Advanced Intelligent Discovery

ADVANCED INTELLIGENT DISCOVERY Open Access

RESEARCH ARTICLE OPEN ACCESS

Composition-Aware Cross-Sectional Integration for Spatial Transcriptomics

Qishi Dong[1] | Zijian Huang[2] | Xuanwu Wang[3] | Zhenghui Feng[2] | Wei Liu[4] | Bingding Huang[1]

[1]School of Artificial Intelligence, Shenzhen Technology University, Shenzhen, China | [2]School of Science, Harbin Institute of Technology-Shenzhen, Shenzhen, China | [3]Department of Mathematics, Hong Kong Baptist University Kowloon Tong, Hong Kong SAR, China | [4]School of Mathematics, Sichuan University, Sichuan, China

Correspondence: Bingding Huang (huangbingding@sztu.edu.cn)

Received: 18 November 2025 | **Revised:** 9 February 2026 | **Accepted:** 20 February 2026

Keywords: cell-type deconvolution | cross-section integration | deep topic model | spatial transcriptomics

ABSTRACT

Spatially resolved transcriptomics (SRT) can profile contiguous tissue sections at near-cellular resolution. However, building multi-slice maps remains difficult because each spot mixes multiple cell types and section-specific batch effects disrupt shared structure. Most pipelines handle cell-type deconvolution, domain detection, and cross-section integration separately, cluster on gene-level variance rather than cellular composition, and rely on post-hoc alignment that misplaces domain boundaries. To overcome these limitations, we present FUSION, a fast, unified, spatial integration and cell-type composition method that leverages single-cell RNA references to unify multiple tasks. FUSION models sequencing read as arising from a latent topic that captures the transcriptomic program of a reference cell type and aggregates read-level topic probabilities into spot-wise compositional embeddings. Clustering directly in this embedding space provides interpretable, composition–driven domains and aligns homogeneous regions across slices. Through a computationally efficient inference scheme augmented with a distance-aware and sparse self-attention kernel, FUSION minimizes runtime and maintains low memory usage. To align embeddings across sections without erasing biological signal, the FUSION pipeline incorporates a Wasserstein GAN module. Across four datasets spanning 10x Visium, ST, Slide-seq, and Stereo-seq, FUSION improves cell-type deconvolution, sharpens domain boundaries, and reconstructs consistent multi-slice architectures, outperforming existing tools while scaling to larger studies.

1 | Introduction

In multicellular organisms, tissue function arises from the precise spatial arrangement of diverse cell types, where cellular identity and activity are tightly linked to their physical microenvironment [1]. Spatially resolved transcriptomics (SRT) couples whole-transcriptome sequencing with in situ coordinates to reveal how diverse cell types are arranged and interact within intact tissues. Imaging assays such as smFISH [2], MERFISH [3], and seqFISH [4] achieve single-cell resolution for a limited gene panel, whereas next-generation sequencing platforms, for example, 10x Visium [5], Slide-seq [6, 7], DBiT-seq [8], and Stereo-seq [9], profile the entire transcriptome but at a multi-cellular resolution. Recent advances in multi-section protocols have extended the application of SRT technologies into three dimensions by serially sampling neighboring sections and digitally registering them in volumetric maps [10, 11]. However, as each next-generation sequencing spot is an aggregation of transcripts from a heterogeneous mixture of cell types, the accurate interpretation of SRT requires cell-type deconvolution to recover the underlying cellular composition and spatial clustering to delineate the coherent tissue domains. Multi-section SRT data further intensifies the analytical burden due to the introduction of section-separable technical bias. Consequently, constructing a faithful multi-section atlas demands

Qishi Dong and Zijian Huang contributed equally to this study.

a pipeline that performs cross-section deconvolution, domain detection, and robust batch correction to harmonize expression profiles across sections while preserving the biological structure.

Despite methodological advances, existing modular pipelines are hamstrung by semantic drift, where each step learns its own latent representations. Thereby, cell-type proportions, domain labels, and batch-aligned embeddings live in incompatible spaces, making it hard to propagate uncertainty or to enforce biological consistency across tasks. In a typical workflow, the cellular composition of each multicellular spot is first inferred via the integration of SRT data with single-cell RNA-seq (scRNA-seq) atlases [12–19]. Subsequently, the latent representation from deconvolution methods is discarded, and the spots are reprojected for domain detection by primarily measuring gene expression similarity [20–25]. In the end, batch-alignment tools have learn another representation for harmonizing adjacent sections [10, 26]. Consequently, clustering in a representation decoupled from cell-type composition often yields boundaries that misalign with histological landmarks, splits mixed spots inconsistently across sections, and undermines downstream analyses such as ligand–receptor inference and trajectory inference [27, 28]. Yet modular pipelines ignore this principle, and cluster spots purely according to gene-level variance. When these tightly linked tasks are solved in isolation, domain boundaries drift away from the underlying cellular ecology, and batch correction no longer respects either structure, leading to biologically inconsistent results.

In addition, scaling SRT analysis beyond a few adjacent sections quickly overwhelms standard pipelines, as the computational and memory footprint grows quadratically with spot size. As data collection expands from single-slice to multi-section or subcellular-resolution profiles [9], spatial graph construction rapidly becomes intractable, with pairwise distances and memory requirements exceeding commodity GPU capacity. Contemporary pipelines typically start analyzing multi-section SRT data with reference-based deconvolution, followed by spatial clustering and batch correction [10, 26], but they require either per-section processing [12] or a unified spatial graph across slices [13, 20], making cross-section alignment computationally costly. In practice, these multiplicative costs translate into tens of gigabytes of GPU memory and multi-hour runtimes, even for mid-sized (8–12 slices with around 4 000 spots in each slice) studies [29, 30]. This effectively caps the capacity of existing pipelines at a handful of sections or forces the use of aggressive down-sampling, which sacrifices biological resolution.

To address the challenges associated with analyzing multi-section SRT data and overcome the drawbacks of existing computational methods, we developed **F**ast, **U**nified, **S**patial **I**ntegration and cell-type c**O**mpositio**N** (FUSION), an efficient method that combines spatially informed clustering, batch integration, and cell-type deconvolution for multi-section SRT data analysis. FUSION models each sequencing read as drawn from a latent topic, which captures the transcriptomic program of a reference cell type. Aggregating the read-level topic probabilities for each spot serves as a cell-type compositional embedding whose softmax normalization yields cell-type proportions. Uniquely, FUSION undertakes interpretable domain detection by clustering spots in the embedding space. By matching these shared embedding signatures, FUSION seamlessly aligns multi-section datasets to yield anatomically coherent domains, without depending solely on transcriptomic similarity. The use of a black-box variational inference algorithm [31] is proposed that propagates neighborhood information through a sparse attention kernel with stochastic mini-gene batch gradient updates, keeping FUSION fast and memory-lean for multi-section SRT datasets. A lightweight Wasserstein GAN [32] further harmonizes section-specific embeddings with minimum signal erasing. We demonstrate the versatility of FUSION using four published spatial transcriptomics (ST) datasets, including the human dorsolateral prefrontal cortex (DLPFC) dataset from 10x Visium, ST data for twelve human squamous cell carcinoma (SCC) slides, two mouse olfactory bulb (MOB) Stereo-seq and Slide-seq slices, and a mouse embryo Stereo-seq dataset from eight time points.

2 | Results

2.1 | Overview of FUSION

FUSION utilizes spot-by-gene expression matrices, spatial coordinates across multiple sections, and scRNA-seq reference profiles to regularize cell type assignments. By casting multi-section SRT counts as a hierarchical topic model that couples cell–type mixtures with the coarse tissue architecture, FUSION extracts spot-separable cell-type compositional embeddings, enabling reference-informed cell-type deconvolution and spatial domain detection (Figure 1a). For spatial spot i of section k, FUSION treats the observed expression $\mathbf{x}_{ki} = (x_{ki1}, \ldots, x_{kiP})^\top$ of P genes as the sum of its W_{ki} single-molecule reads (unique molecular identifiers, UMIs): $\mathbf{x}_{ki} = \sum_d^{W_{ki}} \boldsymbol{x}_{kid}$, where each $\boldsymbol{x}_{kid}$ is a one-hot indicator ($x_{kidp} = 1$ if the d-th UMI maps to gene p). FUSION assumes every UMI $\boldsymbol{x}_{kid}$ is drawn from a latent topic $\boldsymbol{t}_{kid}$ that corresponds to the gene program of a reference cell type. The choice of topic is modulated by a spot-level domain label $\boldsymbol{c}_{ki}$ that captures the coarse tissue architecture shared across sections. Thus, summing the topic probability vector $\boldsymbol{\theta}_{kip}$ (topic probability when observing a UMI map to gene p) over all UMIs produces a cell-type compositional embedding $\boldsymbol{r}_{ki}$ for each spot, and its softmax normalization yields spot-separable proportions $\boldsymbol{p}_{ki}$. As certain cell types may have a spatial preference for specific tissue compartments across sections, FUSION uniquely maps the embeddings $\boldsymbol{r}_{ki}$ to a shared latent domain space via neural networks, assigning domain labels $\boldsymbol{c}_{ki}$ by capturing the cell type resemblance within neighboring microenvironments [33, 34]. Consequently, the resulting domain boundaries track the heterogeneity in cell-type mixtures rather than mere transcriptional variance.

FUSION tackles the computational bottlenecks of multi-section SRT through an efficient black-box variational scheme that operates by mini-batching across genes and has a sparse attention mechanism. At every gradient step, FUSION draws a small set of genes and updates only the corresponding topic probabilities, making the computation cost scale linearly with subset size, while the updates remain unbiased because the model factorizes independently across genes. Spatial information is propagated through a k-nearest sparse-attention operator whose cost scales linearly with the number of spots. These two design choices keep memory use and runtime low enough to process the whole-transcriptome SRT dataset from multiple sections. Training alternates between updating the topic probabilities, as guided by scRNA-seq reference profiles, and refining the spot-separable domain labels with a graph Laplacian regularizer to enforce

spatial smoothness. To erase section-specific artifacts, FUSION applies a lightweight Wasserstein generative adversarial network (WGAN) that aligns the compositional embeddings of each section to a reference section, producing a batch-corrected latent space with negligible extra computation costs. Because the same compositional embeddings underpin deconvolution, domain detection, and batch alignment, FUSION unifies these three tasks in a single latent space. A reference-free variant of FUSION was also developed to provide a similar workflow suitable for when scRNA-seq data are unavailable.

2.2 | Validation Using Simulated Data

We designed two semi-synthetic benchmarks to test accuracy and efficiency for multi-slice SRT integration. Briefly, Scenario 1 draws on a mouse-intestine scRNA-seq reference to generate spatial spots with known cell-type mixtures and domain structure, allowing controlled assessment of domain recovery. Scenario 2 is derived from a mouse embryo (E16.5) SRT atlas [9] by sampling spots from distinct organs and introducing section-specific perturbations to emulate batch effects and realistic scale. In both

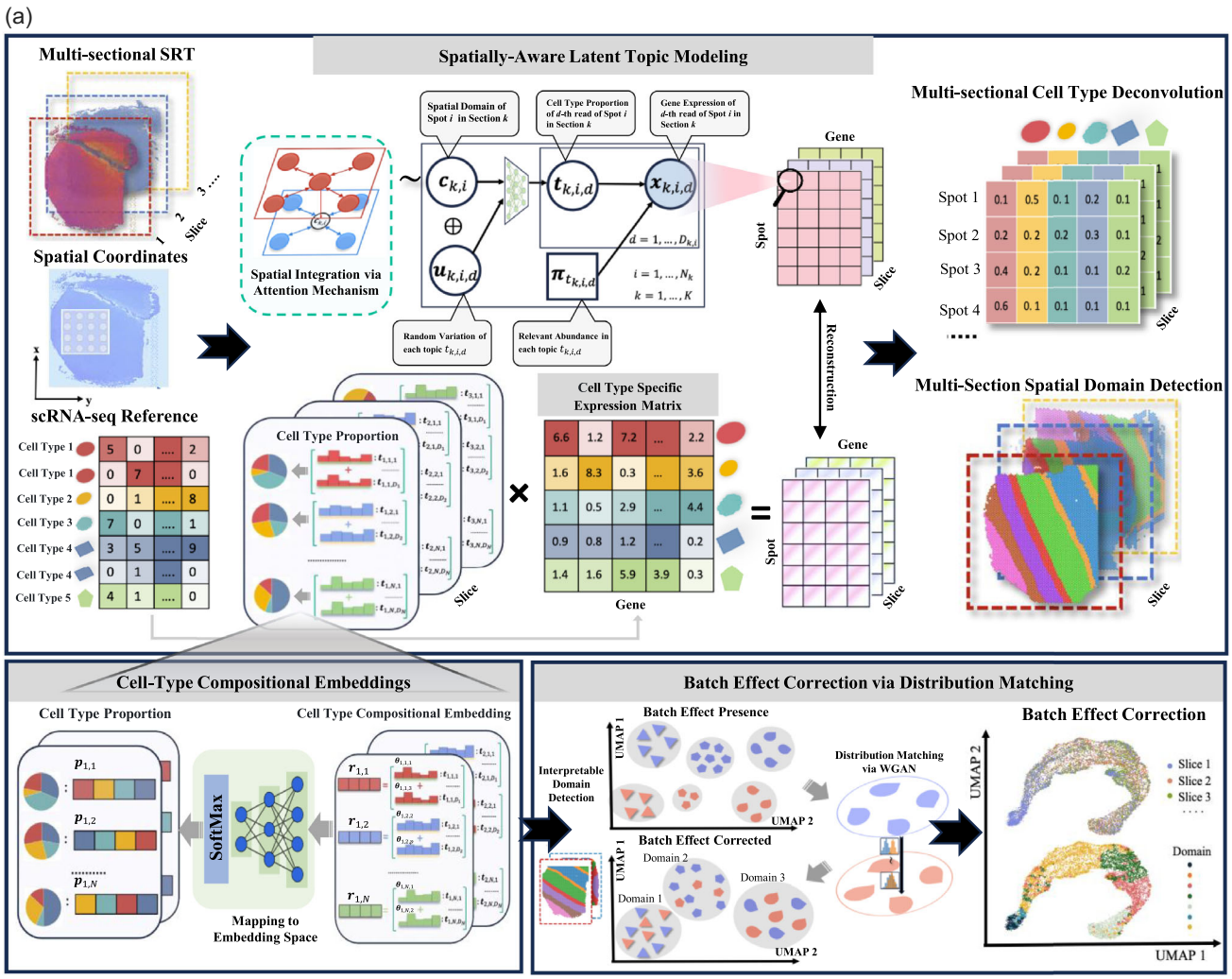

FIGURE 1 | (a) FUSION offers a fast, unified, and interpretable framework for multi-section spatial transcriptomics. It ingests raw UMI counts from multiple tissue sections, per-spot coordinates, and a cell-type–annotated scRNA-seq reference that defines cell-type-specific gene signatures. FUSION directly models each read as originating from a latent topic that captures the transcriptomic program of a reference cell type. Within each spot, a latent spatial domain label is passed through a neural net to sample topics, thereby coupling cellular identity to anatomical context. Aggregating the read-level topic probabilities for each spot forms cell-type compositional embeddings whose softmax-normalization yields quantitative cell-type proportions. Uniquely, FUSION infers domain labels directly from these embeddings with spatial coherence enforced by a distance-aware self-attention kernel and graph regularization, resulting in anatomically coherent, readily interpretable domain detection. All latent probabilities are updated by a black-box variational-inference scheme that operates on mini-batching over genes and sparse attention, keeping memory usage low and making runtime scale almost linearly with the number of reads. A lightweight Wasserstein GAN simultaneously aligns slice-specific embeddings into a common latent space, removing batch effects without erasing biological signal. The unified inference therefore delivers spot-level deconvolution, consistent cross-slice domains, and a batch-corrected 3D tissue map in a single computational pass. (b) Schematic illustration of the two simulation scenarios. The scRNA-seq-derived scenario involved sampling 50 to 160 spots per topic from five cell types across 10 spatial topics to generate three synthetic slices, each containing 900 spots, and applying multiplicative noise at low, medium, and high levels. The scenario derived from real spatial transcriptomics data involved sampling 500 spots to generate a single synthetic slice. For both, three tiers of batch effect strength were applied: Low (0.8−1.2x drift), Medium (0.6−1.4x drift), and High (0.4−1.6x drift). The distribution of ARI values and computational runtime (seconds, log-scale) across 20 independent replicates is shown for each method.

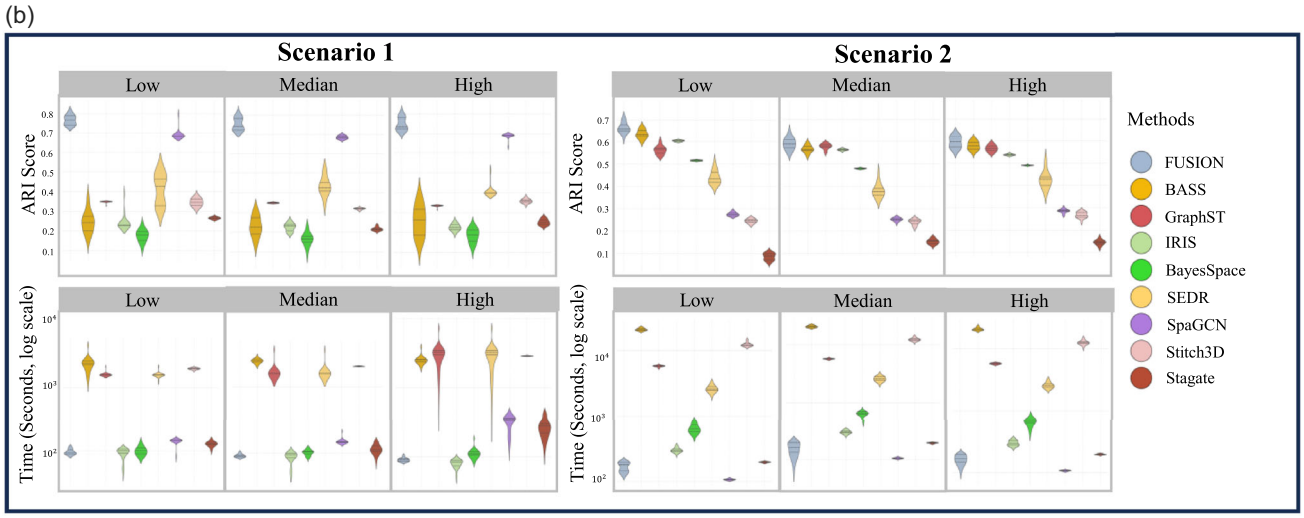

FIGURE 1 | Continued.

scenarios, we created multiple slices under low/medium/high batch-effect settings and compared FUSION against eight state-of-the-art baselines (IRIS [35], BASS [22], BayesSpace [21], SpaGCN [23], SEDR [24], GraphST [20], STAGATE [25], and STitch3D [13]). The simulation details are provided in the "Experimental Section".

To benchmark domain recovery and efficiency, we report Adjusted Rand Index (ARI) and wall-clock runtime on a log scale for Scenario 1 (left) and Scenario 2 (right) under low/medium/high batch settings (Figure 1b). Across these conditions, FUSION consistently demonstrated superior accuracy, attaining at least a moderate advantage over all other methods. In Scenario 1, FUSION achieved median ARIs of 0.77, 0.74, and 0.73 under low/medium/high batch effects, outperforming the next best method (SpaGCN: 0.68, 0.67, 0.69). In addition, FUSION also exhibited competitive runtimes, being among the fastest methods across the benchmarks. Averaging the medians across the three different batch settings, FUSION finished in 90.6 seconds, with a 96.0% runtime reduction compared with BASS, GraphST, SEDR, and Stitch3D (2–133 to 2–470 seconds). In Scenario 2, FUSION led in accuracy (median ARIs = 0.66, 0.59, 0.60), surpassing BASS at low noise (median ARI = 0.63), GraphST at medium (median ARI = 0.58), and BASS at high (median ARI = 0.58). Notably, its accuracy remained around 0.6 even under high noise, while running in a median of 171 to 178 seconds. These simulations indicate that FUSION delivers strong and batch-resilient domain discovery in both controlled and realistically complex settings for multiple slices, with markedly lower computational cost.

2.3 | Application in Human DLPFC Visium Data Analysis

We applied FUSION to the analysis of 12 human DLPFC 10x Visium slices from three donors [27]. Each slice contains manually annotated cortical layers (L1–L6 and white matter) with 33 538 gene expression measurements among a median of 3 844 spots. To evaluate FUSION's performance in cell-type deconvolution, scRNA-seq data from a 10x Chromium analysis of the post-mortem brain tissue [36] were used as reference, covering 44 cell types. We jointly analyzed all 12 integrated slices and compared FUSION against eight state-of-the-art methods [13, 16, 21–25, 35]. Across all slices from different donors, the inferred spatial domains accurately depicted the expected laminar architecture, as evidenced by their sharp alignment with the ground-truth cortical boundaries (Figure 2a; Figure S1–S3). Furthermore, the cell-type compositions inferred by FUSION recapitulated the expected laminar architecture of the DLPFC, with oligodendrocyte lineage cells predominating in white matter (Domain 1) and Layer 6 (Domain 2), typical of regions rich in myelinated axons and subcortical projections [37] (Figure 2a, d). Cortical Layers 1 and 2 (Domains 7 and 6) exhibited elevated astrocyte proportions, consistent with their role in synaptic modulation and neurovascular coupling within superficial cortical layers [38]. By contrast, BayesSpace and CARD showed a blurry boundary between superficial Layer 2 and Layer 3, while SEDR and STAGATE misaligned Layer 1 and 2. IRIS and IRIS-free were unable to identify the white matter region and could barely align Layer 3 in slice 151669. Furthermore, BASS and SpaGCN showed Layer 6 of slices 151669 and 151673 as fragmented and cortical, while CARD, SpaGCN, and SEDR collapsed Layers 2–4, displaying limited multi-section analysis capability. Quantitatively, the domains from FUSION aligned best with the ground-truth partitions (ARI = 0.64 ± 0.09 across 12 integrated slices), surpassing the performance of the reference-based IRIS (ARI = 0.55 ± 0.10), reference-free BASS (ARI = 0.50 ± 0.06), and IRIS-free (ARI = 0.43 ± 0.11) (Figure 2b). FUSION was especially accurate for Donor 1 (mean ARI = 0.70 ± 0.03), and both FUSION and FUSION-free excelled in the analysis of Donor 2 (mean ARI = 0.71 ± 0.05 and 0.58 ± 0.03). FUSION completed its joint analysis of all 12 sections in 834 seconds, outperforming most of the competing methods, whereas SEDR and BASS required more than 2 h with identical hardware.

To evaluate the batch effect correction, we focused on three non-contiguous DLPFC slices 151507, 151669, and 151673 from different donors, providing a stringent test of cross-section harmonization. We visualized the UMAP of inferred compositional embeddings (Figure 2c; Figure S4–S7). The RGB plots for FUSION showed that spots from different slices were well mixed, and there was coherent separation between the identified cortical segments. By contrast, PCA and the benchmark methods resulted in pronounced section-specific clustering and fragmented spatial

organization, confirming the presence of residual batch effects and poor integration. We next quantified the cross-section harmonization using two Local Inverse Simpson's Index (LISI) metrics: iLISI (higher indicates better batch mixing) and cLISI (lower indicates better conservation of biological structure). Across the three representative slices, FUSION achieved the highest iLISI (2.35) and highest ARI (0.59), while attaining the second-lowest cLISI (1.34) for domain separation (Figure 2e). These results indicated that WGAN-aligned embeddings homogenized the batch effects while preserving biologically meaningful heterogeneity between domains.

We performed differentially expressed (DE) gene analysis of the identified spatial domains in slices 151507 and 151673. The domain-specific DE genes aligned with known cortical biology

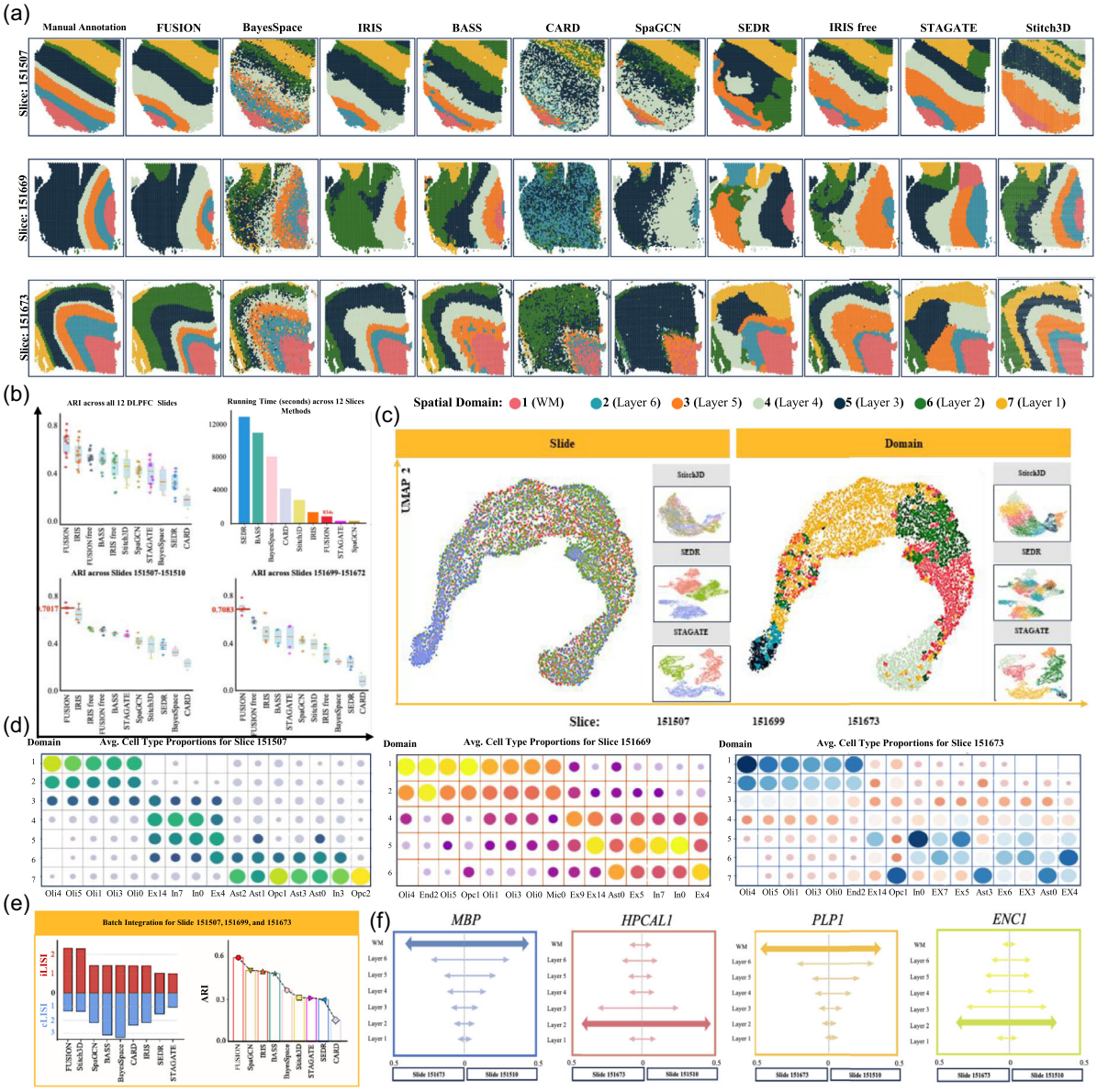

FIGURE 2 | (a) Spatial domain maps are shown for three representative DLPFC slices (151507, 151699, 151673), taken from a joint 12-slice analysis, comparing the manual neuro-anatomical annotations with the segmentations produced by FUSION, BayesSpace, IRIS, BASS, CARD, SpaGCN, SEDR, IRIS-free, STAGATE, and Stitch3D. (b) Top panels: Box and bar plots summarizing ARI scores and running time of benchmarking methods for all 12 DLPFC slices; Bottom panels: Box plots summarizing ARI scores for two subsets of consecutive slices (151507–151510) and (151699–151672), showing ARI values for all slices and the mean ARI value, achieved by each method. (c) Two-dimensional UMAP of FUSION's batch-aligned compositional embedding colored by slice and by inferred spatial domain, with inset UMAPs from Stitch3D, SEDR, and STAGATE for visual comparison. (d) Bubble plots showing the average cell-type composition of each FUSION domain in three DLPFC slices (151507, 151669, 151673); bubble radius encodes proportion and color distinguishes slices. (e) Batch-integration performance on slices 151507, 151699, and 151673 across competing methods, summarized by stacked-bar charts of iLISI and cLISI and a line plot of cross-slice ARI. (f) Layer-resolved strip plots of canonical marker genes (*MBP*, *HPCAL1*, *PLP1*, *ENC1*) for slices 151673 and 151510, each with reproduced marker gradient across sections. Arrow length denotes mean normalized expression within each cortical layer and white matter. (g) Stacked heatmaps showing the FUSION-inferred proportion of oligodendrocytes across six DLPFC sections, revealing the continuous white-matter band in 3-D space. (h) Construction of two meta-genes: spatial expression patterns of canonical markers (*MBP* + *PLP1* – *PTGDS*, top row; *MBP* + *PLP1* – *SYT1*, bottom row) and their resulting meta-gene maps, colored by normalized expression level.

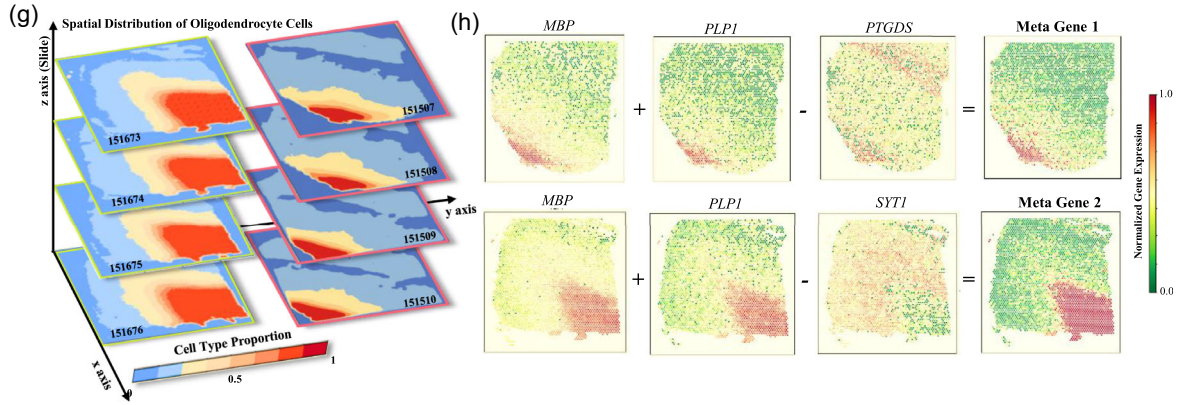

FIGURE 2 | Continued.

features (Figure 2f), that is, *MBP* and *PLP1*, which are canonical markers of the myelinating oligodendrocytes prolific in white matter (Domain 1) [39]; and *HPCAL1* and *ENC1*, the expression of which is enriched in upper Layer 2/3 excitatory neurons [27]. The spatial proportions of oligodendrocytes inferred by FUSION strongly correlated with the regions of white matter across sections, confirming their biological role in myelination (Figure 2g).

To further dissect domain-specific transcriptional activities, we constructed meta-genes combining the expression of key marker genes (Figure 2h). For the white matter region of slice 151507, the meta-gene of *MBP*, *PLP1*, and *PTGDS* outperformed *MBP* alone in white-matter-specificity by suppressing the influence of *PTGDS*, which is expressed in meningeal and subpial astrocytes. Similarly, the meta-gene in slice 151673 excluded *SYT1*, which encodes a synaptic vesicle protein abundant in gray matter, to sharpen the distinction between white and gray matter. As such, the results showed that FUSION successfully characterized the spatial landscape and revealed the cellular structure of the human DLPFC sections.

2.4 | Application in Human SCC ST Data Analysis

We further analyzed human SCC ST data, comprising 12 tissue slices from four patients, each with a median of 629 spots [40]. Using scRNA-seq reference data from the same study, FUSION resolved the spatial and cellular heterogeneity across the slices. In an analysis of serial slices from each patient, FUSION inferred cohesive spatial domains with high intra-domain transcriptional consistency, whereas the competing methods yielded fragmented clusters and anomalous spot assignments that disrupted the overall domain continuity (Figure 3a; Figure S8–S15). The batch-corrected compositional embeddings likewise exhibited superior integration across sections while preserving domain specificity (Figure 3b). UMAP visualization revealed a tight mixing of intra-patient spots, demonstrating effective batch correction, whereas the embeddings from the competing methods exhibited either pronounced section-specific clustering or a failure to reconstruct cohesive spatial domain hierarchies in the embedding space (Figure S16–S18). We quantified the compactness and separation of the UMAP embeddings across different sections via iLISI and cLISI. Across the four patients, FUSION achieved strong cross-section harmonization (mean iLISI = 2.49), while keeping domains well separated (mean cLISI = 1.57) (Figure 3c). Stitch3D showed the highest iLISI (mean iLISI = 2.71) but at the cost of weaker structure preservation (mean cLISI = 1.73). Relative to other methods, FUSION consistently combines strong cross-slice mixing with low cLISI, yielding the best overall balance between batch removal and biological domain conservation.

We then compared the spatial domains of each method using a silver-standard mask of histologist-defined tumor versus non-tumor regions in slice P2_ST_rep2 [35] (Figure 3d). We annotated each detected domain as tumor or non-tumor via DE-based correlation with curated tumor and stromal gene sets. FUSION split the tumor region into four subregions (Domains 1, 6, 10, and 11) that aligned with histological leading edges (ARI = 0.67). In contrast, all other candidate methods failed to separate the immune-infiltrated margins from the tumor edges (Figure 3e; Figure S19). BayesSpace produced relatively fragmented domain segmentations. BASS, IRIS, and STAGATE generated over-smoothed tumor boundaries that were misaligned with the invasive front, while the domains of Stitch3D extended beyond the histological margins, conflating the stromal and tumor microenvironments. Additionally, gene set enrichment analysis (GSEA) of the FUSION-defined domains revealed them to have spatially polarized PI3K-Akt signaling (Figure 3f). The non-tumor stromal regions (Domain 4) showed significant PI3K-Akt pathway activation (normalized enrichment score NES = 2.162, P-adj < 0.001), consistent with tumor-adjacent stromal remodeling and immune crosstalk [41]. In contrast, Domain 6 (tumor core) showed significant negative pathway enrichment (NES = −1.808). This suppression of PI3K-Akt activity in the tumor core aligns with prior reports of pathway downregulation in densely packed tumor regions due to metabolic stress or compensatory feedback mechanisms [42]. The dichotomous pathway activity of these domains underscores FUSION's ability to reveal biologically plausible microenvironmental heterogeneity [43].

In addition, FUSION reconstructed spatially complementary tumor-versus-immune distributions across all slices (Figure 3g; Figure S20). In slice P2_ST_rep2, proliferative tumor cells (Tumor_KC_Basal) were precisely localized at invasive fronts (Domain 10), while differentiated tumor cells (Tumor_KC_Diff) predominated in the tumor core (Domains 6 and 9). Cycling tumor cells (Tumor_KC_Cyc) were spatially proximal to NK cells, suggesting the presence of active immune-mediated tumor surveillance. T cells, B cells, and NK cells were markedly enriched within stromal regions (Domains 4, 5, and 7), where they formed spatial patterns

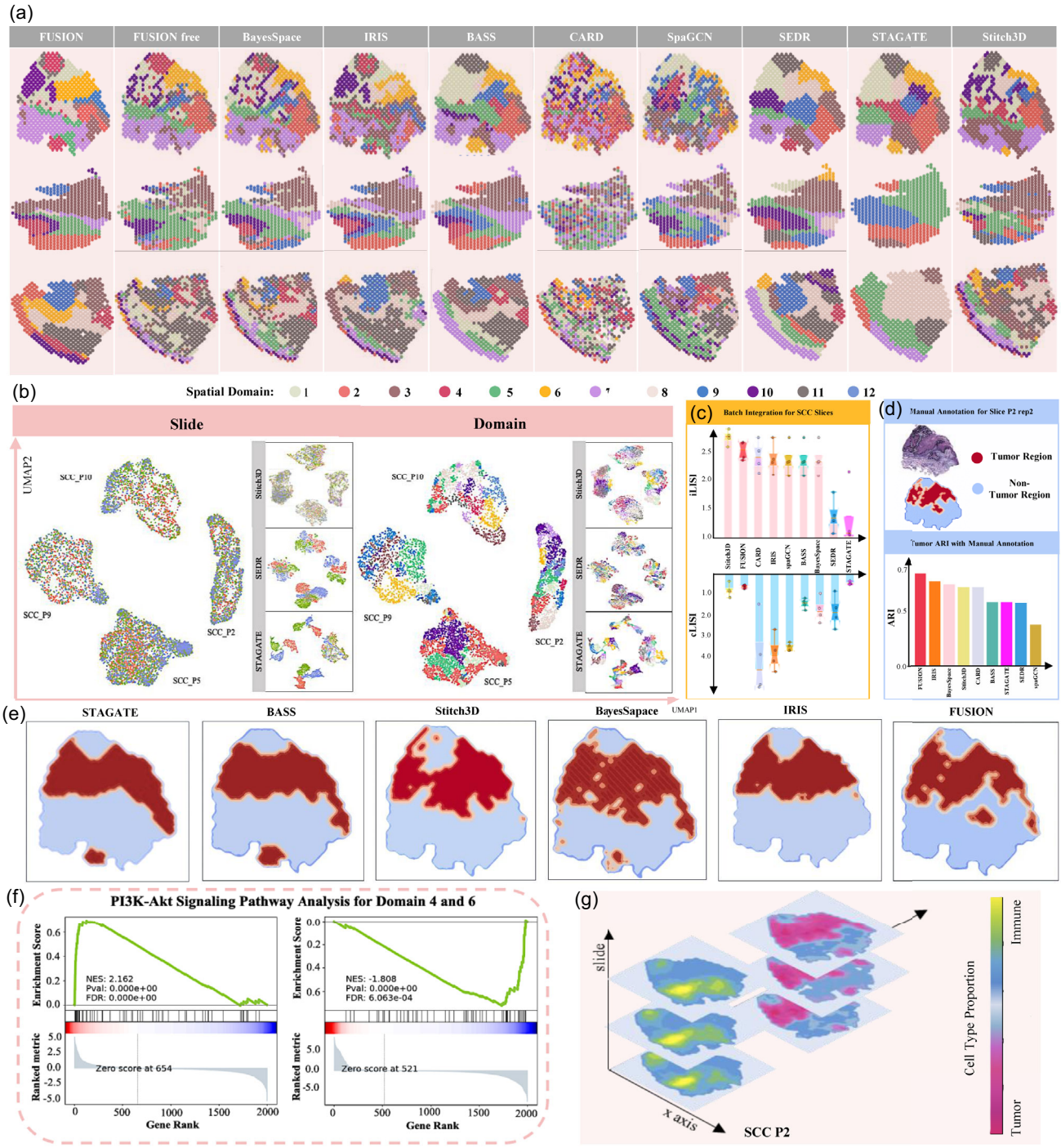

FIGURE 3 | (a) Spatial domain maps for three SCC Spatial-Transcriptomics slides (P2_ST_rep2, P5_ST_rep3, and P10_ST_rep1) generated by FUSION, FUSION-free, BayesSpace, IRIS, BASS, CARD, SpaGCN, SEDR, STAGATE, and Stitch3D; colors indicate the inferred spatial domains. (b) UMAP visualizations of the FUSION compositional embedding for four SCC patients (P2, P5, P9, P10): points colored by slide identity (left) and by FUSION-inferred spatial domain (right); inset panels display the corresponding embeddings returned by Stitch3D, SEDR, and STAGATE for the same four slides. (c) Bar plot of iLISI and cLISI scores measuring batch integration quality for patients P2, P5, P9, and P10, reported for each benchmark method. (d) Top panel: Hematoxylin-and-eosin section of slice P2_ST_rep2 displaying tumor leading edges and the corresponding ground-truth mask separating tumor (red) from non-tumor (blue) tissue; Bottom panel: Bar plot of ARI scores that compare each method's tumor / non-tumor mask in (e) to the manual ground-truth mask. (e). Tumor / non-tumor maps produced by each algorithm after assigning its inferred domains to one of the two classes using domain-specific DEGs. (f) Gene-set enrichment plots for the PI3K–Akt signaling pathway in Domains 4 and 6; the green curve gives the running enrichment score, with the corresponding NES, nominal *p*-value and FDR shown in each panel. (g) Stacked heatmaps of FUSION-estimated tumor-versus-immune cell-type proportions for SCC patient 2, displayed in serial-section order to visualize through-slice consistency.

complementary to tumor-dominated areas [44]. In comparison, the competing methods were unable to clearly separate immune-infiltrated margins from tumor edges and failed to characterize the intricate cellular heterogeneity within these representative slices.

2.5 | Application in Mouse Olfactory Bulb Stereo-Seq and Slide-Seq Data Analysis

To assess the cross-platform integration capability of FUSION, we analyzed paired MOB ST datasets from the Slide-seq and Stereo-seq platforms. These datasets capture the layered architecture of the MOB [45], including the olfactory nerve layer (ONL), rostral migratory stream (RMS), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), and granule cell layer (GCL). Using the 10x Chromium scRNA-seq data from the same tissue as the reference [46], FUSION resolved all seven MOB layers on each platform separately with fine-grained precision, yielding coherent boundaries and the delineation of the subtle dorsomedial-to-ventrolateral sublayers between the GL and EPL [47] (Figure 4a). Cross-platform integration by FUSION further revealed spatially aligned domains consistent with known histological layers (Figure 4b). FUSION successfully reconstructed a distinct MCL (Domain 6) sandwiched between the GCL (Domain 1) and GL (Domain 4) in the joint analysis, preserving cross-platform consistency. In contrast, BayesSpace produced highly fragmented domains with blurred layer boundaries, while IRIS and STAGATE severely over-smoothed the spatial structures, collapsing the distinct layers of GCL into broader regions. BASS, CARD, and STAGATE yielded platform-specific domain labels, indicating substantial batch effects. SpaGCN and SEDR failed to separate the GCL from the neighboring IPL, and Stitch3D could not align the two platforms due to coordinate mismatching.

FUSION's inferred cell-type compositions recapitulated the MOB laminar architecture (Figure 4c), with granule cells (GC 1–6) localized to the GCL (Domain 1) and mitral/tufted cells (M-TC) concentrated in the MCL (Domain 6). These compositions showed strong cross-platform agreement based on the high correlations among the cell-type proportions across shared domains (Figure S21) and close alignments between the inferred PGC proportions in the GL-associated domains and the GC proportions in the GCL (Figure 4d; Figure S21). FUSION's batch-corrected cell-type compositional embeddings further revealed the seamless integration of Slide-seq and Stereo-seq spots in the UMAP space, with well-mixed, yet clearly clustered, spatial spots (Figure 4e). In contrast, embeddings from competing methods retained the platform-specific clustering, even after batch effect correction [26], reflecting their unresolved technical variability (Figure S22, S23). Quantitatively, FUSION achieved the strongest cross-platform integration (iLISI = 1.71) while preserving biological specificity (cLISI = 1.49) (Figure 4f). In contrast, applying Harmony to other methods typically raised iLISI by 40% but typically at the cost of 30% higher cLISI, indicating loss of structure. The WGAN-based alignment of FUSION achieved the best balance between section mixing and structure preservation.

DE analysis of the two datasets (Figure 4g) further confirmed the expression of domain-specific marker genes of the olfactory bulb, that is, *MBP* for myelinating oligodendrocytes [48] (DE genes of Domain 1), *CCK* for neurotransmission [49] (DE genes of Domain 4), and *AUTS2* for neurodevelopmental processes [50] (DE genes of Domain 2). The expression of these genes aligned with their role in axonal insulation and exhibited concordant spatial enrichment patterns in data from both Slide-seq and Stereo-seq platforms. Furthermore, we revealed cross-platform consistency in pathway enrichment via GSEA (Figure 4h). Domain 2 exhibited the activation of lipid-centric pathways (PPAR signaling, arachidonic acid metabolism) in both the Slide-seq and Stereo-seq slices, aligning with the metabolic demands of olfactory signal processing [51]. Domain 8 showed robust cross-platform enrichment for PPAR signaling and lipid metabolism (unsaturated fatty acid biosynthesis in Stereo-seq), consistent with GCs' reliance on lipid homeostasis for inhibitory neurotransmission [52, 53]. The coherence of these pathways with layer-resolved functions demonstrates the ability of FUSION to harmonize cross-platform SRT data while preserving the domain-specific biology.

2.6 | Application in Mouse Embryogenesis Stereo-Seq Data Analysis

To rigorously evaluate FUSION's capacity for large-scale cross-temporal integration, we analyzed comprehensive Stereo-seq ST data encompassing over 540 000 spots across eight developmental stages (E9.5–E16.5) of mouse embryogenesis [9], a subcellular-resolution dataset that poses formidable computational demands. Using *Mus musculus* single-cell atlas data [54] as reference, FUSION accurately recovered the spatial layout of key embryonic structures, including the liver, brain and spinal cord, heart, and choroid plexus across all stages. The inferred domains aligned closely with the tissues annotations, in terms of both shape and relative position, and faithfully reflected the morphogenetic changes seen over developmental time (Figure 5a). The cell-type distributions inferred by FUSION allowed the delineation of coherent developmental trajectories for major organ systems, including the liver parenchyma, neural tube-derived structures (brain and spinal cord), meningeal layers, and cardiac tissues. FUSION captured the spatiotemporal expansion dynamics of organ-specific cell populations, such as intestinal epithelium precursors, differentiating neuronal subtypes, adipogenic mesenchymal stem cells, and vascular endothelial networks, all of which exhibited progressive territorial expansion from E9.5 to E16.5. These reconstructions showed striking consistency with established histogenesis principles [55–58] and accurately recapitulated both the timing of emergence and spatial positioning of developing anatomical structures (Figure 5b; Figure S24). Notably, FUSION revealed coordinated cell-proportion shifts during organ maturation, in which there was a decrease in hematopoietic lineage cell populations (late pro-B cells and B cells) concomitant with an increase in neuron density in Domain 5 (developing forebrain) and expansion in myeloid populations in Domain 18 (cardiac tissue) (Figure 5c). These dynamics align precisely with murine developmental programs, in which neural tissue undergoes diversification while cardiac immune niches are populated by tissue-resident macrophages [59, 60].

The cell-type compositional embeddings generated by FUSION demonstrated continuous temporal progression across the anatomical domains in PCA space, revealing pseudotemporal trajectories that reflected unbroken developmental continuity between stages (Figure 5d). This embedding structure preserved both organ-specific clustering and transitional intermediates between timepoints. Quantitative correlation analysis further confirmed

that there were exceptionally high positive associations in cell-type proportion covariation across the developmental timepoints and spatial domains, with the highest concordance observed among stroma-associated lineages, for example, endothelial cells, fibroblasts, adipose mesenchymal stem cells, and large intestinal epithelium, indicating the conserved developmental programs at

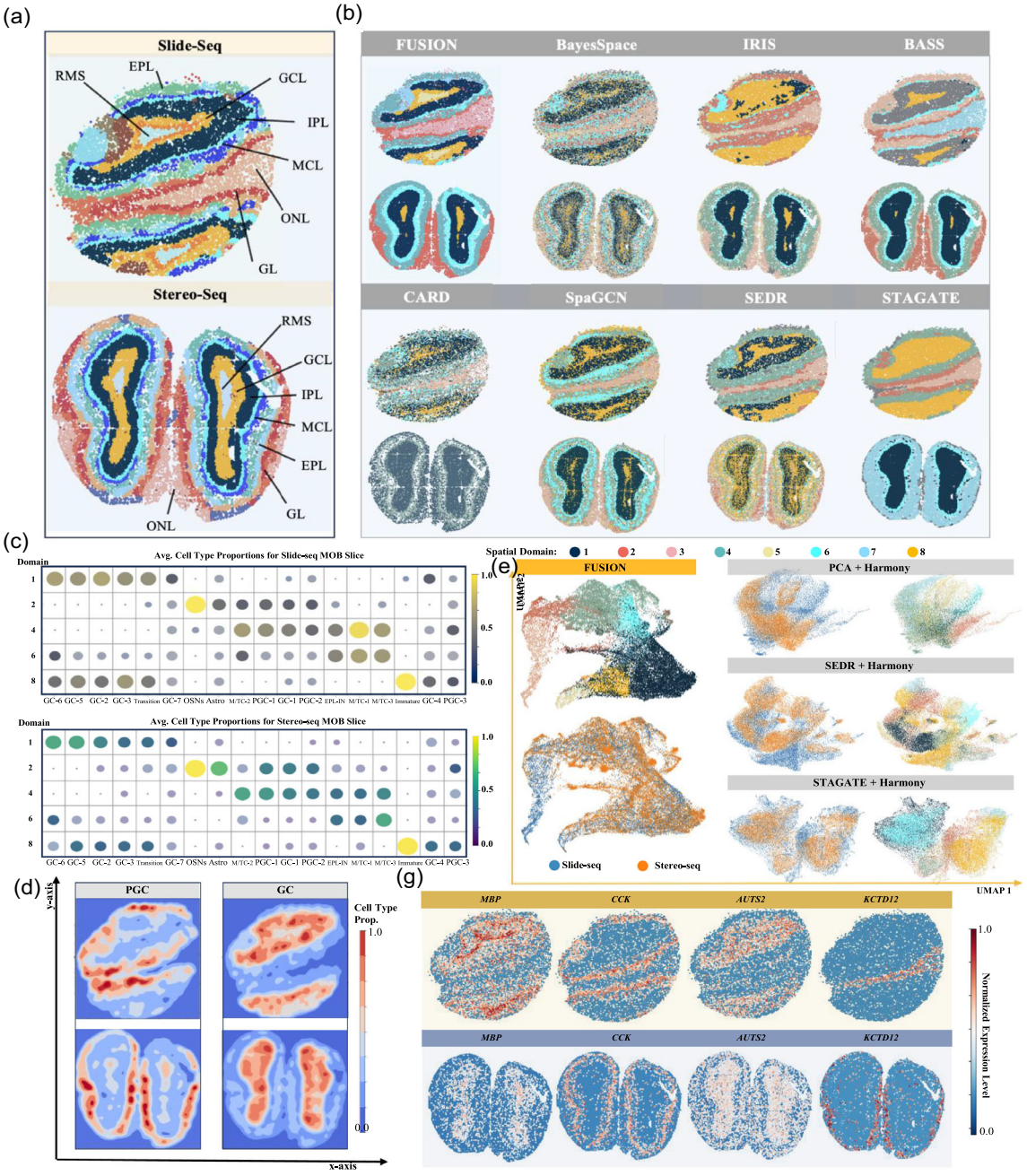

FIGURE 4 | (a) Layer-resolved spatial domain maps for a Slide-seq mouse olfactory-bulb section and a matched Stereo-seq section generated by FUSION on each slice independently with annotated canonical layers—olfactory nerve layer (ONL), rostral migratory stream (RMS), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), and granule cell layer (GCL). (b) Cross-platform domain maps for the two MOB sections obtained from a joint slice analysis with FUSION and seven benchmark methods: BayesSpace, IRIS, BASS, CARD, SpaGCN, SEDR, and STAGATE. (c**)** Bubble plots of average cell-type composition for each FUSION domain in the Slide-seq MOB section and the matched Stereo-seq section; bubble size encodes the mean proportion of the indicated cell type within the domain. (d) Heatmaps of FUSION-inferred cell-type proportions for periglomerular cells (PGC, left) and granule cells (GC, right) in the Slide-seq slice and the Stereo-seq slice. (e**)** UMAP plots of the FUSION compositional embedding colored by spatial domains and by platform, alongside embeddings produced by PCA, SEDR, and STAGATE with Harmony correction for comparison. (f**)** iLISI and cLISI scores for batch integration of the Slide-seq and Stereo-seq MOB slices, shown for each benchmark method with (dark shading) and without (light shading) an additional Harmony correction. (g) Spatial heatmaps of normalized expression level of four representative marker genes: *MBP*, *CCK*, *AUTS2*, and *KCTD12*, displayed for the Slide-seq MOB slice and the matched Stereo-seq slice. (h**)** Bar plots of KEGG pathways significantly enriched in cross-platform Domain 2 and Domain 8; red bars indicate pathways up-regulated in Slide-seq, blue bars those up-regulated in Stereo-seq slices.

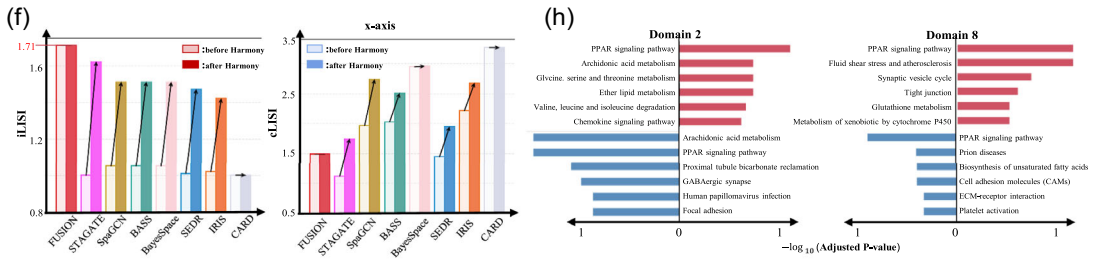

FIGURE 4 | Continued.

play in these populations (Figure 5e). To assess cross-platform consistency, we applied FUSION to the E16.5 Stereo-seq and Visium [5] mouse embryo slice data, and we observed highly concordant cell-type maps for fibroblasts, myeloid cells, and neurons (Figure 5f). Fibroblasts delineated broad mesenchymal layers along the body wall and craniofacial/limb regions, myeloid cells were localized to hematopoietic niches in the fetal liver and spleen primordia, and neurons sharply outlined the brain and spinal cord. Despite differences in resolution and sectioning plane, anatomically consistent distributions appeared in the data

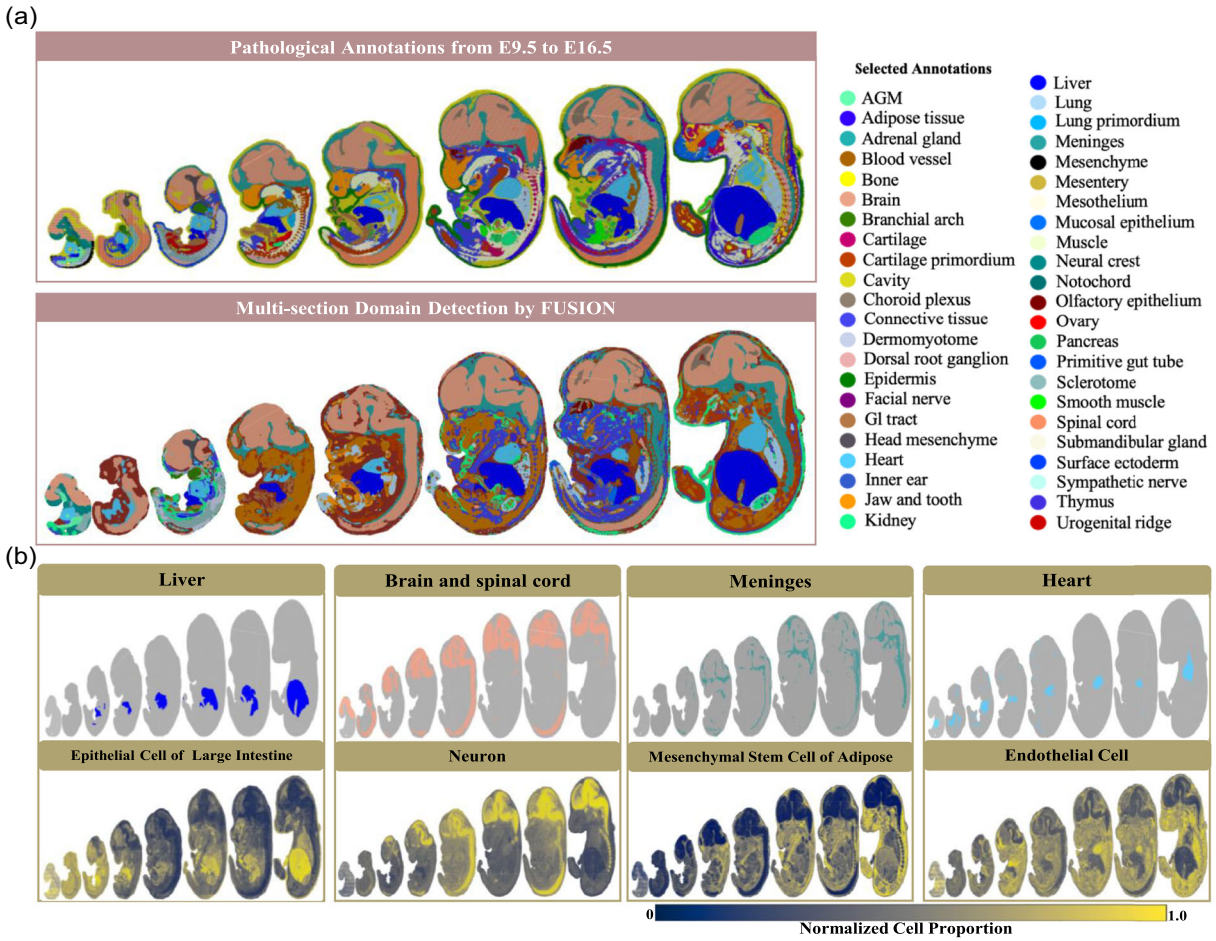

FIGURE 5 | (a) Comprehensive pathological annotations (45 anatomical regions) of mouse embryos from E9.5 to E16.5 alongside multi-section spatial domain detection results generated by FUSION. (b) Isolated views of four domains closest to key anatomical structures (liver, brain and spinal cord, meninges, and heart), with corresponding temporal proportion profiles of associated cell types (epithelial cell of large intestine, neuron, mesenchymal stem cell of adipose, and endothelial cell) across E9.5 to E16.5. (c) Developmental kinetics of five cell types across developmental stages in two FUSION domains from the mouse embryo Stereo-seq series. Stacked bars show the relative abundance of late pro-B cell, B cell, leukocyte, neuron, myeloid cell within Domain 5 (top panel) and Domain 18 (bottom panel) at embryonic days E9.5–E16.5; percentages printed on the first and last bars highlight the temporal shift in dominant cell types. (d) Time-lapse trajectories of FUSION's embeddings across mouse-embryo development. Compositional embeddings for every spot from stages E9.5 – E16.5 are projected onto the first two principal-component (PC) axes; points are colored by embryonic stage and grouped by the spatial domains identified by FUSION. Thin arrows trace the centroid of each domain through developmental time, revealing domain-specific maturation paths in cell-type composition. (e) Pair-wise Pearson correlation matrix of estimated cell-type proportions across the mouse embryo dataset, revealing coordinated abundance patterns among the inferred cell types. (f) Method validation with Visium data comparing with mouse embryo E16.5 in fibroblast, myeloid cell, and neuron.

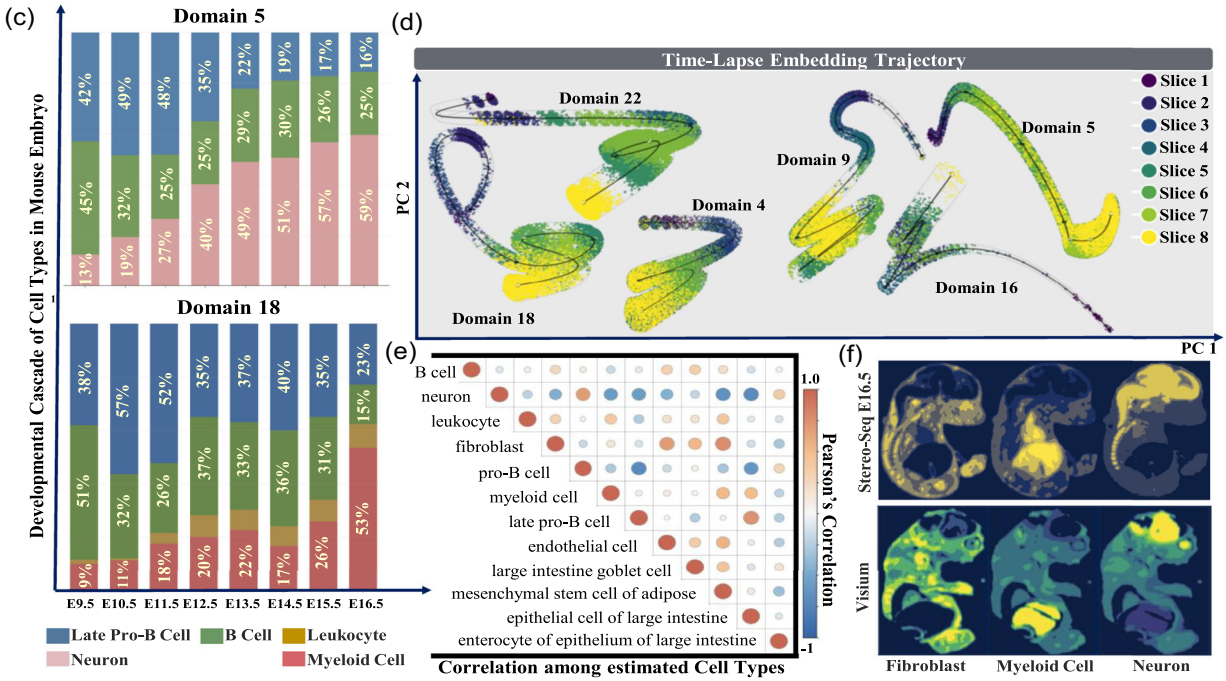

FIGURE 5 | Continued.

from both platforms, providing mutual validation of the deconvolution results and highlighting the robustness of FUSION in integrating and interpreting spatial data across technologies.

3 | Discussion

Spatially resolved transcriptomics is rapidly being applied to multi-section, multi-platform profiling, yet most pipelines still tackle cell-type deconvolution, spatial clustering, and batch correction in isolation. We thus designed FUSION to be a deep probabilistic model that unifies these three tasks for multi-section SRT data in one framework. By modeling gene counts as mixtures of cell-type–specific topics, FUSION derives interpretable, spot-separable embeddings that capture the compositional heterogeneity and align with the tissue microanatomy. Unlike existing methods, FUSION explicitly integrates spatial cohesion through distance-aware self-attention and graph Laplacian regularization, while harmonizing multiple sections through a WGAN to mitigate batch effects without sacrificing biological specificity. In comprehensive benchmarking with four public SRT datasets, FUSION demonstrated superior performance in resolving spatially coherent domains and harmonizing multi-section data, outperforming state-of-the-art tools in various criteria. FUSION can also reveal biologically meaningful cell-type distributions, such as the immune-tumor spatial complementarity in SCC, and uncover pathway polarizations aligned with histological functions.

A distinctive feature of FUSION is that it treats the cell-type composition as the organizing signal for domain discovery. Unlike approaches that rely on static marker gene lists or external references, FUSION dynamically infers cell-type-specific transcriptional topics and clusters directly within the resulting embedding space, automatically reconciling domain boundaries with cellular content and avoiding the over-smoothing of layers or fragmented clustering associated with transcription-only approaches. Computationally, FUSION is engineered for speed and memory efficiency. Its mini-gene black-box variational scheme touches only a small fraction of genes per update, so the computational cost scales with the mini-batch size, rather than with the full transcriptome. The cost of the k-nearest sparse-attention kernel grows linearly with the number of spots, enabling whole-transcriptome training across a multi-section SRT dataset. This scalability was critical for analyzing the mouse embryo Stereo-seq atlas, which spans eight developmental stages and over 540 000 spatial spots. FUSION efficiently handled the more than 20-fold increase in spot count from the early (E9.5) to late (E16.5) stages, and its formulation accommodated the spatial discontinuities caused by morphogenetic transformations. Together, these efficiencies enabled FUSION to integrate complex spatiotemporal data that exceeded the computational limits of existing pipelines.

While FUSION represents a significant advancement in multi-section spatial transcriptomic analysis, there are several challenges remaining that warrant further investigation. First, FUSION currently assumes there is a consistent resolution across sections, but emerging sequencing technologies are increasingly producing multi-resolution datasets, where subcellular and multicellular measurements coexist within or across tissue sections. Though FUSION demonstrated a promising cross-platform-harmonization ability in its analysis of Stereo-seq and Slide-seq MOB datasets, its graph-based spatial regularization may struggle in hybrid-resolution settings, potentially blurring critical nanoscale tissue features. Second, FUSION's adversarial harmonization framework maps the embeddings of all sections to a single anchor section, which risks overfitting to the anchor's technical artifacts or biological idiosyncrasies. A more robust method would involve mutual alignments via the iterative harmonization of all sections against a dynamically learned consensus embedding, rather than a static anchor. However, our current implementation prioritizes computational tractability and scalability in large datasets. Potential improvements to mitigate this limitation include adopting cycle-

consistent adversarial networks or optimal transport-based multi-reference frameworks. Furthermore, the performance of FUSION can be sensitive to hyperparameters such as topic number, adjacency radius, and balancing weights. Developing automatic model-selection criteria and more stable optimization schedules would make the framework easier to deploy for non-specialists. Taken together, future expansions could allow FUSION to provide a robust solution for dissecting cell-type architecture and spatial domains in complex, heterogeneous, and multi-section SRT studies.

4 | Experimental Section

4.1 | Model Specification

FUSION treats all profiled sections as one dataset and builds a single hierarchy that simultaneously deconvolves cell types, assigns spatial domains, and aligns batches within a common latent space. It simultaneously models the gene expression counts of K SRT slices, each of which is profiled for P genes. We denote $\boldsymbol{X}_k$ as the k-th profiled spot-by-gene spatial expression matrix, consisting of n_k spots. The corresponding spatial spots coordinate is indexed as $i = 1, \ldots, n_k$ with $\mathbf{s}_{ki} \in \mathbb{R}^2$. As such, $\mathbf{x}_{ki} = (x_{ki1}, \ldots, x_{kiP})^\top$ represents the gene count vector for spatial spot i. FUSION models the gene expression as the aggregation of single-molecule reads in SRT sequencing. The library size W_{ki} is the total number of counted reads (UMI) of $\mathbf{x}_{ki} = \sum_d^{W_{ki}} \mathbf{x}_{kid}$, where $\boldsymbol{x}_{kid}$ is a one-hot indicator vector and $x_{kidp} = 1$ if a sequencing read records the p-th gene expression. FUSION assumes each read, $\boldsymbol{x}_{kid}$, is associated with one of the T latent topic labels, $t_{kid} \in \{1, \ldots, T\}$, and one spatial–domain label, $\mathbf{c}_{ki} \in \{0,1\}^L$ for L, domains to capture the macroscopic tissue context shared across slides. To perform multi-section SRT analysis, FUSION requires K transcriptomic profiles $\{\mathbf{X}_k\}_{k=1}^K$, and their corresponding spatial coordinates, along with optional scRNA-seq reference data containing information on the T cell types for deconvolution. FUSION casts multi-sectional SRT counts as a hierarchical topic model that links cell-type mixtures to the coarse-tissue architecture. The generative model is given by:

$$
\begin{aligned}
\mathbf{x}_{kid} &\sim \text{Multinomial}(\boldsymbol{\pi}_{t_{kid}}),\\
\boldsymbol{\pi}_t &\sim \text{Dirichlet}(\boldsymbol{\alpha}_0) \in \Delta^{P-1},\\
t_{kid} &\sim \text{Multinomial}(\boldsymbol{\theta}_{kid}),\\
\boldsymbol{\theta}_{kid} &= \text{Softmax}\big(f_{\boldsymbol{\beta}}([\boldsymbol{u}_{kid}; \mathbf{c}_{ki}])\big) \in \Delta^{T-1},\\
\boldsymbol{u}_{kid} &\sim \mathcal{N}(\mathbf{0}, \alpha\mathbf{I}),\\
{}_{\mathbf{c}}ki &\sim \text{Cat}(L^{-1}\mathbf{1}_L).
\end{aligned}
$$

For every topic $t = 1, \ldots, T$, FUSION draws a gene–probability vector $\boldsymbol{\pi}_t$ from a Dirichlet prior, providing a shared dictionary of cell-type signatures. Each individual UMI $\boldsymbol{x}_{kid}$ at spot (k,i) is then generated in two steps: (i) a latent topic label $\boldsymbol{t}_{kid}$ is sampled from a categorical distribution whose logits are produced by a two-layer MLP $f_{\boldsymbol{\beta}}[\boldsymbol{u}_{kid}; \boldsymbol{c}_{ki}]$; (ii) depending on that label, the UMI is drawn from Multinomial $(1, \boldsymbol{\pi}_{t_{kid}})$. The MLP takes as input a standard Gaussian embedding, $\boldsymbol{u}_{kid} \sim \mathcal{N}(\mathbf{0}, \alpha\mathbf{I})$, that captures the UMI-specific variability and concatenates it with a one-hot spatial-domain indicator, $\boldsymbol{c}_{ki} \sim \text{Cat}(L^{-1}\mathbf{1}_L)$, shared across all slides, thereby linking topic usage to tissue architecture.

The inference algorithm for FUSION is designed to perform approximate Bayesian inference for models with non-conjugate priors and non-linear functions (via the neural network). We proposed the use of black-box variational inference (BBVI), a gradient-based optimization approach that approximates the posterior distribution. The variational distributions factorize over the latent variables as follows:

$$
\begin{aligned}
&q(\{\boldsymbol{t}_{kid}, \boldsymbol{u}_{kid}, \boldsymbol{c}_{ki}\}_{k,i,d}) = \prod_{k=1}^{K}\prod_{i=1}^{n_k} q(\boldsymbol{c}_{ki}) \prod_{p=1}^{P} q(\boldsymbol{t}_{kip}) q(\boldsymbol{u}_{kip}), \quad \text{with}\\
&q(\boldsymbol{t}_{kip}) = \text{Softmax}(\boldsymbol{\theta}_{kip}), \quad q(\boldsymbol{c}_{ki}) = \text{Softmax}(\text{Atten}\,(\boldsymbol{\rho}_{ki})),
\end{aligned}
$$

where $q(\boldsymbol{t}_{kip})$ is defined as the topic probability for the UMI of gene p in spot i in slice k and $q(\boldsymbol{c}_{ki})$ as the domain probability. To incorporate spatial information, FUSION uses the following attention mechanism to build the variational distribution for the spatial domain indicators $\boldsymbol{c}_{ki}$:

$$
\begin{aligned}
\text{Atten}\,(\boldsymbol{\rho}_{ki}) &= f_{\boldsymbol{\phi}}\left(\sum_{j \in \mathcal{N}(\boldsymbol{s}_{ki})} \boldsymbol{W}(\boldsymbol{s}_{ki}, \boldsymbol{s}_{kj}) \cdot \boldsymbol{\rho}_{kj}\right),\\
\boldsymbol{W}(\boldsymbol{s}_{ki}, \boldsymbol{s}_{kj}) &\propto \exp\left(-\gamma \cdot D(\boldsymbol{s}_{ki}, \boldsymbol{s}_{kj}) / \sqrt{L}\right)
\end{aligned}
$$

where $\mathcal{N}(\boldsymbol{s}_{ki})$ denotes neighbor subset of spot $\boldsymbol{s}_{ki}$, and $f_{\boldsymbol{\phi}}$ denotes two-layer MLP with parameters ϕ. The attention weight matrix $\boldsymbol{W}$ is computed based on the compound distance $D(\cdot,\cdot)$, based on the Euclidean distance between the spots $\boldsymbol{s}_{ki}$, $\boldsymbol{s}_{kj}$ and the cosine similarity of their posterior probabilities. The loss function of BBVI is based on the evidence lower bound (ELBO), which is computed as:

$$
\begin{aligned}
\mathcal{L}_{ELBO} \propto{}& \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{p=1}^{P} x_{kip} \sum_{t=1}^{T} \theta_{kipt} \cdot \mathbb{E}_{\boldsymbol{\pi}_t} \log\, \pi_{tp}\\
&- \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{l=1}^{L} \text{Atten}\,(\boldsymbol{\rho}_{ki})[l] \cdot \log\, \text{Atten}\,(\boldsymbol{\rho}_{ki})[l]\\
&- \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{p=1}^{P} x_{kip} \sum_{t=1}^{T} \theta_{kipt}\\
&\cdot \Bigg[\log\, \theta_{kipt} - \sum_{l=1}^{L} \text{Atten}\,(\boldsymbol{\rho}_{ki})[l]\\
&\cdot \mathbb{E}_{\boldsymbol{u}} \log\, \text{Softmax}(f_{\beta}([\boldsymbol{u}, c_{kil} = 1])[t])\Bigg],
\end{aligned}
$$

where $\boldsymbol{\theta}_{kip} = [\theta_{kip1}, \ldots, \theta_{kipT}]$ and $\boldsymbol{\rho}_{ki} = [\rho_{ki1}, \ldots, \rho_{kiL}]$ are trainable parameters. Formally, the optimal variational parameters and loss function of FUSION are defined as follows:

$$
\begin{aligned}
\{\boldsymbol{\theta}^*_{kip}, \boldsymbol{\rho}^*_{ki}\}_{k,i,p} = \arg\min_{\boldsymbol{\theta},\, \boldsymbol{\rho}}\ & \alpha \cdot \sum_{k=1}^{K} \|\boldsymbol{X}_k - \boldsymbol{P}_k \cdot \boldsymbol{B}^T\|_2\\
& - \beta \cdot \mathcal{L}_{ELBO} + \sum_{k=1}^{K} \text{Tr}\,(\mathbf{R}_k^T \cdot \mathbf{L}_k \cdot \mathbf{R}_k),
\end{aligned}
$$

where α and β are weight coefficients, and we have defined

$$
\begin{aligned}
&\mathbf{r}_{ki} = \sum_{p=1}^{P} x_{kip} \cdot [\theta_{kip1}, \ldots, \theta_{kipT}]^T, \quad \mathbf{p}_{ki} = \text{Softmax}\,(\mathbf{r}_{ki}),\\
&\mathbf{R}_k = [\mathbf{r}_{k1}, \ldots, \mathbf{r}_{kn_k}]^T, \quad \boldsymbol{P}_k = [\mathbf{p}_{k1}, \ldots, \mathbf{p}_{kn_k}]^T, \quad \mathbf{L}_k = \mathbf{D}_k - \mathbf{A}_k.
\end{aligned}
$$

FUSION tackles the computational bottlenecks of multi-section SRT by pairing sparse spatial attention with mini-gene batching. For the spot dimension, we consider an m-nearest neighbor graph, so the self-attention update for each domain logit $\boldsymbol{\rho}_{ki}$ involves only $m \ll \sum_k n_k$ neighbors. We choose $m = 30$ for our experiments. The cost of one update is therefore $O(m \cdot \sum_k n_k)$ rather than $O((\sum_k n_k)^2)$, and memory scales with the edge list rather than the full distance matrix. For the gene dimension, each BBVI iteration samples a small subset of $\tilde{p}$ genes and updates only their topic weights $\boldsymbol{\theta}_{kip}$. Because the ELBO factorizes over genes, such minibatching reduces both GPU memory and runtime from $P \times \sum_k n_k$ to $\tilde{p} \times \sum_k n_k$ while still providing unbiased gradient estimates. Together, these choices mean the increase in FUSION's memory footprint remains linear with batch size and enables single-GPU training on tens of sections at the whole-transcriptome resolution.

FUSION generates the spot-separable embeddings $\boldsymbol{r}_{ki}$ by weighting the gene expression counts $\boldsymbol{x}_{kip}$ with the topic probability parameters $\boldsymbol{\theta}_{kip}$. These embeddings are then normalized to obtain spot-separable cell-type proportions, $\boldsymbol{p}_{ki}$, at each spatial location. As these embeddings depend on the topic probabilities $\boldsymbol{\theta}_{\text{kip}}$, any change in the topic mixture immediately reshapes the embedding and the probability that the spot belongs to each spatial domain. During model training, we alternate the two optimization phases with Adam. In the topic training phase, we update $\boldsymbol{\theta}_{\text{kip}}$ by maximizing the ELBO (as in variational LDA), optionally adding an NMF regularizer that anchors the inferred topics probabilities to scRNA-seq reference profiles when such data are available. In the domain training phase, we refine the domain logits $\boldsymbol{\rho}_{ki}$ while holding $\boldsymbol{\theta}_{kip}$ fixed. A graph-Laplacian penalty $\text{Tr}\left(\mathbf{R}_k^\top \mathbf{L}_k \mathbf{R}_k\right)$ encourages neighboring spots with similar embeddings $\boldsymbol{r}_{ki}$ to share the same domain. The alternation stops when both the ELBO and the Laplacian term stabilize, typically within 50 epochs. Because domain probabilities are updated directly from the compositional embeddings, the resulting boundaries reflect shifts in the cell-type composition rather than gene-level variance, while the sparse Laplacian keeps domains spatially coherent. Full optimization details are described in the Supplementary Notes.

To remove section-specific batch effects while preserving biological signals, FUSION additionally fits a slice-conditioned Wasserstein GAN to harmonize compositional embeddings across sections. After the BBVI step we treat the spot embeddings $\mathbf{R}_k = \left[\mathbf{r}_{k1}, \ldots, \mathbf{r}_{kn_k}\right]^\top$ as empirical samples from a distribution $\mathbb{P}_k$. Embeddings from one reference section ($k = 1$) serve as the "real" distribution $\mathbb{P}_1$; for every other section ($k > 1$), we train a generator G_k (10-layer MLP with ReLU activation, shared weights across spots) that maps $\boldsymbol{r}_{ki}$ to $\hat{\mathbf{r}}_{ki} = G_k(\mathbf{r}_{ki})$. A single discriminator D tries to assign higher scores to $\mathbf{r} \sim \mathbb{P}_1$ than to $\hat{\mathbf{r}} \sim \mathbb{P}_k$, and the generator is updated to minimize the Wasserstein distance. Because the inputs are the cell-type compositional embeddings, alignment operates in a biologically interpretable space rather than on raw gene counts, and adversarial updates are computed on mini-batches of spots, adding negligible overhead to the BBVI. After 50 alternating steps the transformed embeddings $\hat{\mathbf{R}}_k$ are statistically indistinguishable from $\boldsymbol{R}_1$, yielding a batch-corrected, cross-slice latent space that feeds directly into all downstream analyses.

4.2 | Simulation Details

The simulated data were generated from two different biological sources to create two distinct multi-slice integration scenarios with known ground truth.

In scenario 1, we generated a multi-slice dataset from scRNA-seq data. We utilized the *Mouse musculus* single-cell atlas data [54] containing 2 837 cells and 35 100 genes. From this dataset, we randomly selected five major cell types with sufficient representation: endothelial cell (231), late pro-B cell (177), pro-B cell (171), myeloid cell (103), and neuron (61). We designed 10 different spatial topics (patterns of cell type composition) and assigned each topic a distinct mixture of the five cell types. For each topic, we generated a varying number of spots (ranging from 50 to 160) by sampling cells from the single-cell data according to the topic's cell type mixture. This procedure created spatially organized domains with known cellular composition gradients. To simulate multiple slices with varying technical quality, we generated three synthetic slices by applying different levels of multiplicative noise to gene expression counts: low (5% standard deviation, 10% probability), medium (10% standard deviation, 20% probability), and high (20% standard deviation, 30% probability). This resulted in a multi-slice dataset comprising 2 700 spots in total (900 spots per slice) with preserved biological structure but varying technical noise levels.

In scenario 2, the simulated data were generated from a real ST dataset of a mouse embryo at stage E16.5 [9]. The original dataset contained 121 767 spots and 23 761 genes. To construct biologically meaningful scenarios with known anatomical annotations, a total of 42 909 spots were isolated from six major organs, including the brain, choroid plexus, heart, kidney, liver, and lung. Using stratified sampling with organs as strata, we selected a total of 5 000 spots, distributed as follows: brain (2 025), choroid plexus (201), heart (434), kidney (251), liver (1 651), and lung (438). To simulate varying degrees of batch effects, we applied three different noise levels by applying three levels of scaling factors to the count matrix: low (0.8–1.2), medium (0.6–1.4), and high (0.4–1.6). These ranges control the variability between simulated slices and their divergence from the original sample. Higher noise levels result in reduced inter-slice correlation and greater deviation from the original data. We generated five synthetic slices by applying multiplicative noise and random drift to the gene expression count matrix, along with Gaussian noise to the spatial coordinates. Because all five slices were derived from the same set of 5 000 spots, inter-slice correlation was naturally preserved. This resulted in a multi-slice dataset comprising 25 000 spots in total.

With FUSION, each of the eight state-of-the-art spatial data-integration methods was executed independently, with 20 replicates per condition to ensure statistical robustness. Performance was evaluated using the ARI to measure clustering accuracy when compared to the known cellular compositions (scenario 1) or anatomical annotations (scenario 2). Computational runtime was also recorded to assess scalability and efficiency.

4.3 | Downstream Analysis

4.3.1 | Differentially Expressed Gene Analysis

For every simulated and real multi-section ST dataset, we first removed low-quality spots (total UMI count < 200) and genes

detected in fewer than 10 spots using `sc.pp.filter_cells` and `sc.pp.filter_genes` using `Python` package *Scanpy*. Quality-controlled data were library-size–normalized with `sc.pp.normal ize_total` and log-transformed via `sc.pp.log1p`. We then performed domain–specific differential expression based on the detected domain labels. Specifically, we called `sc.tl.rank_genes_gr oups` (method = "wilcoxon") for each annotated domain, treating all remaining spots as the background population. Using *Scanpy* to report Benjamini–Hochberg–adjusted *p*-values and log 2 fold changes for every gene; we retained the DE genes whose adjusted *p*-values were < 0.001.

4.3.2 | Gene-Set Enrichment Analysis of Domain-Specific DE Genes

To elucidate the biological programs underlying each spatial domain, we subjected the domain-specific DE genes to GSEA using the `Python` package *gseapy*. For every domain on each slide, we first ranked the DE genes according to their assigned $\log_{10}$ Benjamini–Hochberg–adjusted *p*-values. The resulting ranked list was analyzed by `gseapy.prerank` (permutation_num = 1, 000; min_size = 15; max_size = 500) using the KEGG pathway collections with gene_sets = ["KEGG_2016","KEGG_2013"]. Enriched gene sets with a false-discovery rate of <0.05 were considered significant. For each domain, we retained the top 10 pathways by NES and visualized them with `gseapy.barplot`, enabling a direct comparison of functional themes across adjacent domains.

4.3.3 | Domain Annotation Based on Context-Specific DE Genes

To annotate each predicted domain, for example, tumor or non-tumor, we first derived two reference signatures by performing DE analysis between the pathologist-defined tumor and non-tumor regions and retained the top 300 up-regulated genes for each class (denoted $\mathcal{G}^{\text{tum}}$ and $\mathcal{G}^{\text{n.t.}}$). For every predicted domain, d, we carried out an analogous DE comparison of the spots in d against all remaining spots to obtain its domain-specific gene set, $\mathcal{G}_d$ (top $M = 50$ genes). We then quantified the similarity between d and each reference class, $c \in \{\text{tum}, \text{n.t.}\}$, via the normalized overlap, $S_{d,c} = |\mathcal{G}_d \cap \mathcal{G}^c|/M \in [0, 1]$, and assigned the label $\hat{c}_d = \arg\max_c S_{d,c}$. A high correspondence between these automatically generated labels and the ground-truth tumor mask provided downstream validation of the spatial clustering results.

4.4 | Methods for Comparison

We compared FUSION with eight spatial-domain detection methods: (1) BayesSpace implemented in the `R` package BayesSpace (version 1.14.0), using the `spatialCluster()` function with q equal to domain number and $nrep = 10\,000$ iterations from https://bioconductor.org/packages/BayesSpace; (2) IRIS and its reference-free variant IRISfree implemented in the `R` package IRIS, invoked with `IRIS()` or `IRISfree()` at default settings from https://github.com/YingMa0107/IRIS; (3) BASS implemented in the `R` package BASS (version 1.1.0), executed via `BASS()` with 5 000 burn-in and 10 000 MCMC iterations from https://github.com/zhengli09/BASS; (4) CARD implemented in the `R` package CARD (version 1.2.0), estimating spot-level cell-type proportions through `CARD_deconvolution()` and deriving domain labels by *k*-means clustering as described in the original paper (https://yma-lab.github.io/CARD/); (5) SpaGCN implemented in the `Python` package SpaGCN (version 1.2.7), run with the high-level `SpaGCN.clustering()` workflow in the project tutorial (https://github.com/jianhuupenn/SpaGCN); (6) SEDR implemented in the `Python` package SEDR (version 1.0), trained for 300 epochs using `SEDR.train_model()` and clustered in latent space with Leiden (resolution 0.8) from https://github.com/JinmiaoChenLab/SEDR; (7) STAGATE implemented in the `Python` package STAGATE, producing graph-attention embeddings via `STAGATE.train()` (3 000 epochs) followed by Leiden clustering (resolution 1.0) from https://github.com/zhanglabtools/STAGATE; and (8) STitch3D implemented in the `Python` package STitch3D (version 1.0.3), aligning slides with `STitch3D.utils.align_spots()`, fitting a 3-D auto-encoder with `model.train()`, and clustering the resulting coordinates at Leiden resolution 0.6 (https://github.com/YangLabHKUST/STitch3D).

Briefly, BayesSpace imposes a Markov-random-field prior on a finite mixture of multivariate t distributions, encouraging neighboring spots to share a label. IRIS leverages cell-type compositions learned from a scRNA-seq reference and jointly segments multiple slices; IRISfree removes this dependency via an unsupervised optimization step. BASS fits a Bayesian hierarchical model that simultaneously clusters single cells and aggregates them into higher-level tissue domains, borrowing strength across sections. CARD is a spatially informed autoregressive deconvolution method that returns spot-level cell-type proportions; by clustering these proportions, we obtain coarse tissue compartments. SpaGCN builds a weighted graph that fuses gene expression, spatial adjacency, and histology features, then applies a graph convolutional network followed by deep embedded clustering. SEDR couples a deep autoencoder with a variational graph autoencoder to learn a spatially aware latent space, whereas STAGATE replaces the fixed graph weights with an adaptive graph-attention mechanism, markedly sharpening boundaries. Finally, STitch3D extends autoencoder–based embedding to the 3D setting, jointly modeling aligned slices and a scRNA-seq reference to delineate coherent spatial regions across depth.

4.5 | Evaluation Metrics

We comprehensively benchmarked FUSION on real multi-section SRT datasets. For the evaluation of cell-type deconvolution, we quantified the agreement of the estimated cell type composition matrices using the average column-wise Pearson correlation (Avg Corr). To assess the identified spatial domain, we compared the inferred domain labels with the reference anatomic annotations via the ARI. Finally, to evaluate the effectiveness of cross-section batch correction, we computed the integration Local Inverse Simpson's Index (iLISI), which measures how well spots from different tissue sections are mixed in the learned embedding. Higher Avg Corr, ARI, and iLISI values indicate better performance in their respective tasks of deconvolution accuracy, domain delineation, and batch-effect removal.

ARI ARI measures how often the predicted domain labels agree with the ground-truth labels on *pairs* of spots, while correcting

for agreements that could occur by chance. Denote the unadjusted Rand index (the raw proportion of agreeing pairs) by

$$\mathrm{RI} = \frac{(\#\text{ spot pairs on which the two labelings agree})}{\binom{N}{2}},$$

Because RI is biased upward when many clusters are present, ARI subtracts its expected value under random labelings and rescales to the theoretical maximum:

$$\mathrm{ARI} = \frac{\mathrm{RI} - \mathrm{expected(RI)}}{\max(\mathrm{RI}) - \mathrm{expected(RI)}}$$

Hence ARI = 1 indicates perfect concordance, 0 corresponds to random agreement, and negative values signal worse-than-random alignment.

iLISI To quantify cross-section mixing in a learned embedding, we compute a LISI score per spot using section labels. For each spot i, let $\mathcal{N}_k(i)$ be its k-nearest neighbors and let $b \in \{1, \ldots, B\}$ index sections. The local section distribution is

$$p_{i,b} = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} \mathbf{1}\{\text{section } (j) = b\}, \quad \sum_{b=1}^{B} p_{i,b} = 1$$

The spot-level and global scores are

$$\mathrm{LISI}_i = \left(\sum_{b=1}^{B} p_{i,b}^2 \right)^{-1}, \quad \mathrm{iLISI} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{LISI}_i.$$

We use $k = 30$. Higher iLISI (approaching B) indicates better cross-section mixing and lower values reflect residual batch structure.

Conservation Local Inverse Simpson's Index (cLISI) To assess preservation of biological structure (e.g., domains or cell types), we apply the same LISI definition using biological labels instead of sections. Let $y \in \{1, \ldots, C\}$ index biological classes and define

$$q_{i,y} = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} \mathbf{1}\{\mathrm{label}(j) = y\}, \quad \sum_{y=1}^{C} q_{i,y} = 1$$

with

$$\mathrm{cLISI}_i = \left(\sum_{y=1}^{C} q_{i,y}^2 \right)^{-1}, \quad \mathrm{cLISI} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{cLISI}_i$$

Lower cLISI (approaching 1) indicates that neighbors share the same biological label (good structure conservation) and higher cLISI reflects over-mixing that blurs biological boundaries.

Pearson Correlation across Shared Domains To assess the cell-type proportions for homologous (shared) domains across different tissue slices of multi-section SRT data, we proposed the following correlation metric. Let B be the number of slices, K the number of cell types, and $\mathcal{D}$ the set of domains that are identified on *at least two* slices. For domain $d \in \mathcal{D}$ and slice $b \in \{1, \ldots, B\}$, denoted by $S_d^{(b)} \subseteq \{1, \ldots, N_b\}$ the index set of spots on slice b that are assigned to domain d, and let $\hat{\mathbf{c}}_i \in \mathbb{R}^K$ be the estimated composition vector of spot i. The slice-averaged composition for domain d on slice b is

$$\hat{\boldsymbol{\mu}}_d^{(b)} = \frac{1}{|S_d^{(b)}|} \sum_{i \in S_d^{(b)}} \hat{\mathbf{c}}_i.$$

For every domain d that appears on a pair of slices (b_1, b_2), we compute the Pearson correlation between their slice-level means:

$$\mathrm{Corr}_d^{(b_1, b_2)} = \frac{(\hat{\boldsymbol{\mu}}_d^{(b_1)} - \overline{\mu}_d \mathbf{1})^\top (\hat{\boldsymbol{\mu}}_d^{(b_2)} - \overline{\mu}_d \mathbf{1})}{\|\hat{\boldsymbol{\mu}}_d^{(b_1)} - \overline{\mu}_d \mathbf{1}\|_2 \, \|\hat{\boldsymbol{\mu}}_d^{(b_2)} - \overline{\mu}_d \mathbf{1}\|_2},$$
$$\overline{\mu}_d = \tfrac{1}{2} \left(\mathbf{1}^\top \hat{\boldsymbol{\mu}}_d^{(b_1)} + \mathbf{1}^\top \hat{\boldsymbol{\mu}}_d^{(b_2)} \right) \Big/ K.$$

When a domain occurs on more than two slices, we evaluate all unordered slice pairs and average their correlations. The overall agreement of cell-type compositions across shared domains is summarized by

$$\mathrm{AvgCorr} = \frac{1}{\sum_{d \in \mathcal{D}} \binom{m_d}{2}} \sum_{d \in \mathcal{D}} \sum_{(b_1, b_2) \in \binom{\mathcal{B}_d}{2}} \mathrm{Corr}_d^{(b_1, b_2)},$$

where $m_d = |\mathcal{B}_d|$ is the number of slices on which domain d appears and $\binom{\mathcal{B}_d}{2}$ denotes all $\binom{m_d}{2}$ unordered slice pairs. A higher AvgCorr indicates more consistent cell-type compositions for homologous domains across tissue sections.

4.6 | Real Datasets

4.6.1 | Human DLPFC Data by 10x Visium Technology

We analyzed the publicly available human DLPFC 10x Visium dataset [27] using FUSION and compared its performance to that of competing methods. We profiled 12 contiguous 10-μm tissue sections (two pairs per block from three neurotypical adult donors, with the second pair collected 300-μm posterior to the first), resulting in a total of 45 044 bar-coded spots and expression counts for 33 538 genes. Each section is accompanied by manual histological annotations that delineate the six cortical layers (L1–L6) and the underlying white matter, which we adopted as ground-truth domains for benchmarking the spatial clustering. For cell-type deconvolution we used the 10x Chromium scRNA-seq dataset from the post-mortem brain [36], comprising 70 634 cells and 17 926 genes for 44 transcriptional subtypes from 7 major cell classes: astrocyte (Ast, $n = 3\,392$), endothelial cell (End, $n = 121$), excitatory neuron (Ex, $n = 34\,976$), inhibitory neuron (In, $n = 9\,196$), microglia (Mic, $n = 1\,920$), oligodendrocyte (Oli, $n = 18\,235$), oligodendrocyte precursor cell (Opc, $n = 2\,627$), and pericyte (Per, $n = 167$). Throughout our experiments, we fixed the number of spatial domains to $K = 7$ for all competing methods and set the loss-weight coefficients to $\alpha = 10$ and $\beta = 1$ for FUSION.

4.6.2 | Human SCC Data from ST

We analyzed the publicly available human cutaneous SCC dataset generated with the first-generation ST platform [40] using FUSION and currently used methods. Four lesions (patients P2, P5, P9, and P10) were profiled at 100-μm resolution, yielding 8 671 barcoded spots with a median of 15 168 detected genes.

Hematoxylin-eosin staining followed by dermatopathologist review produced five histological compartments: (i) tumor core, (ii) invasive leading edge, (iii) cancer-associated stroma, (iv) inflamed immune foci, and (v) adjacent normal epidermis/dermis. To benchmark the spatial clustering, we collapsed these annotations into a binary mask (*tumor* vs. *non-tumor*). The top 1 000 highly variable genes per lesion were selected with the R package `Seurat`. Predicted domains were subsequently annotated as tumor or non-tumor by computing the overlap score between their domain-specific DE genes (top 50 genes) and reference tumor/non-tumor DE gene signatures (top 300 genes per class). For deconvolution, we used the matched scRNA-seq atlas from the same study, comprising 48, 164 cells and 32 738 genes. Fine-grained clustering identified 24 transcriptional subtypes that we aggregated into seven macro cell classes: malignant keratinocytes, non-malignant keratinocytes, immune (lymphoid + myeloid), fibroblasts, endothelial cells, melanocytes, and appendage/glandular epithelium. We fixed the number of spatial domains at $K = 12$; for FUSION we adjusted the loss-weight coefficients to $\alpha = 20$ and $\beta = 1$.

4.6.3 | Mouse Olfactory Bulb Data from Stereo-Seq and Slide-Seq v2 Technologies

We analyzed the publicly released MOB Slide-seq v2 datasets [6] using FUSION and other methods. The coronal section was arrayed on barcoded bead surfaces, yielding a total of 20 193 spots with expression counts for 21 220 genes. We also analyzed the MOB Stereo-seq dataset released with the original Stereo-seq paper [9]. After quality control, sagittal Section 1 yielded 27 106 spots with expression counts for 19 109 genes. Each section is accompanied by a histology-based mask that delineates the canonical MOB laminae—ONL, GL, EPL, MCL, IPL, GCL, and RMS—which we treat as ground-truth domains for clustering benchmarks. For cell-type deconvolution, we obtained the scRNA-seq data from 10× Chromium [46], comprising 21 746 cells, 18 560 genes, and 18 transcriptional subtypes spanning excitatory neurons (mitral/tufted, external tufted, superficial tufted), inhibitory interneurons (granule cells, periglomerular subclasses), astrocytes, oligodendrocytes, endothelial cells, and microglia. In all experiments, we fixed the number of spatial domains to $K = 8$ to match the anatomical layers for cross-platform analysis and set the FUSION loss-weight coefficients to $\alpha = 10$ and $\beta = 1$.

4.6.4 | Mouse Embryo Data from Stereo-Seq and Visium Technologies

We analyzed the publicly available mouse embryonic development dataset generated with Stereo-seq technology [9] using FUSION and other methods. Eight embryonic stages (E9.5 to E16.5) were profiled at bin50 (50 μm) and segmented at a single-cell resolution, yielding 5 910 to 121 760 analyzable spots from E9.5 to E16.5 (representing a >20-fold increase). Anatomical structures (e.g., neural tube, heart, somites, limb buds, liver) were annotated using in situ gene expression patterns and morphological landmarks. For comparison, we also analyzed the FFPE Mouse Embryo tissue blocks from the Visium platform, which contain 6 434 spots profiled with an 11-mm capture area. For benchmarking the spatial clustering, we used these annotations as ground truth domains. The top 1,000 highly variable genes per stage were selected with the R package `Seurat`. For deconvolution, we employed the Tabula Muris single-cell transcriptomic atlas of adult mouse organs [61], comprising 100,605 cells from 20 organs. We aggregated the 117 annotated cell types into 18 broad classes including: epithelial cells (lung, liver, kidney), endothelial cells, immune cells (T cells, B cells, macrophages), mesenchymal cells (fibroblasts, adipocytes), muscle cells (cardiac, skeletal), neuronal cells, and specialized organ cell types (hepatocytes, pancreatic acinar cells). In all experiments, we fixed the number of spatial domains to $K = 8$ to match the anatomical layers for multi-section analysis and set the FUSION loss-weight coefficients to $\alpha = 1$ and $\beta = 1$.

Acknowledgments

This work was supported by Shenzhen Medical Research Fund (D250402003), Basic Research Fund in Shenzhen Natural Science Foundation (grant no. JCYJ20240813104924033), National Natural Science Foundation of China (12401361), the Fundamental Research Funds for the Central Universities (1082204112J06), and Natural Science Foundation of Sichuan Province (2025ZNSFSC0809).

Funding

This study was supported by Shenzhen Medical Research Fund (D250402003), Basic Research Fund in Shenzhen Natural Science Foundation (JCYJ20240813104924033), National Natural Science Foundation of China (12401361), Fundamental Research Funds for the Central Universities (1082204112J06), and Natural Science Foundation of Sichuan Province (2025ZNSFSC0809).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All datasets used in this study are publicly available. These include human DLPFC data by 10x Visium available at (`http://spatial.libd.org/spatialLIBD/`), with Chromium scRNA-seq dataset from the post-mortem brain at (`https://www.synapse.org/#!Synapse:syn18485175`); human SCC ST data is available at (`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144240`), with matched scRNA-seq atlas at (`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144236`); The mouse olfactory bulb (MOB) Stereo-seq and Slide-seqV2 dataset are available at (`https://github.com/JinmiaoChenLab/SEDR_analyses`) and (`https://singlecell.broadinstitute.org/single_cell/study/SCP815`), respectively, with the mouse olfactory bulb scRNA-seq data available at GEO accession (`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121891`); mouse embryo stereo-seq data for eight stages are available at (`https://db.cngb.org/stomics/mosta/download/`), with matched scRNA-seq reference data at (`https://www.ebi.ac.uk/gxa/sc/experiments/E-ENAD-15/downloads`); mouse embryo Visium data is available at (`https://www.10xgenomics.com/datasets/visium-cytassist-mouse-embryo-11-mm-capture-area-ffpe-2-standard`).

Code Availability

The FUSION package and source code were publicly available at https://github.com/SpatialOmics-pj/FUSION.git. Codes and datasets for reproducing the analysis can be found on the same website.

References

1. V. Marx, "Method of the Year 2020: Spatially Resolved Transcriptomics,"*Nature Methods* 18 (2021): 9–14, https://www.nature.com/articles/s41592-020-01042-x.

2. A. Raj, P. Van Den Bogaard, S. A. Rifkin, A. Van Oudenaarden, and S. Tyagi, "Imaging Individual Mrna Molecules Using Multiple Singly Labeled Probes," *Nature Methods* 5 (2008): 877–879.

3. J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, et al., "Molecular, Spatial, and Functional Single-Cell Profiling of the Hypothalamic Preoptic Region," *Science* 362 (2018): eaau5324, https://api.semanticscholar.org/CorpusID:53566396.

4. T. Lohoff, S. Ghazanfar, A. Missarova, et al., "Integration of Spatial and Single-Cell Transcriptomic Data Elucidates Mouse Organogenesis," *Nature Biotechnology* 40 (2021): 74–85, https://api.semanticscholar.org/CorpusID:237434255.

5. 10x genomics. visium spatial gene expression, 2019, https://www.10xgenomics.com/products/spatial-gene-expression.

6. S. G. Rodriques, R. R. Stickels, A. Goeva, et al., "Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution," *Science* 363 (2019): 1463–1467.

7. R. R. Stickels, E. Murray, P. Kumar, et al., "Highly Sensitive Spatial Transcriptomics at Near-Cellular Resolution with Slide-Seqv2," *Nature Biotechnology* 39 (2021): 313–319.

8. Y. Liu, M. Yang, Y. Deng, et al., "High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue," *Cell* 183 (2020): 1665–1681.

9. A. Chen, S. Liao, M. Cheng, et al., "Spatiotemporal Transcriptomic Atlas of Mouse Organogenesis Using Dna Nanoball-Patterned Arrays," *Cell* 185 (2022): 1777–1792.

10. K. Clifton, M. Anant, G. Aihara, et al., "Stalign: Alignment of Spatial Transcriptomics Data Using Diffeomorphic Metric Mapping," *Nature Communications* 14 (2023): 8123.

11. X. Liu, R. Zeira, and B. J. Raphael, "*Paste2: Partial Alignment of Multi-Slice Spatially Resolved Transcriptomics Data*," bioRxiv (2023).

12. D. M. Cable, E. Murray, L. S. Zou, et al., "Robust Decomposition of Cell Type Mixtures in Spatial Transcriptomics," *Nature Biotechnology* 40 (2022): 517–526.

13. G. Wang, J. Zhao, Y. Yan, Y. Wang, A. R. Wu, and C. Yang, "Construction of a 3d Whole Organism Spatial Atlas by Joint Modelling of Multiple Slices with Deep Neural Networks," *Nature Machine Intelligence* 5 (2023): 1200–1213.

14. R. Dong and G.-C. Yuan, "Spatialdwls: Accurate Deconvolution of Spatial Transcriptomic Data," *Genome Biology* 22 (2021): 145.

15. M. Elosua-Bayes, P. Nieto, E. Mereu, I. Gut, and H. Heyn, "Spotlight: Seeded Nmf Regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes," *Nucleic Acids Research* 49 (2021): e50–e50.

16. Y. Ma and X. Zhou, "Spatially Informed Cell-Type Deconvolution for Spatial Transcriptomics," *Nature Biotechnology* 40 (2022): 1349–1359.

17. V. Kleshchevnikov, A. Shmatko, E. Dann, et al., "Cell2Location Maps Fine-Grained Cell Types in Spatial Transcriptomics," *Nature Biotechnology* 40 (2022): 661–671.

18. R. Lopez, B. Li, H. Keren-Shaul, et al., "Destvi Identifies Continuums of Cell Types in Spatial Transcriptomics Data," *Nature Biotechnology* 40 (2022): 1360–1369.

19. Q. Dong, Y. Yang, Z. Luo, H. Shen, X. Shi, and J. Liu, "Robust Spatial Cell-Type Deconvolution with Qualitative Reference for Spatial Transcriptomics," *Small Methods* 9 (2025): 2401145.

20. Y. Long, K. S. Ang, M. Li, et al., "Spatially Informed Clustering, Integration, and Deconvolution of Spatial Transcriptomics with Graphst," *Nature Communications* 14 (2023): 1155.

21. E. Zhao, M. R. Stone, X. Ren, et al., "Spatial Transcriptomics at Subspot Resolution with Bayesspace," *Nature Biotechnology* 39 (2021): 1375–1384.

22. Z. Li and X. Zhou, "Bass: Multi-Scale and Multi-Sample Analysis Enables Accurate Cell Type Clustering and Spatial Domain Detection in Spatial Transcriptomic Studies," *Genome Biology* 23, no. 1 (2022): 168.

23. J. Hu, X. Li, K. Coleman, et al., "Spagcn: Integrating Gene Expression, Spatial Location and Histology to Identify Spatial Domains and Spatially Variable Genes by Graph Convolutional Network," *Nature Methods* 18 (2021): 1342–1351.

24. H. Xu, H. Fu, Y. Long, et al., "Unsupervised Spatially Embedded Deep Representation of Spatial Transcriptomics," *Genome Medicine* 16 (2024): 12.

25. K. Dong and S. Zhang, "Deciphering Spatial Domains from Spatially Resolved Transcriptomics with an Adaptive Graph Attention Auto-Encoder," *Nature Communications* 13 (2022): 1739.

26. I. Korsunsky, N. Millard, J. Fan, et al., "Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony," *Nature Methods* 16 (2019): 1289–1296.

27. K. R. Maynard, L. Collado-Torres, L. M. Weber, et al., "Transcriptome-Scale Spatial Gene Expression in the Human Dorsolateral Prefrontal Cortex," *Nature Neuroscience* 24 (2021): 425–436.

28. Z. Li, T. Wang, P. Liu, and Y. Huang, "Spatialdm for Rapid Identification of Spatially Co-Expressed Ligand–receptor and Revealing Cell–cell Communication Patterns," *Nature Communications* 14 (2023): 3995.

29. V. Singhal, N. Chou, J. Lee, et al., "Banksy Unifies Cell Typing and Tissue Domain Segmentation for Scalable Spatial Omics Data Analysis," *Nature Genetics* 56 (2024): 431–441.

30. B. Duan, S. Chen, X. Cheng, and Q. Liu, "Multi-Slice Spatial Transcriptome Domain Analysis with Spado," *Genome Biology* 25 (2024): 73.

31. R. Ranganath, S. Gerrish, and D. Blei, "Black Box Variational Inference," *Artificial Intelligence and Statistics* (PMLR, 2014), 814–822.

32. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, (PMLR, 2017), 214–223.

33. A. Bove, D. Gradeci, Y. Fujita, S. Banerjee, G. Charras, and A. R. Lowe, "Local Cellular Neighborhood Controls Proliferation in Cell Competition," *Molecular Biology of the Cell* 28 (2017): 3215–3228.

34. C. R. Stoltzfus, J. Filipek, B. H. Gern, et al., "Cytomap: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues," *Cell Reports* 31 (2019): 107523–107523, https://api.semanticscholar.org/CorpusID:203400366.

35. Y. Ma and X. Zhou, "Accurate and Efficient Integrative Reference-Informed Spatial Domain Detection for Spatial Transcriptomics," *Nature Methods* 21 (2024): 1231–1244.

36. H. Mathys, J. Davila-Velderrain, Z. Peng, et al., "Single-Cell Transcriptomic Analysis of Alzheimer's Disease," *Nature* 570 (2019): 332–337.

37. V. Gallo and B. Deneen, "Glial Development: The Crossroads of Regeneration and Repair in the Cns," *Neuron* 83 (2014): 283–308.

38. O. A. Bayraktar, T. Bartels, S. Holmqvist, et al., "Astrocyte Layers in the Mammalian Cerebral Cortex Revealed by a Single-Cell In Situ Transcriptomic Map," *Nature Neuroscience* 23 (2020): 500–509.

39. P. L. Ståhl, F. Salmén, S. Vickovic, et al., "Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics," *Science* 353 (2016): 78–82.

40. A. L. Ji, A. J. Rubin, K. Thrane, et al., "*Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma*,"Cell182 (2020): 497–514.

41. Z. Fang, Q. Meng, J. Xu, et al., "Signaling Pathways in Cancer-Associated Fibroblasts: Recent Advances and Future Perspectives," *Cancer Communications* 43 (2023): 3–41.

42. A. Schneider, R. H. Younis, and J. S. Gutkind, "Hypoxia-Induced Energy Stress Inhibits the Mtor Pathway by Activating an Ampk/redd1 Signaling Axis in Head and Neck Squamous Cell Carcinoma," *Neoplasia* 10 (2008): 1295–1302.

43. D. F. Quail and J. A. Joyce, "Microenvironmental Regulation of Tumor Progression and Metastasis," *Nature Medicine* 19 (2013): 1423–1437.

44. L. Keren, M. Bosse, D. Marquez, et al., "A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging," *Cell* 174 (2018): 1373–1387.

45. S. Nagayama, R. Homma, and F. Imamura, "Neuronal Organization of Olfactory Bulb Circuits," *Frontiers in Neural Circuits* 8 (2014): 98.

46. B. Tepe, M. C. Hill, B. T. Pekarek, et al., "Single-Cell Rna-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons," *Cell Reports* 25 (2018): 2689–2703.

47. K. Kosaka and T. Kosaka, "Synaptic Organization of the Glomerulus in the Main Olfactory Bulb: Compartments of the Glomerulus and Heterogeneity of the Periglomerular Cell," *Anatomical Science International* 80 (2005): 80–90.

48. J. M. Boggs, "Myelin Basic Protein: A Multifunctional Protein," *Cellular and Molecular Life Sciences CMLS* 63 (2006): 1945–1961.

49. B. E. Maley, "Immunohistochemical Localization of Neuropeptides and Neurotransmitters in the Nucleus Solitarius," *Chemical Senses* 21 (1996): 367–376.

50. K. Hori, K. Yamashiro, T. Nagai, et al., "Auts2 Regulation of Synapses for Proper Synaptic Inputs and Social Communication," *IScience* 23 (2020): 101183.

51. M. Pajares, I. Rojo, A. Manda, G. Boscá, L, and A. Cuadrado, "Inflammation in Parkinson's Disease: Mechanisms and Therapeutic Implications," *Cells* 9 (2020): 1687.

52. N. Nagan and R. A. Zoeller, "Plasmalogens: Biosynthesis and Functions," *Progress in Lipid Research* 40 (2001): 199–229.

53. F. W. Pfrieger and N. Ungerer, "Cholesterol Metabolism in Neurons and Astrocytes," *Progress in Lipid Research* 50 (2011): 357–371.

54. T. M. Consortium, "Coordination Schaum Nicholas 1 Karkanias Jim 2 Neff Norma F. 2 May Andrew P. 2 Quake Stephen R. quake@ stanford. edu 2 3 f Wyss-Coray Tony twc@ stanford. edu 4 5 6 g Darmanis Spyros spyros. darmanis@ czbiohub. org 2 h, O., coordination Batson Joshua 2 Botvinnik Olga 2 Chen Michelle B. 3 Chen Steven 2 Green Foad 2 Jones Robert C. 3 Maynard Ashley 2 Penland Lolita 2 Pisco Angela Oliveira 2 Sit Rene V. 2 Stanley Geoffrey M. 3 Webber James T. 2 Zanini Fabio 3, L. & data analysis Batson Joshua 2 Botvinnik Olga 2 Castro Paola 2 Croote Derek 3 Darmanis Spyros 2 DeRisi Joseph L. 2 27 Karkanias Jim 2 Pisco Angela Oliveira 2 Stanley Geoffrey M. 3 Webber James T. 2 Zanini Fabio 3, C. Single-cell transcriptomics of 20 mouse organs creates a tabula muris," *Nature* 562 (2018): 367–372,

55. K. Uemura, M. Saka, T. Nakagawa, et al., "L-mbp Is Expressed in Epithelial Cells of Mouse Small Intestine," *The Journal of Immunology* 169 (2002): 6945–6950.

56. K. Tiklová, A. K. Björklund, L. Lahti, et al., "Single-Cell Rna Sequencing Reveals Midbrain Dopamine Neuron Diversity Emerging during Mouse Brain Development," *Nature Communications* 10 (2019): 581.

57. K. Dasgupta and J. Jeong, "Developmental Biology of the Meninges," *Genesis* 57 (2019): e23288.

58. D. M. DeLaughter, A. G. Bick, H. Wakimoto, et al., "Single-Cell Resolution of Temporal Gene Expression during Heart Development," *Developmental Cell* 39 (2016): 480–490.

59. Y. Wang, D. Chen, D. Xu, et al., "Early Developing b Cells Undergo Negative Selection by Central Nervous System-Specific Antigens in the Meninges," *Immunity* 54 (2021): 2784–2794.

60. M. Nahrendorf, "Myeloid Cells in Cardiovascular Organs," *Journal of Internal Medicine* 285 (2019): 491–502.

61. N. Schaum, J. Karkanias, N. F. Neff, et al., "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris: The Tabula Muris Consortium," *Nature* 562 (2018): 367.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Fig. S1:** The spatial domain maps generated from the joint analysis of all 12 integrated slices for DLPFC tissue for 151507-151510 using all candidate methods. IRIS free and FUSION free are alternative methods of IRIS and FUSION, respectively, that do not use scRNA-seq data from 10x Chromium analysis of the post-mortem brain tissue as reference. **Supporting Fig. S2:** The spatial domain maps generated from the joint analysis of all 12 integrated slices for DLPFC tissue for 151669-151672 using all candidate methods. IRIS free and FUSION free are alternative methods of IRIS and FUSION, respectively, that do not use scRNA-seq data from 10x Chromium analysis of the post-mortem brain tissue as reference. **Supporting Fig. S3:** The spatial domain maps generated from the joint analysis of all 12 integrated slices for DLPFC tissue for 151673-151676 using all candidate methods. IRIS free and FUSION free are alternative methods of IRIS and FUSION, respectively, that do not use scRNA-seq data from 10x Chromium analysis of the post-mortem brain tissue as reference. **Supporting Fig. S4:** The UMAP embeddings across the non-contiguous slices 151507, 151669, and 151673. The RGB plots show spots from the three slices, while the 8-color embedding plots indicate section identity, where the colors illustrate the annotation labels and the coordinates are the values of the embeddings from FUSION, FUSION without WGAN, BayesSpace, CARD, and IRIS. **Supporting Fig. S5:** The UMAP embeddings across the non-contiguous slices 151507, 151669, and 151673. The RGB plots showed spots from the three slices, while the 8-color embedding plots indicates the section identity where the colors illustrate the annotation labels and the coordinates are the value of the embeddings from SpaGCN, SEDR, and STAGATE. **Supporting Fig. S6:** The UMAP embeddings across the non-contiguous slices 151507, 151669, and 151673. The RGB plots showed spots from the three slices, while the multi-color embedding plots indicates the topic domains and the corresponding coordinates inferred from FUSION, FUSION without WGAN, BayesSpace, CARD, and IRIS. **Supporting Fig. S7:** The UMAP embeddings across the non-contiguous slices 151507, 151669, and 151673. The RGB plots showed spots from the three slices, while the multi-color embedding plots indicates the topic domains and the corresponding coordinates inferred from SpaGCN, SEDR, and STAGATE. **Supporting Fig. S8:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using FUSION, FUSION free, BayesSpace, IRIS, BASS, and CARD. The plots express the results of the three tissue slices from Patient 2. FUSION free are the alternative method of FUSION, but not using scRNA-seq data from 10x Chromium on the post-mortem brain tissue as reference. **Supporting Fig. S9:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using SpaGCN, SEDR, STAGATE, and Stitch3D. The plots express the results of the three tissue slices from Patient 2. **Supporting Fig. S10:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using FUSION, FUSION free, BayesSpace, IRIS, BASS, and CARD. The plots express the results of the three tissue slices from Patient 5. FUSION free are the alternative method of FUSION, but not using scRNA-seq data from 10x Chromium on the post-mortem brain tissue as reference. **Supporting Fig. S11:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using SpaGCN, SEDR, STAGATE, and Stitch3D. The plots express the results of the three tissue slices from Patient 5. **Supporting Fig. S12:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using FUSION, FUSION free, BayesSpace, IRIS, BASS, and CARD. The plots express the results of the three tissue slices from Patient 9. FUSION free are the alternative method of FUSION, but not using scRNA-seq data from 10x Chromium on the post-mortem brain tissue

as reference. **Supporting Fig. S13:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using SpaGCN, SEDR, STAGATE, and Stitch3D. The plots express the results of the three tissue slices from Patient 9. **Supporting Fig. S14:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using FUSION, FUSION free, BayesSpace, IRIS, BASS, and CARD. The plots express the results of the three tissue slices from Patient 10. FUSION free are the alternative method of FUSION, but not using scRNA-seq data from 10x Chromium on the post-mortem brain tissue as reference. **Supporting Fig. S15:** The spatial domain maps generated from the joint analysis of all 12 tissue slices for SCC tissue across four patients using SpaGCN, SEDR, STAGATE, and Stitch3D. The plots express the results of the three tissue slices from Patient 10. **Supporting Fig. S16:** The UMAP embeddings of slices for SCC tissue across four patients (patient 2 to patient 10). The RGB plots showed spots from the three slices (0,1, 2), while the multi-color embedding plots indicate the section identity where the colors illustrate the annotation labels and the coordinates are the value of the embeddings from FUSION, FUSION without WGAN, and BayesSpace. **Supporting Fig. S17:** The UMAP embeddings of slices for SCC tissue across four patients (patient 2 to patient 10). The RGB plots showed spots from the three slices (0,1, 2), while the multi-color embedding plots indicate the topic domains and the corresponding coordinates inferred from BASS, IRIS, and CARD. **Supporting Fig. S18:** The UMAP embeddings of slices for SCC tissue across four patients (patient 2 to patient 10). The RGB plots showed spots from the three slices (0,1, 2), while the multi-color embedding plots indicate the topic domains and the corresponding coordinates inferred from SpaGCN, SEDR, and STAGATE. **Supporting Fig. S19:** Tumor regions of SCC from slices produced by the spatial domains estimated from FUSION, BayesSpace, SEDR, CARD, BASS, STAGATE, Stitch3D, IRIS, and SpaGCN. The top part of the graph is the manual annotation of the tumor region with a silver-standard mask of histologist-defined tumor versus non-tumor regions in in slice P2_ST_rep2. **Supporting Fig. S20:** The spatially complementary tumor-versus-immune distributions across all slices of SCC tissue from patients 2, 9, and 10. The three figures in the first row show the proportions of the three tumor-related cells from patient 2, differentiated tumor cells (Tumor_KC_Diff), proliferative tumor cells (Tumor_KC_Basal), and cycling tumor cells (Tumor_KC_Cyc). The three figures in the second row show the proportions of the three immune cells from patient 2, T cell, B cells, and NK cells. Figures in the bottom row show the proportions of the three immune cells in green and the three tumor cells in red from patients 2, 9, and 10. **Supporting Fig. S21:** a. Heatmap of the paired MOB cell-type proportions by FUSION for all 8 cell types. For each pair of figures, the top figure is the Stereo-seq cell type proportion and the bottom figure is the Slide-seq proportion; b. Bar plot of cell type correlations between matched FUSION domains in the Slide-seq and Stereo-seq sections. **Supporting Fig. S22:** The UMAP embeddings calculated with different methods from the paired mouse olfactory bulb (MOB) spatial transcriptomics datasets of Slide-seq and Stereo-seq platforms. The two-color plots showed spots from the two MOB slices, while the multi-color embedding plots indicate the topic domains and the corresponding coordinates. Figures in the top part express the embeddings of FUSION, BayesSpace, BASS, and IRIS. Figures in the bottom part express the embeddings of FUSION and Harmony, plus the remaining three methods. **Supporting Fig. S23:** The UMAP embeddings calculated with different methods from the paired mouse olfactory bulb (MOB) spatial transcriptomics datasets of Slide-seq and Stereo-seq platforms. The two-color plots showed spots from the two MOB slices, while the multi-color embedding plots indicate the topic domains and the corresponding coordinates. Figures in the top part express the embeddings of CARD, SpaGCN, SEDR, and STAGATE. Figures in the bottom part express the embeddings of the Harmony, plus the above four methods. **Supporting Fig. S24:** The cross-temporal fluctuations of cell type proportions for 12 cell types of mouse embryo from E9.5 to E16.5 by FUSION. **Supporting Table S1:** Implementation and Reproducibility Details. **Supporting Table S2:** Preprocessing and Quality Control Checklist. **Supporting Table S3:** FUSION Hyperparameters (Smoothing&Alignment) Checklist. **Supporting Table S4:** Baseline Methods Comparison Framework Checklist.