensure sufficient supervision signal during early training stage and further improve matching quality, SpatialFuser introduces a Mutual Top-K Neighbours (MKNs) strategy based on the soft assignment matrix generated by the Sinkhorn algorithm. Specifically, the model initially selects MNNs from the soft couplings to construct high-confidence correspondences. However, strong batch effects in the early

stages may result in insufficient MNNs. To address this, the matching criteria are relaxed to include mutual top-K neighbours, enhancing the supervision signal and ensuring stable training. These MKNs serve as anchors to guide the fusion layer.

Fusion layer

In the fusion layer design, a shared multilayer perceptron (MLP) is connected to two independent MCGATE modules to project the low-dimensional embeddings of spatial omics data from different experimental platforms and modalities into a unified latent space. To prevent the model from memorizing batch-specific patterns, the embeddings are firstly concatenated and then randomly permuted along the row prior to being input into the MLP, encouraging it to learn a generalizable mapping function.

The fusion layer employs contrastive learning for spatial multi-omics data integration. Specifically, high-confidence cross-slice correspondences derived from the matching layer serve as anchor-positive pairs to guide the correction of batch effects or modality biases. To enhance cell-type discriminability, negative samples are randomly drawn from non-neighbouring regions within the same tissue slice. Given that the number of spatial spots typically far exceeds the local neighbourhood size, this strategy is both efficient and effective in minimizing the risk of false negatives.

The training process employs an improved triplet loss[31] as the modality fusion objective, simultaneously promoting the clustering of similar samples (anchor-positive) and the separation of dissimilar ones (anchor-negative), while enforcing both absolute and relative distance constraints to improve training stability:

$$Loss_{fuse} = \left\|E_i - E_p^i\right\|_2 + \frac{1}{N_a}\sum_{i=1}^{N_a} \max\left\{\left\|E_i - E_p^i\right\|_2 - \left\|E_a^i - E_n^i\right\|_2 + \tau, 0\right\}, \quad (17)$$

where E_i is the fused embedding of spot i, serving as anchor, E_p^i and E_n^i denote the embeddings of the corresponding positive and negative spots respectively, and τ is a margin parameter. To prevent embedding collapse and preserve spatial representational fidelity, a reconstruction loss is incorporated into the integration process. Meanwhile,

to address potential inconsistencies caused by the dual independent encoder design of SpatialFuser, we introduce an embedding direction constraint loss $Loss_{dir}$ to align the global orientation of the representations across slices:

$$Loss_{dir} = 1 - \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{E_1[i]\cdot E_2[j]}{\|E_1[i]\|\cdot\|E_2[j]\|}. \tag{17}$$

The total objective function of the integration process combines the fusion objective, the reconstruction loss, and the embedding direction constraint loss, and is formulated as:

$$Loss_{integrate} = Loss_{fuse} + \beta_{rec}\, Loss_{rec} + \beta_{dir} Loss_{dir}, \tag{18}$$

where β_{rec} and β_{dir} are the weight of the reconstruction loss and the embedding direction constraint loss, respectively $\beta_{rec} = 1$, $\beta_{dir} = 0.1$ in default). The embeddings integrated by the fusion layer are used in the next iteration to guide the matching layer in identifying more accurate cross-slice correspondences, enabling progressively improved alignment and integration across training rounds.

Coordinate re-matching modules (optional)

Spatial registration of spot coordinates is a key preparatory step for aligning and integrating spatial omics slices. SpatialFuser formulates spatial omics data as 2D point clouds and corrects coordinate discrepancies via rigid transformations defined by a rotation matrix M and a translation t. The coordinate registration problem is formulated as the following optimization objective:

$$argmin_{M,t}\, \|Y_1 - Y_2 \cdot M - t\|^2, \tag{19}$$

where $Y_1 \in \mathbb{R}^{n\times 2}$ and $Y_2 \in \mathbb{R}^{m\times 2}$ denote the spatial coordinate matrices of the two datasets to be aligned, with n and m are the numbers of spots in each set.

SpatialFuser provides multiple solution strategies to accommodate different scenarios. The ICP algorithm is a classical method for solving point cloud registration problems[72, 73]. It first establishes spot correspondences by finding the nearest neighbour for each spot in one slice, and then addresses the above optimization problem

via singular value decomposition, iteratively updating the rotation matrix M and translation vector t by minimizing the Euclidean distance between matched spot pairs. The ICP algorithm has been widely applied to spatial transcriptomic slice alignment, particularly in cases where the slices display well-defined structural grids or regular anatomical boundaries.

However, ICP is not consistently reliable on slices with uneven spot distributions. To address this limitation, SpatialFuser also incorporates the NDT algorithm[74], which models the reference slice by partitioning it into a 2D grid. Each grid cell $V_{i,j}$ contains the spots within its bounds, and a multivariate Gaussian distribution $\mathcal{N}(\mu_{i,j}, \Sigma_{i,j})$ is fitted to the spots in each non-empty grid cell. Instead of relying on pairwise Euclidean distances, NDT evaluates the Mahalanobis distance between each transformed spot $z_t \in \mathbb{R}^2$ and the Gaussian distribution of the corresponding grid cell in the reference slice. The coordinate pre-matching is achieved by minimizing the cumulative deviation:

$$\min_{R,t} \sum_{k=1}^{K} \left(z_k - \mu_{i,j}\right)^T \Sigma_{i,j}^{-1} \left(z_k - \mu_{i,j}\right), \tag{20}$$

where K is the total spot number of the transformed slice. This probabilistic approach enables more robust registration in scenarios involving sparse, noisy, or irregular spot distributions[75].

Whole-slice Alignment

We propose a whole-slice alignment strategy to support downstream analysis and performance comparison against existing alignment methods. By default, the slice with fewer spatial spots is mapped to the one with more spots. For each spot i in the source slice, a candidate set is first identified in the target slice by selecting all spots within a spatial distance radius r, and we further compute the cosine similarity between spot i and its candidate spots based on the fused embeddings. Inspired by previous work[24], we introduce a probabilistic quality control mechanism. Rather than randomly pairing

spots across both slices, we randomly sample 1000 spots from the source slice and, for each sampled spot, randomly select a nearby spot within radius r in the target slice to compute cosine similarities, in order to form a stricter null distribution. The 95th percentile of this null distribution is used as the threshold for quality control. In scenarios where the source and target slices have the same resolution, we retain the candidate with the highest similarity score as the final alignment result for each source spot.

Implementation details of SpatialFuser

Spatial domain detection task. To accommodate datasets with varying complexity, SpatialFuser adopts different MCGATE configurations. For high-throughput data such as 10X Visium, Stereo-seq, and MAGIC-seq, a two-layer MCGATE is used, with hidden dimensions set to 512 and 32 respectively. For low-throughput data such as CODEX and MALDI-MSI, a single-layer MCGATE with 32 hidden units is employed. The MCGATE is trained using the Adam optimizer, with the ELU function selected as the activation. The model is trained for 500 steps by default, during which the Adaptive Adjacency Expansion Layer is updated every 100 epochs to refine long-range relationships and their associated attention weights. Both the learning rate and the long-range attention coefficient α are customized for each dataset.

Alignment and integration task. The training process is divided into two stages: pre-training and contrastive learning. In the pre-training stage, a fusion layer is connected to two independent MCGATE encoders and jointly trained with a reconstruction loss for 500 epochs by default, producing high-quality embeddings for downstream alignment and integration. For slices with semantically similar molecular profiles, shared features across slices are used. The matching layer and the fusion layer are then iteratively optimized to align the embeddings into a shared latent space. Specifically, the matching layer first identifies anchor-positive pairs based on the pre-trained embeddings. These identified pairs are then used to guide the fusion layer in integrating

the modalities. By default, triplet pairs are updated every 20 epochs, and this iterative process continues until 500 steps. The learning rate is dataset-specific defined.

Clustering, differential expression analysis, and spatial trajectory inference

Clustering was performed by constructing a shared nearest neighbour graph based on the spatially resolved data, followed by cluster identification using the Louvain[35], Leiden[34], or Mclust[33] algorithms. The default resolution parameter used in benchmarking for the Louvain and Leiden clustering algorithms was set to 0.1, while for other downstream analyses it was adjusted to the optimal value for each case. Cluster biomarkers and differentially expressed genes between groups were determined using the Wilcoxon rank-sum test, as implemented in the rank_genes_groups() function of the Scanpy[65] package. To infer spatial trajectories, we applied partition-based graph abstraction via the paga() function in Scanpy.

Spatial distribution detection across diverse experimental platforms and modalities

Data description. The 10X Visium DLPFC dataset[32] comprises 12 tissue sections from 3 individuals, providing whole-transcriptome coverage with 33,438 gene features. Each section contains between 3498 and 4789 spatially resolved spots. The mouse somatosensory cortex dataset[2], generated using the osmFISH platform, includes 5328 spots at single-cell resolution, based on the targeted detection of 33 genes. The CODEX human bladder cancer dataset[4] consists of 75 tumour tissue sections from 31 patients, profiling 35 protein markers across nuclear and membrane compartments. After quality control, ~360,000 epithelial cells, ~140,000 immune cells, and ~90,000 stromal cells were retained. A 1440-dimensional feature matrix was initially constructed using statistical summaries (e.g., percentages and quantile distributions) of protein expression. For the detection of cell type spatial distributions, the 210308_TMA2_reg6 slice was selected, and the average nuclear and membrane expression levels of each protein were

extracted as input features. All datasets include detailed annotations of tissue regions or cell types, which serve as ground truth for evaluation.

Hyperparameter settings. The hyperparameters for SpatialFuser were optimized through grid search. For all baseline methods, we adopted the recommended data pre-processing steps and hyperparameter settings as specified in their official implementation and documentation for DLPFC dataset, except for STAligner. Since STAligner's embedding model structure is very similar to that of STAGATE, we referred to the documentation of STAGATE for its parameter settings. The documentation links for these methods are shown as follows:

- SpaGCN: https://github.com/jianhuupenn/SpaGCN/blob/master/tutorial/tutorial.md
- STAGATE: https://stagate.readthedocs.io/en/latest/index.html
- SEDR: https://sedr.readthedocs.io/en/latest/index.html

Hyperparameter robustness. We tested SpatialFuser's robustness to the key hyperparameters of MCGATE including: (1) number of encoder and decoder layers, (2) dimension of layers, (3) dimension of MCGATE embedding. We ran MCGATE on all 12 slices of DLPFC dataset and every experiment was run with different random seeds. To identify a suitable attention architecture for the DLPFC dataset, we benchmarked different numbers of attention heads based on a selected network structure from the robustness testing.

Spatiotemporal alignment and integration

Data description. The BaristaSeq mouse visual cortex dataset[9] comprises three tissue slices with similar domain distributions and morphological structures, containing 4491, 3545, and 3390 spatial spots, respectively. A total of 78 transcript features were retained, along with detailed anatomical region annotations. Slices 1 and 2 were selected for integration and alignment. The Stereo-seq axolotl regenerative telencephalon dataset[10] includes 18 brain tissue slices with single-cell resolution, capturing over 20,000 genes across various stages of regeneration and development.

Two slices from developmental stages 54 and 57, containing 2929 and 4410 spatial spots respectively, were selected for alignment and integration. The original annotations were used as references for evaluation.

Hyperparameter settings. The hyperparameters for SpatialFuser were optimized through grid search, while the hyperparameter settings for all other methods followed the official guidelines and tutorials provided by their respective authors:

- SALT: https://slat.readthedocs.io/en/latest/tutorials.html
- SPACEL: https://spacel.readthedocs.io/en/latest/index.html
- STAligner: https://staligner.readthedocs.io/en/latest/index.html

To ensure a fair comparison of the integration and alignment performance on consecutive BaristaSeq slices, we conducted a grid search for several key parameters related to both graph construction and network architecture, which are critical yet not explicitly specified in the original documentation of the compared methods. Specifically, SPACEL constructs the spatial graph using 8-nearest neighbours and employs a two-layer spline, with a hidden layer size of 256 and an output embedding dimension of 16 (Supplementary Fig. 6). STAligner builds the graph using 20-nearest neighbours and adopts a two-layer encoder with node dimensions set to 64 and 30, respectively (Supplementary Fig. 7). For contrastive learning-based integration process, the number of nearest neighbours used when constructing MNNs was set to 10. Notably, while typical GCNs tend to suffer from over-fitting, gradient vanishing, or over-smoothing beyond 4–5 layers, SLAT consistently demonstrated improved performance with increased depth, achieving its best results with over 20 LGCN layers (Supplementary Fig. 8). Therefore, we report SLAT's performance using 20 layers GCN with mlp_hidden = 128 and hidden_size = 64 in the experiments on the BaristaSeq dataset. Before training, SLAT constructs a KNN graph with K=10.

Spatial ATAC-RNA-seq data integration

Data description. The spatial ATAC-RNA-seq dataset[14] of the E13 mouse embryo comprises 2187 spots with genome-wide co-profiling of epigenome and transcriptome. The transcriptomic assay measures 15,748 genes, while the epigenomic assay includes 32,437 accessible chromatin peaks. Gene-level chromatin accessibility was further summarized into activity scores for 24,017 genes derived from the ATAC signal. Manual annotation of anatomical regions was not provided by the authors. Instead, the dataset includes eight major ATAC-defined clusters, 14 RNA-defined clusters, and 14 integrative clusters based on joint analysis of spatial ATAC and RNA data.

Hyperparameter settings. For the high-throughput spatial ATAC-RNA-seq data, we constructed a 4-nearest neighbour graph using the top 3000 spatially variable features, which were normalized and log-transformed before being used as input. The number of layers in MCGATE was set to two, with hidden dimensions of 512 and 32, respectively. We first pre-trained two 4-head MCGATE model separately on the ATAC and RNA modalities for 500 epochs using a learning rate of 1e-3 and 1e-4. Further, we set the reconstruction loss weight $\beta_{rec} = 50$ and the directional loss weight $\beta_{dir} = 0.1$ and trained the fusion and matching layers for 500 steps with a learning rate of 1e-4 to achieve integration and alignment. For all other methods, we followed the official protocols and usage instructions provided by the respective authors:

- MISO: https://github.com/kpcoleman/miso/blob/main/tutorial/tutorial.ipynb
- MultiGATE: https://multigate.readthedocs.io/en/latest/index.html
- SpatialGlue: https://spatialglue-tutorials.readthedocs.io/en/latest/index.html
- SIMO: https://github.com/ZJUFanLab/SIMO/tree/main

For fair comparison, parameters for each method were tuned within the ranges recommended in the original studies. For MISO, pixel_size_raw was estimated using the 50 μm grid resolution and the pixel distance between adjacent spots to align image pixel and physical scales for spot-level image feature extraction. For MultiGATE, the learning rate was set to 1e-3 to ensure stable integration, and model performance showed minimal sensitivity to bp_width (Supplementary Fig. 12), which was therefore

set to 400. As SpatialGlue uses data-type-specific weight_factors schemes, we applied the Spatial-epigenome-transcriptome training mode following the official documentation. For SIMO, we followed the original protocol in which four adjacent pixels were merged to construct the spatial transcriptomics data, while the original data were treated as single-cell inputs for mapping. Original clustering labels were incorporated as structural priors during training. We observed that increasing the number of mapped cells led to progressive loss of spatial coherence and disruption of tissue organization (Supplementary Fig. 11). Therefore, top_num was set to 3 for result presentation and comparison.

Marker–Region Concordance. To assess the agreement between identified cortical plate (CP) clusters and CP-associated molecular patterns, we derived a CP marker enrichment score from ATAC data by aggregating normalized chromatin accessibility signals of the canonical CP markers Tbr1 and Satb2 via the score_genes() function in Scanpy. Marker-enriched spots were defined using quantile-based thresholds. For each threshold, cluster purity was defined as the fraction of cluster spots classified as marker-enriched, while coverage was defined as the fraction of marker-enriched spots captured by the cluster.

Efficiency Evaluation. We evaluated computational efficiency on a single NVIDIA GeForce RTX 4090 GPU with 24 GB memory. To minimize the influence of hardware-related variability, the training of each method was repeated five times, and we report the mean and standard deviation to provide a more robust estimate of practical runtime. For methods consisting of multiple computational stages, we recorded stage-specific runtimes and present stacked runtimes.

Spatial alignment and integration of weakly correlated modalities across resolutions

Data description. The multi-omics mouse cerebellum dataset[15] comprises three consecutive tissue slices, each representing a distinct spatial omics modality: MAGIC-

seq for spatial transcriptomics, PLATO for spatial proteomics, and MALDI-MSI for spatial metabolomics. The MAGIC-seq slice captured 1677 spatial spots at a resolution of 32 μm, profiling 16,116 genes. The PLATO slice provides high-throughput proteomic profiling aligned with the transcriptomic slice, identifying 5722 protein groups after AI enhancement and quality control. The MALDI-MSI slice achieves higher spatial resolution, capturing 3908 spots with 491 metabolite peaks. Spatial domain annotations are available for all three slices, though slight differences exist in domain distributions across modalities. Notably, the slices with different resolutions exhibit significant discrepancies in spatial coordinates.

Hyperparameter settings. In the experiments on the multi-omics mouse cerebellum dataset, we set K=4 for KNN graph construction and employed the NDT algorithm for coordinate registration prior to training. During the pretraining stage, we extracted the top 3000 spatially variable features from both the MAGIC-seq transcriptomic slice and the PLATO proteomic slice. These features were normalized and log-transformed before being input into separate two-layer MCGATE models with hidden dimensions of 512 and 32. The number of attention heads was set to 4, and the models were trained for 500 steps using the Adam optimizer with learning rates of 3e-3 and 1e-3, respectively. Considering the relatively low throughput of the MALDI-MSI metabolomic data, we used all available features as input and employed a single-layer MCGATE model with a 32-dimensional embedding and four attention heads, trained for 500 steps with a learning rate of 5e-4. In the integration and alignment stage, we set $\beta_{rec} = 50$ and $\beta_{dir} = 0.1$, and trained the fusion and matching layers for 200 steps using the Adam optimizer with a learning rate of 3e-3.

Data availability

All experimental data of this study have already been published and are accessible within the corresponding articles and public repositories. Specifically, the 10X Visium DLPFC dataset is available in the spatialLIBD package[76]

(http://spatial.libd.org/spatialLIBD); the osmFISH mouse somatosensory cortex dataset is available at http://linnarssonlab.org/osmFISH/; the CODEX human bladder cancer dataset is available from the Aquila database[77] (https://aquila.cheunglab.org/); the BaristaSeq mouse visual cortex dataset is available from the SODB database[78] (https://gene.ai.tencent.com/SpatialOmics/); the Stereo-seq axolotl regenerative telencephalon dataset is available from the STOmicsDB database[79] (https://db.cngb.org/stomics/artista/); the spatial ATAC–RNA-seq E13 mouse embryo dataset is available at https://ki.se/en/mbb/oligointernode; and the spatial multi-omics mouse cerebellum dataset is available from the original article.

Code availability

The SpatialFuser framework was implemented in the "spatialFuser" Python package, which is open-source for the research community and can be accessible at https://github.com/liwz-lab/SpatialFuser.

Competing interests

To maximize the impact of this study, Sun Yat-sen University has submitted a patent application to the State Intellectual Property Office of Chia (SIPO).

Acknowledgements

This work was supported by the grants of National Natural Science Foundation of China (92474107 and 32570798), National Key R&D Program of China (2021YFF1200903), Major Project of Guangzhou National Laboratory of China (GZNL2024A01003), and Guangdong Basic and Applied Basic Research Foundation of China (2022B1515120077).

References

1. Qiu, X., et al., Spatiotemporal modeling of molecular holograms. Cell, 2024. **187**(26): p. 7351-7373. e61.
2. Codeluppi, S., et al., Spatial organization of the somatosensory cortex revealed by osmFISH. Nature methods, 2018. **15**(11): p. 932-935.
3. Zhang, M., et al., Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature, 2021. **598**(7879): p. 137-143.
4. Black, S., et al., CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nature protocols, 2021. **16**(8): p. 3802-3835.
5. Ptacek, J., et al., Multiplexed ion beam imaging (MIBI) for characterization of the tumor microenvironment across tumor types. Laboratory Investigation, 2020. **100**(8): p. 1111-1123.
6. Williams, C., et al., Spatial insights into tumor immune evasion illuminated with 1000-plex RNA profiling with CosMx Spatial Molecular Imager. Cancer Res, 2023. **83**(6765): p. 10.
7. Rohner, T.C., D. Staab, and M. Stoeckli, MALDI mass spectrometric imaging of biological tissue sections. Mechanisms of ageing and development, 2005. **126**(1): p. 177-185.
8. Ståhl, P.L., et al., Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science, 2016. **353**(6294): p. 78-82.
9. Chen, X., et al., Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. Nucleic acids research, 2018. **46**(4): p. e22-e22.
10. Wei, X., et al., Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. Science, 2022. **377**(6610): p. eabp9444.
11. Zhu, J., et al., Custom microfluidic chip design enables cost-effective three-dimensional spatiotemporal transcriptomics with a wide field of view. Nature Genetics, 2024. **56**(10): p. 2259-2270.
12. Liao, S., et al., Integrated spatial transcriptomic and proteomic analysis of fresh frozen tissue based on stereo-seq. bioRxiv, 2023: p. 2023.04. 28.538364.
13. Deng, Y., et al., Spatial profiling of chromatin accessibility in mouse and human tissues. Nature, 2022. **609**(7926): p. 375-383.
14. Zhang, D., et al., Spatial epigenome–transcriptome co-profiling of mammalian tissues. Nature, 2023. **616**(7955): p. 113-122.

15. Hu, B., et al., High-resolution spatially resolved proteomics of complex tissues based on microfluidics and transfer learning. Cell, 2025. **188**(3): p. 734-748. e22.

16. Sun, C., et al., Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer. Nature Communications, 2023. **14**(1): p. 2692.

17. Chen, A., et al., Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell, 2022. **185**(10): p. 1777-1792. e21.

18. Li, H., et al., Spatially resolved genome-wide joint profiling of epigenome and transcriptome with spatial-ATAC-RNA-seq and spatial-CUT&Tag-RNA-seq. Nature Protocols, 2025: p. 1-35.

19. Hu, J., et al., SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nature Methods, 2021. **18**(11): p. 1342-1351.

20. Dong, K. and S. Zhang, Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nature Communications, 2022. **13**(1): p. 1739.

21. Xu, H., et al., Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Medicine, 2024. **16**(1): p. 12.

22. Yang, P., et al., Spatial integration of multi-omics single-cell data with SIMO. Nature communications, 2025. **16**(1): p. 1265.

23. Zhou, X., K. Dong, and S. Zhang, Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. Nature Computational Science, 2023. **3**(10): p. 894-906.

24. Xia, C.-R., et al., Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. Nature Communications, 2023. **14**(1): p. 7236.

25. Coleman, K., et al., Resolving tissue complexity by multimodal spatial omics modeling with MISO. Nature methods, 2025. **22**(3): p. 530-538.

26. Long, Y., et al., Deciphering spatial domains from spatial multi-omics with SpatialGlue. Nature Methods, 2024. **21**(9): p. 1658-1667.

27. Miao, J., et al., MultiGATE: integrative analysis and regulatory inference in spatial multi-omics data via graph representation learning. Nature Communications, 2025. **16**(1): p. 9403.

28. Hu, Y., et al., Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. Genome Biology, 2024. **25**(1): p. 212.

29. Lais, P., et al., Image guided construction of a common coordinate framework for spatial transcriptome data. Scientific Reports, 2025. **15**(1): p. 18074.

30. Sarlin, P.-E., et al. Superglue: Learning feature matching with graph neural networks. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

31. Cheng, D., et al. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. in Proceedings of the iEEE conference on computer vision and pattern recognition. 2016.

32. Maynard, K.R., et al., Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nature neuroscience, 2021. **24**(3): p. 425-436.

33. Fraley, C., et al., mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. 2012, Technical report.

34. Traag, V.A., L. Waltman, and N.J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 2019. **9**(1): p. 1-12.

35. Blondel, V.D., et al., Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008. **2008**(10): p. P10008.

36. Gouin III, K.H., et al., An N-Cadherin 2 expressing epithelial cell subpopulation predicts response to surgery, chemotherapy and immunotherapy in bladder cancer. Nature communications, 2021. **12**(1): p. 4906.

37. Yuan, Y., Spatial heterogeneity in the tumor microenvironment. Cold Spring Harbor perspectives in medicine, 2016. **6**(8): p. a026583.

38. Wolf, F.A., et al., PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome biology, 2019. **20**(1): p. 59.

39. Long, Y., et al., Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nature Communications, 2023. **14**(1): p. 1155.

40. Xu, H., et al., SPACEL: deep learning-based characterization of spatial transcriptome architectures. Nature Communications, 2023. **14**(1): p. 7603.

41. Hoshikawa, M., et al., Structure and expression of a novel fibroblast growth factor, FGF-17, preferentially expressed in the embryonic brain. Biochemical and biophysical research communications, 1998. **244**(1): p. 187-191.

42. Dhokia, V. and S. Macip, A master of all trades–linking retinoids to different signalling pathways through the multi-purpose receptor STRA6. Cell death discovery, 2021. **7**(1): p. 358.

43. Xavier-Neto, J., et al., Signaling through retinoic acid receptors in cardiac development: Doing the right things at the right times. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2015. **1849**(2): p. 94-111.

44. Hu, H., Chemorepulsion of neuronal migration by Slit2 in the developing mammalian forebrain. Neuron, 1999. **23**(4): p. 703-711.

45. Li, Z., et al., Cross-modality representation and multi-sample integration of spatially resolved omics data. bioRxiv, 2024: p. 2024.06. 10.598155.

46. Zembrzycki, A., et al., Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain. Neural development, 2007. **2**(1): p. 8.

47. Ma, S., et al., Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell, 2020. **183**(4): p. 1103-1116. e20.

48. Meijer, M., et al., Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility. Neuron, 2022. **110**(7): p. 1193-1210. e13.

49. Barbieri, A.M., et al., A homeobox gene, vax2, controls the patterning of the eye dorsoventral axis. Proceedings of the National Academy of Sciences, 1999. **96**(19): p. 10729-10734.

50. Tang, X., et al., Inwardly rectifying potassium channel Kir4. 1 is responsible for the native inward potassium conductance of satellite glial cells in sensory ganglia. Neuroscience, 2010. **166**(2): p. 397-407.

51. Manuel, M.N., et al., Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. Frontiers in cellular neuroscience, 2015. **9**: p. 70.

52. Chen, J., et al., MYT1L is required for suppressing earlier neuronal development programs in the adult mouse brain. Genome research, 2023. **33**(4): p. 541-556.

53. Chen, S., et al., Integration of spatial and single-cell data across modalities with weakly linked features. Nature Biotechnology, 2024. **42**(7): p. 1096-1106.

54. Lundberg, E. and G.H. Borner, Spatial proteomics: a powerful discovery tool for cell biology. Nature Reviews Molecular Cell Biology, 2019. **20**(5): p. 285-302.

55. Alexandrov, T., Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence. Annual review of biomedical data science, 2020. **3**(1): p. 61-87.

56. Maheras, K.J., et al., Absence of claudin 11 in CNS myelin perturbs behavior and neurotransmitter levels in mice. Scientific reports, 2018. **8**(1): p. 3798.

57. Scolding, N., et al., Myelin-oligodendrocyte glycoprotein (MOG) is a surface marker of oligodendrocyte maturation. Journal of neuroimmunology, 1989. **22**(3): p. 169-176.

58. Lappe-Siefke, C., et al., Disruption of Cnp1 uncouples oligodendroglial functions in axonal support and myelination. Nature genetics, 2003. **33**(3): p. 366-374.

59. Vandereyken, K., et al., Methods and applications for single-cell and spatial multi-omics. Nature Reviews Genetics, 2023. **24**(8): p. 494-515.

60. Liu, X., et al., Spatial multi-omics: deciphering technological landscape of integration of multi-omics and its applications. Journal of Hematology & Oncology, 2024. **17**(1): p. 72.

61. Kiessling, P. and C. Kuppe, Spatial multi-omics: novel tools to study the complexity of cardiovascular diseases. Genome medicine, 2024. **16**(1): p. 14.

62. Smith, S. and G. Karypis. Tensor-matrix products with a compressed sparse tensor. in Proceedings of the 5th Workshop on Irregular Applications: Architectures and Algorithms. 2015.

63. Auddy, A., D. Xia, and M. Yuan, Tensor methods in high dimensional data analysis: Opportunities and challenges. arXiv preprint arXiv:2405.18412, 2024.

64. Bergenstråhle, L., et al., Super-resolved spatial transcriptomics by deep data fusion. Nature biotechnology, 2022. **40**(4): p. 476-479.

65. Wolf, F.A., P. Angerer, and F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis. Genome biology, 2018. **19**(1): p. 15.

66. Ioffe, S. and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in International conference on machine learning. 2015. pmlr.

67. Bjorck, N., et al., Understanding batch normalization. Advances in neural information processing systems, 2018. **31**.

68. Haghverdi, L., et al., Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nature biotechnology, 2018. **36**(5): p. 421-427.

69. Vaswani, A., et al., Attention is all you need. Advances in neural information processing systems, 2017. **30**.

70. Panaretos, V.M. and Y. Zemel, Statistical aspects of Wasserstein distances. Annual review of statistics and its application, 2019. **6**(1): p. 405-431.

71. Cuturi, M., Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems, 2013. **26**.

72. Dong, Z., et al., Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 2020. **163**: p. 327-342.

73. Umeyama, S., Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on pattern analysis and machine intelligence, 2002. **13**(4): p. 376-380.

74. Biber, P. and W. Straßer. The normal distributions transform: A new approach to laser scan matching. in Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453). 2003. IEEE.

75. Magnusson, M., The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection. 2009, Örebro universitet.

76. Pardo, B., et al., spatialLIBD: an R/Bioconductor package to visualize spatially-resolved transcriptomics data. BMC genomics, 2022. **23**(1): p. 434.

77. Zheng, Y., et al., Aquila: a spatial omics database and analysis platform. Nucleic Acids Research, 2023. **51**(D1): p. D827-D834.

78. Yuan, Z., et al., SODB facilitates comprehensive exploration of spatial omics data. Nature Methods, 2023. **20**(3): p. 387-399.

79. Xu, Z., et al., STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. Nucleic acids research, 2024. **52**(D1): p. D1053-D1061.