ARCADE: CROSS-SECTION MULTIMODAL LEARNING FOR MOSAIC SPATIAL DECONVOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial multi-omics profiles tissues across multiple molecular modalities, but real-world experiments often exhibit heterogeneous modality coverage across sections, giving rise to *mosaic spatial multi-omics*, where different sections contain different subsets of molecular modalities. Deconvolving such data creates an incomplete multi-view latent mixture inference problem, where cell-type compositions must be recovered under heterogeneous modality coverage. Existing spatial deconvolution methods typically require RNA measurements, whereas multimodal integration methods mainly focus on representation alignment without explicitly modeling cell-type compositions. We propose ARCADE, a unified framework for mosaic spatial deconvolution that integrates cross-section multimodal information, spatial structure, and single-cell RNA references through heterogeneous graph learning and reference-guided mixture inference. A directional–magnitude factorization further separates directional information for cross-modal alignment from abundance-related information for composition inference. Experiments on real spatial multi-omics datasets demonstrate consistent improvements over widely used spatial deconvolution methods, with ARI gains over the best competing method averaging 18% across RNA-observed datasets and reaching 98% in the section where RNA is entirely unavailable. Reproducible code is available at `https://anonymous.4open.science/r/submission-D956`.

1 INTRODUCTION

Spatially resolved transcriptomics (SRT) provides molecular measurements while preserving spatial context within tissues, enabling systematic investigation of cellular organization and tissue architecture (Ståhl et al., 2016; Rodriques et al., 2019; Liu et al., 2022; 2023). A fundamental problem in spatial omics analysis is *spatial deconvolution*, which aims to recover the underlying cell-type composition from spatially mixed molecular signals. Existing approaches address this problem by leveraging single-cell RNA-sequencing (scRNA-seq) references, spatial structures, or probabilistic generative models (Cable et al., 2022; Kleshchevnikov et al., 2022; Ma & Zhou, 2024; Dong et al., 2025; Gaspard-Boulinc et al., 2025). However, their RNA-centric design limits their applicability when complementary information is available or when RNA measurements are missing.

This limitation becomes increasingly important as recent spatial multi-omics technologies extend SRT by profiling transcriptomic signals together with additional molecular modalities, such as chromatin accessibility, histone modifications, or protein abundance (Ben-Chetrit et al., 2023; Liu et al., 2024; Zhang et al., 2025). In practice, different tissue sections may be profiled using different experimental protocols, giving rise to *mosaic spatial multi-omics*, where different sections contain different subsets of molecular modalities (Figure 1). For example, one section may contain RNA and protein measurements, another may contain RNA only, and a third may contain protein measurements but no RNA. However, existing deconvolution approaches are primarily designed for RNA-based spatial transcriptomics and cannot directly infer cell-type compositions in sections where RNA is unavailable. We refer to cell-

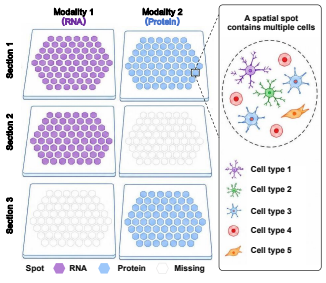


Figure 1: Illustration of mosaic spatial multi-omics and spot-level cell-type mixtures.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

type composition inference under such heterogeneous and incomplete modality coverage as *mosaic spatial deconvolution*, which requires information to be transferred across sections and molecular modalities.

To address such modality-transfer problem, recently proposed spatial mosaic integration methods (Yang et al., 2026; Yan et al., 2026; Zheng et al., 2026; Liu et al., 2026) provide initial solutions to heterogeneous modality configurations by learning shared spot-level representations. At the representation level, this setting can be viewed as incomplete multi-view learning with partially observed modalities, whose goal is to recover shared latent representations from partially observed views (Li et al., 2019; Neverova et al., 2015; Tian et al., 2020). Graph-based domain adaptation methods further motivate the use of structural relationships to transfer information across heterogeneous data distributions (Wu et al., 2020). However, representation alignment alone is insufficient for deconvolution, which requires recovering latent cellular compositions from molecular mixtures. Moreover, most existing formulations assume that each observation corresponds to a single semantic instance. This assumption does not hold in mosaic spatial deconvolution, where each spatial observation is a spatially structured mixture of latent cell populations and the goal is to infer an cell-type composition.

Collectively, mosaic spatial deconvolution poses three fundamental challenges: *incomplete modality coverage*, *latent cell-type composition inference*, and *cross-section cross-modal alignment*. To address these challenges, we propose ARCADE (**A**daptive **R**NA-referenced **C**ross-section Analysis for mosaic spatial **DE**convolution), which formulates the problem as incomplete multi-view latent mixture inference and jointly learns multimodal representations and cell-type compositions through a reference-guided probabilistic framework. Specifically, ARCADE uses RNA as an identity-defining anchor by leveraging widely available annotated scRNA-seq references (Travaglini et al., 2020; Han et al., 2018; Consortium, 2020), while transferring complementary information from paired bridge sections to sections lacking RNA. In real-data experiments, ARCADE consistently outperforms widely used spatial deconvolution methods across multiple evaluation metrics, improving ARI over the best competing method by 18% on average across the two datasets with RNA-observed sections and by 98% in the RNA-missing section. A detailed discussion of related work is provided in Appendix A.

Our contributions are summarized as follows:

- We formulate **mosaic spatial deconvolution** as an **incomplete multi-view latent mixture inference problem**, where spatial sections may contain different subsets of molecular modalities. This formulation provides a unified perspective for cell-type composition inference under heterogeneous modality coverage.
- We propose a **cross-section and cross-modal representation learning framework** that integrates spatial structure, molecular similarity, and cross-section correspondence through multi-relational graph learning and contrastive alignment. A directional–magnitude factorization further decouples directional information for cross-modal alignment from abundance information relevant to cell-type composition, enabling information transfer without discarding composition-relevant signals.
- To the best of our knowledge, ARCADE is the first **joint representation–deconvolution framework for mosaic spatial multi-omics**, coupling multimodal representation learning with reference-guided cell-type composition inference under incomplete modality coverage, including sections where RNA measurements are entirely unavailable.

2 METHOD

We represent a mosaic spatial multi-omics dataset as $\mathcal{D} = \{(\mathcal{X}_b, \mathcal{S}_b, \mathcal{H}_b)\}_{b=1}^{B}$, where $\mathcal{S}_b \in \mathbb{R}^{N_b \times 2}$ contains the coordinates of the N_b spots in section b, $\mathcal{H}_b \subseteq \mathcal{M}$ denotes its observed modality set that may vary across sections, with $\mathcal{M}$ being the global set of molecular modalities,, and $\mathcal{X}_b = \{X_b^{(m)} \in \mathbb{R}^{N_b \times p_m} : m \in \mathcal{H}_b\}$ contains the corresponding molecular measurements, with p_m denoting the number of features for modality m. For spot s in section b, let

$$\mathbf{p}_{bs} = (p_{bs1}, \ldots, p_{bsC})^\top \in [0,1]^C, \text{ such that } \sum_{c=1}^{C} p_{bsc} = 1,$$

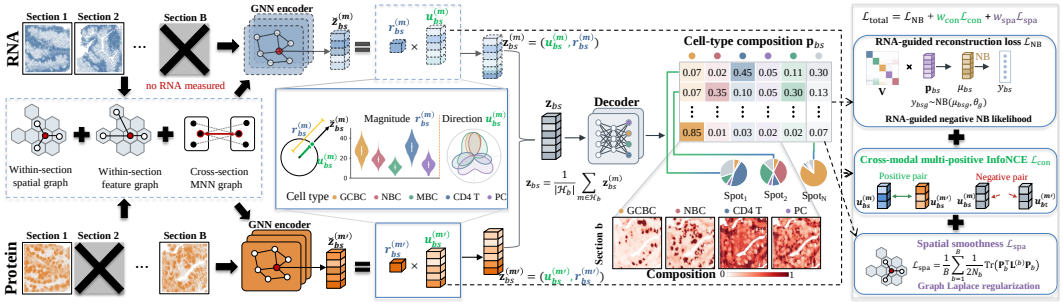


Figure 2: Workflow of ARCADE. Given a spatial mosaic multi-omics dataset and reference expression matrix $\mathbf{V}$, ARCADE integrates multi-relational graph learning, directional–magnitude factorization, and reference-guided mixture modeling to infer cell-type compositions $\mathbf{p}_{bs}$ under incomplete modality coverage.

denote its latent cell-type composition over C cell types. Let $\mathbf{V} = (v_{cg}) \in \mathbb{R}_+^{C\times G}$ denote the observed cell-type-specific reference expression profiles over G informative genes, which are provided as fixed inputs to the model. Our goal is to estimate the latent cell-type compositions $\{\mathbf{p}_{bs}\}_{b,s}$ by integrating the observed molecular measurements with the reference expression profiles, within-section spatial dependencies, and cross-section relationships derived from shared modalities.

Overview of ARCADE. Figure 2 illustrates the workflow of ARCADE. To enable deconvolution under mosaic modality coverage, ARCADE integrates within-section spatial and molecular relationships with cross-section correspondences through multi-relational graph learning, allowing information to propagate across sections with different observed modalities. A key design is the directional–magnitude factorization of modality-specific representations: the directional component facilitates cross-modal alignment, while the magnitude component preserves complementary abundance information relevant to cell-type composition. The resulting representations are aggregated and decoded into cell-type compositions under reference guidance. ARCADE is trained end-to-end with reference-guided reconstruction, cross-modal contrastive alignment, and spatial regularization. Detailed descriptions of each module are provided in the following sections.

2.1 ARCADE ARCHITECTURE

Multi-relational graph construction. For each modality m, we construct a multi-relational graph $\mathcal{G}^{(m)}$ over all spots where modality m is observed, with adjacency matrix

$$\mathbf{A}^{(m)} = \mathbf{A}^{(m)}_{\text{spa}} + w_f \mathbf{A}^{(m)}_{\text{feat}} + w_g \mathbf{A}^{(m)}_{\text{cross}}.$$

Here, $\mathbf{A}^{(m)}_{\text{spa}}$ encodes within-section spatial neighborhoods for sections where modality m is observed. The modality-specific matrix $\mathbf{A}^{(m)}_{\text{feat}}$ captures within-section molecular similarity in the corresponding feature space. Both $\mathbf{A}^{(m)}_{\text{spa}}$ and $\mathbf{A}^{(m)}_{\text{feat}}$ are block diagonal across sections. In contrast, $\mathbf{A}^{(m)}_{\text{cross}}$ connects mutual nearest neighbors across sections sharing modality m, providing cross-section correspondences based on the shared molecular view. Together, these relations capture complementary spatial, molecular, and cross-section structures for modality-specific representation learning. Combined with cross-modal alignment, the cross-section correspondences enable information transfer through bridge sections to sections with incomplete modality coverage. The weights w_f and w_g control the contributions of within-section molecular similarity and cross-section correspondence, respectively. Further graph construction details are provided in Appendix B.1.

Multi-hop graph propagation. For each modality m, we apply LightGCN-style propagation (He et al., 2020) to the multi-relational graph $\mathcal{G}^{(m)}$. Let $N^{(m)} = \sum_{b:\, m\in\mathcal{H}_b} N_b$. The normalized adjacency matrix is defined as

$$\widehat{\mathbf{A}}^{(m)} = \widetilde{\mathbf{D}}_m^{-1/2} \widetilde{\mathbf{A}}^{(m)} \widetilde{\mathbf{D}}_m^{-1/2}, \qquad \widetilde{\mathbf{A}}^{(m)} = \mathbf{A}^{(m)} + \mathbf{I}_{N^{(m)}},$$

where $\widetilde{\mathbf{D}}_m$ is the corresponding degree matrix. Let $\mathbf{F}^{(m)}$ denote the preprocessed low-dimensional feature representation for modality m, with its construction detailed in Appendix C.2. We initialize

$\mathbf{H}_m^{(0)} = \mathbf{F}^{(m)}$ and propagate node representations as

$$\mathbf{H}_m^{(k)} = \widehat{\mathbf{A}}^{(m)}\mathbf{H}_m^{(k-1)}, \quad k = 1, \ldots, K.$$

The representations from all propagation layers are concatenated to capture multi-hop neighborhood information:

$$\mathbf{H}_m^{\text{cat}} = \text{Concat}\left(\mathbf{H}_m^{(0)}, \ldots, \mathbf{H}_m^{(K)}\right).$$

A modality-specific two-layer MLP then maps $\mathbf{H}_m^{\text{cat}}$ to a common d-dimensional latent space, yielding the representation $\breve{\mathbf{z}}_{bs}^{(m)}$ for spot s in section b.

Directional–magnitude factorization. Cross-modal alignment typically relies on L_2-normalized representations, which removes magnitude information that may encode cell-type-specific molecular abundance. To preserve this information, ARCADE factorizes each modality-specific representation into its direction and magnitude (as shown in Figure 2):

$$\mathbf{z}_{bs}^{(m)} = \left[\left(\frac{\breve{\mathbf{z}}_{bs}^{(m)}}{\|\breve{\mathbf{z}}_{bs}^{(m)}\|_2}\right)^\top, \|\breve{\mathbf{z}}_{bs}^{(m)}\|_2\right]^\top \equiv \left[\left(\mathbf{u}_{bs}^{(m)}\right)^\top, r_{bs}^{(m)}\right]^\top.$$

Here, $\mathbf{u}_{bs}^{(m)}$ captures the relative pattern of latent molecular features, providing information for distinguishing different cell-type mixtures and enabling cross-modal contrastive alignment. In contrast, the magnitude $r_{bs}^{(m)}$ retains abundance-related variation discarded by normalization, which can help distinguish mixtures with similar relative patterns but different overall molecular signal levels. This information can be particularly relevant for cell types whose presence or abundance is associated with stronger or weaker signals in the corresponding modality. ARCADE therefore jointly uses both components for cell-type composition inference, consistent with our ablation results showing impaired recovery of specific cell types when the magnitude component is removed.

Partial-modality aggregation. For section b, ARCADE aggregates only the representations of the observed modalities:

$$\mathbf{z}_{bs} = \frac{1}{|\mathcal{H}_b|}\sum_{m \in \mathcal{H}_b} \mathbf{z}_{bs}^{(m)}.$$

Since all modality-specific encoders map their inputs to the same latent space, this operation supports both bridge sections with multiple modalities and sections with only a single modality.

2.2 Reference-guided Latent Mixture Inference

To infer cell-type compositions from the learned multimodal representations, ARCADE combines a simplex composition decoder with a reference-guided negative-binomial observation model. For spot s in section b, the composition is decoded from the aggregated representation $\mathbf{z}_{bs}$ as

$$\mathbf{p}_{bs} = \text{Softmax}\left(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{z}_{bs} + \mathbf{b}_1) + \mathbf{b}_2\right).$$

Reference-guided negative-binomial modeling. To anchor the inferred cell-type compositions to the observed RNA signals and cell-type-specific reference profiles, ARCADE employs an RNA reference-guided negative-binomial model (Cable et al., 2022). For an RNA-observed spot, let $y_{bsg} = [X_b^{(\text{RNA})}]_{sg}$ denote the raw RNA count of gene g at spot s in section b. We model

$$y_{bsg} \sim \text{NB}(\mu_{bsg}, \theta_g), \qquad \mu_{bsg} = n_{bs}\left(\sum_{c=1}^{C} b_c\, p_{bsc}\, v_{cg}\right)\exp(\alpha_{bs} + \gamma_g).$$

Here, n_{bs} and v_{cg} are fixed quantities representing the observed library size and reference expression, respectively. The remaining model quantities are learned from the data: $b_c > 0$ is a cell-type-specific RNA size factor, α_{bs} captures spot-specific capture efficiency, γ_g accounts for gene-specific reference bias, and $\theta_g > 0$ is the gene-specific dispersion parameter. For identifiability, $(b_1, \ldots, b_C)$ is constrained to have unit geometric mean. These parameters, together with the GNN parameters and cell-type compositions $\mathbf{p}_{bs}$, are jointly learned during end-to-end training.

Cross-modal alignment in bridge sections. Since cross-modal alignment should be invariant to modality-specific intensity variation, we apply the contrastive objective to the directional components $\mathbf{u}_{bs}^{(m)}$. Let $\mathcal{B}_{\text{bridge}}$ denote the set of bridge sections in which multiple modalities are measured at matched spots. For an anchor representation $\mathbf{u}_{bs}^{(m)}$, representations from other modalities at the same spot form positive pairs:

$$\mathcal{P}_{bs}^{(m)} = \{(b, s, m') : m' \in \mathcal{H}_b \setminus \{m\}\}.$$

Representations from unmatched spots are used as negative candidates:

$$\mathcal{N}_{bs}^{(m)} = \{(b, t, m') : t \neq s,\ m' \in \mathcal{H}_b\}.$$

We propose a multi-positive InfoNCE objective (Tian et al., 2020):

$$\mathcal{L}_{\text{con}} = \frac{1}{|\mathcal{A}|} \sum_{(b,s,m)\in\mathcal{A}} \left[-\frac{1}{|\mathcal{P}_{bs}^{(m)}|} \sum_{(b,s,m')\in\mathcal{P}_{bs}^{(m)}} \log \frac{\exp\left((\mathbf{u}_{bs}^{(m)})^\top \mathbf{u}_{bs}^{(m')}/\tau\right)}{\sum_{(b',t,\tilde{m})\in\mathcal{P}_{bs}^{(m)}\cup\mathcal{N}_{bs}^{(m)}} \exp\left((\mathbf{u}_{bs}^{(m)})^\top \mathbf{u}_{b't}^{(\tilde{m})}/\tau\right)} \right],$$

where $\mathcal{A} = \{(b, s, m) : b \in \mathcal{B}_{\text{bridge}},\ s = 1, \ldots, N_b,\ m \in \mathcal{H}_b\}$.

The contrastive objective aligns RNA and auxiliary-modality representations within bridge sections. For each shared modality m, the cross-section graph $\mathbf{A}_{\text{cross}}^{(m)}$ then propagates these aligned representations to sections that contain m but lack RNA. The shared cell-type composition decoder is applied to the representation constructed from the available modalities, allowing cell-type compositions to be inferred without a direct RNA likelihood in every section.

2.3 Model Training

ARCADE is trained end-to-end by jointly optimizing reference-guided composition inference, cross-modal alignment, and spatial coherence.

Reference-guided likelihood. For sections with observed RNA measurements, we minimize the negative log-likelihood of the raw RNA counts:

$$\mathcal{L}_{\text{NB}} = -\frac{1}{|\mathcal{B}_{\text{RNA}}|G} \sum_{b\in\mathcal{B}_{\text{RNA}}} \frac{1}{N_b} \sum_{s=1}^{N_b} \sum_{g=1}^{G} \log p_{\text{NB}}\left(y_{bsg} \mid \mu_{bsg}, \theta_g\right),$$

where $\mathcal{B}_{\text{RNA}}$ denotes the set of RNA-observed sections. This likelihood anchors the inferred compositions to the observed RNA counts and cell-type-specific reference profiles.

Spatial regularization. To encourage locally coherent cell-type compositions, we define

$$\mathcal{L}_{\text{spa}} = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{2N_b} \operatorname{Tr}\left(\mathbf{P}_b^\top \mathbf{L}^{(b)} \mathbf{P}_b\right),$$

where $\mathbf{P}_b = \left[\mathbf{p}_{b1}^\top; \ldots; \mathbf{p}_{bN_b}^\top\right] \in \mathbb{R}^{N_b \times C}$, and $\mathbf{L}^{(b)}$ is the spatial graph Laplacian of section b. This term penalizes large composition differences between neighboring spots, thereby promoting local spatial coherence.

Overall objective. Combining the reference-guided likelihood, cross-modal contrastive loss, and spatial regularization gives

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{NB}} + w_{\text{con}}\mathcal{L}_{\text{con}} + w_{\text{spa}}\mathcal{L}_{\text{spa}}.$$

All trainable parameters are jointly optimized using Adam (Kingma & Ba, 2015); optimization details are provided in Appendix B.2.

3 EXPERIMENTS

3.1 DATASETS

We evaluate ARCADE on two real spatial multi-omics datasets with complementary modality configurations. The mouse brain dataset contains RNA paired with three epigenomic modalities, with RNA available in all sections, and is used to assess the benefit of complementary modalities beyond RNA. The human tonsil dataset contains RNA and protein with heterogeneous modality coverage, including an RNA-missing section, and is used to evaluate cross-section transfer under missing RNA. Spatial-domain annotations are supported by the Allen Mouse Brain Atlas (P56) (Sunkin et al., 2012) and the human tonsil atlas (Massoni-Badosa et al., 2024). Dataset sources, preprocessing, and annotation details are provided in Appendix C.1–C.2.

3.2 BASELINES AND EVALUATION PROTOCOL

Baselines. We compare ARCADE against five widely used RNA-based deconvolution methods: Tangram (Biancalani et al., 2021), RCTD (Cable et al., 2022), cell2location (Kleshchevnikov et al., 2022), CARD (Ma & Zhou, 2024) and STitch3D (Wang et al., 2023). Existing spatial mosaic integration methods, such as SpaMosaic (Yan et al., 2026) and SpaMode (Zheng et al., 2026) primarily focus on learning spot-level representations under heterogeneous modality configurations, rather than directly estimating cell-type compositions. Therefore, they are not included as end-to-end deconvolution baselines. Instead, we separately evaluate whether their learned representations can support cell-type composition transfer under RNA-missing settings in Section 3.4.

Metrics. Since ground-truth spot-level cell-type compositions are unavailable, we evaluate agreement with curated spatial-domain annotations using eight complementary metrics: LISI and SC for local separation and domain purity; PAS and CHAOS for spatial coherence; SMS and BEAR for global spatial-pattern and local-boundary agreement; and ARI and NMI for global domain recovery. Lower values are better for LISI, PAS, and CHAOS, whereas higher values are better for the remaining metrics. Detailed definitions, implementation, and discussion of their complementary properties are provided in Appendix C.3.

3.3 ALL-SECTION-RNA-AVAILABLE RESULTS

One of the key advantages of ARCADE is its ability to exploit complementary multimodal information beyond RNA measurements while preserving cell-type composition inference. To quantify this benefit, we evaluate ARCADE on the mouse brain dataset, where RNA is available in all sections and additional epigenomic modalities provide complementary molecular signals. This setting isolates whether integrating complementary multimodal information can improve deconvolution accuracy when RNA measurements are available in all sections.

Does multimodal information improve deconvolution accuracy? The results of ARCADE and competing methods are summarized in Table 1. ARCADE achieves the best performance on seven out of eight evaluation metrics compared with the five widely used spatial deconvolution methods that rely solely on RNA modality. These results demonstrate that incorporating complementary molecular modalities provides additional information beyond RNA measurements and improves the accuracy of cell-type composition estimation. Furthermore, ARCADE consistently outperforms its RNA-surrogate variant, ARCADE-RS, which uses the same architecture, objectives, and optimization procedure but replaces non-RNA modalities in bridge sections with duplicated RNA inputs. By controlling for the number of modalities and model complexity, this comparison demonstrates that the gain comes from integrating complementary molecular modalities rather than simply increasing input information.

Is the estimated cell-type composition by ARCADE biologically meaningful? Beyond quantitative accuracy, ARCADE produces biologically cell-type composition maps. To investigate the spatial enrichment of different cell types, we visualize the estimated cell-type compositions from ARCADE (Figure 3). ARCADE successfully recovers biologically meaningful spatial distributions of major cell populations. Specifically, oligodendrocytes and oligodendrocyte pre-

Table 1: Deconvolution performance on the mouse brain dataset. Metrics are computed per section and averaged over the three sections. Best values in bold, second best underlined. The last row reports the relative improvement of ARCADE over the best competitor, where the improvement is calculated according to the favourable direction of each metric.

METHOD	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
TANGRAM	3.732	0.242	0.138	0.500	0.523	−0.054	0.085	0.147
RCTD	2.443	0.206	0.130	<u>0.626</u>	<u>0.624</u>	0.043	0.280	0.363
CELL2LOCATION	4.309	0.126	**0.076**	0.379	0.537	−0.207	0.056	0.156
CARD	1.975	0.155	0.107	0.537	0.493	0.004	0.221	0.320
STITCH3D	2.375	<u>0.1064</u>	0.079	0.451	0.571	−0.047	0.107	0.225
ARCADE-RS	<u>1.955</u>	0.138	0.094	0.613	0.621	<u>0.071</u>	<u>0.293</u>	<u>0.440</u>
ARCADE	**1.614**	**0.1060**	<u>0.077</u>	**0.632**	**0.632**	**0.130**	**0.367**	**0.510**
IMPROVEMENT (%)	+21.1	+0.4	−1.3	+1.0	+1.3	+83.1	+25.3	+15.9

cursor cells (Oligo/OPC) are enriched in the ccg/aco regions, excitatory cortical neurons (Exc-Ctx) are predominantly localized across cortical layers 1–6, inhibitory neurons (Inh-Meis2) are enriched in the CP/ACB regions, and neuroblasts (Nb) show higher abundance in the VL region. These spatial enrichment patterns are consistent with previously reported biological findings (Sunkin et al., 2012; Zeisel et al., 2018). Compared with the competing methods, ARCADE recovers spatial domains with clear boundaries and strong agreement with the manual annotation, whereas competing methods generally produce noisier boundaries, fragmented domains, or weaker recovery of biologically meaningful spatial organization (Figures S3–S5 in Appendix C.4).

3.4 RNA-MISSING-SECTION RESULTS

A unique capability of ARCADE is its ability to infer cell-type compositions in sections lacking RNA measurements by transferring information from RNA-containing bridge sections. We evaluate this capability on the human tonsil dataset, where sections 1 and 3 contain paired RNA and protein measurements, section 2 contains RNA only, and section 4 contains protein only. Thus, the cell-type compositions of section 4 must be inferred without direct RNA observations.

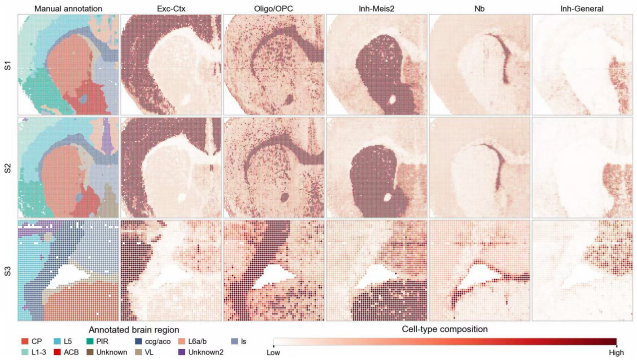


Figure 3: Spatial distribution of annotated brain regions and ARCADE-estimated cell-type compositions on the mouse brain dataset. The five representative cell types shown exhibit distinct spatial enrichment patterns consistent with known anatomical organization.

Can ARCADE accurately infer cell-type compositions under missing RNA? We jointly train ARCADE on all four sections using their available modalities. For comparison with RNA-based deconvolution methods, which require direct RNA measurements, we first evaluate performance on the three RNA-containing sections. As shown in Table 2, ARCADE achieves the best performance on six of eight metrics, with substantial improvements in composition-based metrics including ARI and NMI. These results indicate that jointly modeling heterogeneous modalities does not compromise deconvolution accuracy when RNA is available.

We next focus on the RNA-missing section, which represents the primary challenge of this dataset. Unlike RNA-reference-based deconvolution methods, ARCADE enables cell-type composition inference without direct RNA observations by transferring information through RNA-containing multimodal bridge sections. The inferred compositions closely match manual annotations (Figures S6 and S7 in Appendix C.5), demonstrating that ARCADE can recover spatially organized cell-type mixtures even when the reference modality is completely unavailable.

Table 2: Deconvolution performance on the human tonsil dataset. Metrics are computed per section and averaged over the first three sections, which contain RNA; arrows indicate the favourable direction. Best values in bold, second best underlined.

METHOD	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
TANGRAM	3.031	0.593	0.342	0.645	0.629	−0.007	0.176	0.286
RCTD	2.063	0.378	0.226	<u>0.6865</u>	0.624	0.054	0.261	0.424
CELL2LOCATION	<u>2.002</u>	<u>0.353</u>	<u>0.224</u>	0.617	**0.659**	0.054	0.260	0.4295
CARD	2.247	0.416	0.254	0.649	0.598	0.023	0.241	0.404
STITCH3D	2.009	0.354	0.227	**0.6867**	0.648	<u>0.068</u>	<u>0.290</u>	<u>0.4299</u>
ARCADE	**1.863**	**0.305**	**0.201**	0.664	<u>0.656</u>	**0.092**	**0.321**	**0.488**
IMPROVEMENT (%)	+7.5	+15.7	+11.4	-3.4	-0.5	+35.3	+10.7	+13.5

Table 3: Performance on human tonsil section 4, which contains protein only. ARCADE operates on this section natively; baseline estimates are obtained by transferring RNA-based deconvolution results from RNA-containing sections using spatial mosaic representations. Best values in bold. Second best underlined. The last row reports the relative improvement of ARCADE over the best baseline, where the improvement is calculated according to the favourable direction of each metric.

METHOD	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
TANGRAM	3.758	0.607	0.329	0.565	<u>0.633</u>	−0.020	0.044	0.134
RCTD	3.254	0.547	<u>0.316</u>	0.600	0.519	<u>−0.019</u>	0.087	0.196
CELL2LOCATION	3.400	0.558	0.341	0.532	0.608	−0.033	<u>0.151</u>	<u>0.224</u>
CARD	3.555	0.593	0.353	0.572	0.544	−0.051	0.081	0.174
STITCH3D	<u>3.170</u>	<u>0.544</u>	0.324	<u>0.615</u>	0.566	−0.027	0.039	0.190
ARCADE	**1.850**	**0.316**	**0.210**	**0.709**	**0.639**	**0.089**	**0.299**	**0.477**
IMPROVEMENT (%)	+71.4	+72.2	+50.5	+15.3	+0.9	+568.4	+98.0	+112.9

Can representation alignment alone recover compositions under missing RNA? A natural alternative to ARCADE is to decouple representation learning and composition inference: first learn cross-section aligned representations using mosaic integration methods, and then transfer cell-type compositions estimated by RNA-based deconvolution methods to RNA-missing sections. We evaluate this strategy using SpaMosaic followed by a k-nearest neighbor (kNN) composition transfer model.

As shown in Table 3, representation-only transfer is insufficient for accurate cell-type composition inference in the RNA-missing section. Although RNA-based deconvolution methods perform reasonably when RNA is directly available, their transferred predictions degrade substantially when RNA is missing. In contrast, ARCADE jointly learns multimodal representations and latent cell-type compositions, enabling information from RNA-containing bridge sections to directly guide composition inference. Additional qualitative comparisons (Figure S9) and alternative transfer strategies (Tables S2 and S3) are provided in Appendix C.5, further supporting the advantage of joint representation and composition learning over representation-only transfer.

3.5 MODEL ANALYSIS

We further investigate the contribution, robustness, and scalability of ARCADE through ablation studies, sensitivity analysis, and runtime evaluation.

Component analysis. We evaluate the contributions of cross-modal contrastive alignment, spatial regularization, and directional–magnitude factorization on the mouse brain dataset (Table 4). Removing the contrastive loss causes the largest performance degradation, highlighting its importance for cross-modal knowledge transfer. Spatial regularization provides additional improvements, while removing the magnitude component degrades five of eight metrics. On the human tonsil dataset, removing the magnitude component further prevents naive B cells from being identified as dominant in any spot, compared with 579 spots under the full model (Figure 4(a)). RNA and protein marker patterns further support the biological relevance of the inferred naive B-cell-dominant regions in both

Table 4: Ablation of the ARCADE components on mouse brain dataset. Metrics are averaged over all sections; rows below give relative changes (%) with respect to the full model. Arrows indicate the favourable direction, so a change opposing the arrow denotes degradation. "w/o mag." uses only the directional components, discarding magnitude information.

CONFIGURATION	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
W/O $\mathcal{L}_{\text{con}}$	+28.8%	+35.4%	+27.8%	−5.2%	−11.9%	−70.5%	−29.7%	−24.3%
W/O $\mathcal{L}_{\text{spa}}$	+1.6%	+2.8%	+1.9%	+0.3%	+0.3%	−10.4%	−1.7%	−2.4%
W/O MAG.	−0.8%	−7.1%	−6.7%	−9.2%	−0.8%	−28.2%	−6.3%	−4.8%

RNA-containing and RNA-missing sections (Appendix Figure S8). Together, these results demonstrate the complementary contribution of magnitude information to cell-type composition inference.

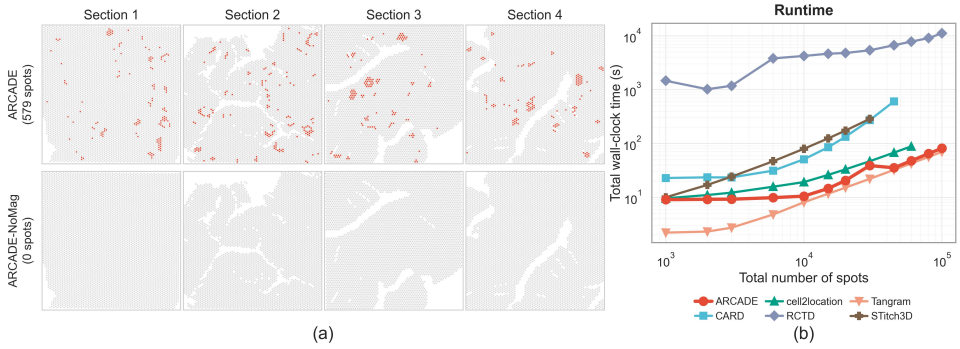


Figure 4: (a) Effect of discarding the magnitude component, illustrated for naive B cells on the human tonsil dataset. Spatial heatmap in each panel highlights spots dominated by naive B cells (colored), with remaining spots shown in gray. (b) Scalability with increasing numbers of spots; several competing methods fail to scale to larger numbers of spots because of out-of-memory errors.

Scalability. We evaluate ARCADE with increasing numbers of spots by downsampling or oversampling the human tonsil dataset. ARCADE scales approximately linearly with dataset size, processing 100,000 spots in less than 100 seconds while remaining computationally competitive with RNA-based deconvolution methods, despite jointly modeling multiple modalities (Figure 4(b)).

Sensitivity analysis to graph construction choices. We further examine the sensitivity of ARCADE to graph construction hyperparameters, including neighborhood sizes (k_{feat}, k_{mnn}) and graph fusion weights (w_f, w_g). As shown in Table S4 and Figures S10–S11 in Appendix C.6, ARCADE maintains stable performance across a broad range of neighborhood sizes, while the sensitivity to graph fusion weights reflects the tissue-dependent contributions of molecular similarity and cross-section correspondence.

4 CONCLUSION

We introduce ARCADE, a unified framework for mosaic spatial deconvolution under heterogeneous and incomplete modality coverage. In contrast to spatial deconvolution methods that typically require RNA measurements and spatial mosaic integration methods that primarily learn aligned representations, ARCADE jointly performs multimodal representation learning and reference-guided cell-type composition inference. By integrating spatial, molecular, and cross-section relationships with cross-modal alignment and latent mixture inference, ARCADE enables deconvolution even in sections where RNA is entirely unavailable. Experiments on two real spatial multi-omics datasets demonstrate the benefits of jointly modeling representation alignment and cell-type composition inference under mosaic modality coverage. More broadly, ARCADE provides a latent mixture inference framework for incomplete multi-view learning with structured spatial observations.

REFERENCES

Alma Andersson, Joseph Bergenstråhle, Michaela Asp, Ludvig Bergenstråhle, Aleksandra Jurek, José Fernández Navarro, and Joakim Lundeberg. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications biology*, 3(1):565, 2020.

Nir Ben-Chetrit, Xiang Niu, Ariel D Swett, Jesus Sotelo, Maria S Jiao, Caitlin M Stewart, Catherine Potenski, Paulius Mielinis, Patrick Roelli, Marlon Stoeckius, et al. Integration of whole transcriptome spatial profiling with protein markers. *Nature biotechnology*, 41(6):788–793, 2023.

Tommaso Biancalani, Gabriele Scalia, Lorenzo Buffoni, Raghav Avasthi, Ziqing Lu, Aman Sanger, Neriman Tokcan, Charles R Vanderburg, Åsa Segerstolpe, Meng Zhang, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature Methods*, 18(11): 1352–1362, 2021.

Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, pp. 3121–3124. IEEE, 2010.

Dylan M Cable, Evan Murray, Luli S Zou, Aleksandrina Goeva, Evan Z Macosko, Fei Chen, and Rafael A Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology*, 40(4):517–526, 2022.

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11221–11229, 2024.

The Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590–595, 2020.

Helena L Crowell, Laura Llaó-Cid, Gerard Frigola, Samuel Gunz, Irene Ruano, Patricia Lorden, Max Ruiz, Marta Kulis, José Ignacio Martin-Subero, Holger Heyn, et al. A transcriptional map of human tonsil architecture: beyond the sum of (single cell) parts. *European Journal of Immunology*, 56(1):e70121, 2026.

Qishi Dong, Yi Yang, Ziye Luo, Haipeng Shen, Xingjie Shi, and Jin Liu. Robust spatial cell-type deconvolution with qualitative reference for spatial transcriptomics. *Small Methods*, 9(5): 2401145, 2025.

Rui Dong and Guo-Cheng Yuan. Spatialdwls: accurate deconvolution of spatial transcriptomic data. *Genome biology*, 22(1):145, 2021.

Marc Elosua-Bayes, Paula Nieto, Elisabetta Mereu, Ivo Gut, and Holger Heyn. Spotlight: seeded nmf regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic acids research*, 49(9):e50–e50, 2021.

Lucie C Gaspard-Boulinc, Luca Gortana, Thomas Walter, Emmanuel Barillot, and Florence MG Cavalli. Cell-type deconvolution methods for spatial transcriptomics. *Nature Reviews Genetics*, 26(12):828–846, 2025.

Jeffrey M Granja, M Ryan Corces, Sarah E Pierce, S Tansu Bagdatli, Hani Choudhry, Howard Y Chang, and William J Greenleaf. Archr is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature genetics*, 53(3):403–411, 2021.

Xiaoping Han, Renying Wang, Yincong Zhou, Lijiang Fei, Huiyu Sun, Shujing Lai, Assieh Saadatpour, Ziming Zhou, Haide Chen, Fang Ye, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*, 172(5):1091–1107, 2018.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 639–648, 2020.

Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology*, 40(5):661–671, 2022.

Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature methods*, 16(12):1289–1296, 2019.

Haoyang Li, Juexiao Zhou, Zhongxiao Li, Siyuan Chen, Xingyu Liao, Bin Zhang, Ruochi Zhang, Yu Wang, Shiwei Sun, and Xin Gao. A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nature communications*, 14(1):1548, 2023.

Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. Flexible multi-view representation learning for subspace clustering. In *Ijcai*, volume 2019, pp. 2916–2922, 2019.

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11169–11178. IEEE, 2021.

Wei Liu, Xu Liao, Yi Yang, Huazhen Lin, Joe Yeong, Xiang Zhou, Xingjie Shi, and Jin Liu. Joint dimension reduction and clustering analysis of single-cell rna-seq and spatial transcriptomics data. *Nucleic acids research*, 50(12):e72–e72, 2022.

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

Wei Liu, Dewei Xiang, Xiaolu Jiang, Yihao Bai, Shuyue Li, Wenlin Li, Shaogao Lv, Qian Li, Jiakun Jiang, and Jin Liu. Spatial multi-omics imputation and embedding with spamie. *Cell Reports Methods*, 6(7):101456, 2026.

Xiaojie Liu, Ting Peng, Miaochun Xu, Shitong Lin, Bai Hu, Tian Chu, Binghan Liu, Yashi Xu, Wencheng Ding, Li Li, et al. Spatial multi-omics: deciphering technological landscape of integration of multi-omics and its applications. *Journal of Hematology & Oncology*, 17(1):72, 2024.

Yahui Long, Kok Siong Ang, Raman Sethi, Sha Liao, Yang Heng, Lynn van Olst, Shuchen Ye, Chengwei Zhong, Hang Xu, Di Zhang, et al. Deciphering spatial domains from spatial multi-omics with spatialglue. *Nature Methods*, pp. 1–10, 2024.

Romain Lopez, Baoguo Li, Hadas Keren-Shaul, Pierre Boyeau, Merav Kedmi, David Pilzer, Adam Jelinski, Ido Yofe, Eyal David, Allon Wagner, et al. Destvi identifies continuums of cell types in spatial transcriptomics data. *Nature biotechnology*, 40(9):1360–1369, 2022.

Ying Ma and Xiang Zhou. Accurate and efficient integrative reference-informed spatial domain detection for spatial transcriptomics. *Nature Methods*, 21(7):1231–1244, 2024.

Ramon Massoni-Badosa, Sergio Aguilar-Fernández, Juan C Nieto, Paula Soler-Vila, Marc Elosua-Bayes, Domenica Marchese, Marta Kulis, Amaia Vilas-Zornoza, Marco Matteo Bühler, Sonal Rashmi, et al. An atlas of cells in the human tonsil. *Immunity*, 57(2):379–399, 2024.

Natalia Neverova, Christian Wolf, Graham Taylor, and Florian Nebout. Moddrop: adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (8):1692–1706, 2015.

HANA Řezanková. Different approaches to the silhouette coefficient calculation in cluster evaluation. In *21st international scientific conference AMSE applications of mathematics and statistics in economics*, pp. 1–10, 2018.

Samuel G Rodriques, Robert R Stickels, Aleksandrina Goeva, Carly A Martin, Evan Murray, Charles R Vanderburg, Joshua Welch, Linlin M Chen, Fei Chen, and Evan Z Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

Lulu Shang and Xiang Zhou. Spatially aware dimension reduction for spatial transcriptomics. *Nature communications*, 13(1):7203, 2022.

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

Susan M Sunkin, Lydia Ng, Chris Lau, Tim Dolbeare, Terri L Gilbert, Carol L Thompson, Michael Hawrylycz, and Chinh Dang. Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic acids research*, 41(D1):D996–D1008, 2012.

Jingjing Tang, Qingqing Yi, Saiji Fu, and Yingjie Tian. Incomplete multi-view learning: Review, analysis, and prospects. *Applied Soft Computing*, 153:111278, 2024.

Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *European conference on computer vision*, pp. 776–794. Springer, 2020.

Kyle J Travaglini, Ahmad N Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V Sit, Stephen Chang, Stephanie D Conley, Yasuo Mori, Jun Seita, et al. A molecular cell atlas of the human lung from single-cell rna sequencing. *Nature*, 587(7835):619–625, 2020.

Gefei Wang, Jia Zhao, Yan Yan, Yang Wang, Angela Ruohao Wu, and Can Yang. Construction of a 3d whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nature Machine Intelligence*, 5(11):1200–1213, 2023.

Mike B Wright. Speeding up the hungarian algorithm. *Computers & Operations Research*, 17(1): 95–96, 1990.

Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of the web conference 2020*, pp. 1457–1467, 2020.

Lulu Yan and Xiaoqiang Sun. Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics*, 39(1):btac805, 2023.

Xuhua Yan, Zhaoyu Fang, Kok Siong Ang, Lynn van Olst, Alex Edwards, Thomas Watson, Ruiqing Zheng, Di Zhang, Rong Fan, Min Li, et al. Mosaic integration of spatial multi-omics with spamosaic. *Nature Genetics*, pp. 1–12, 2026.

Qixing Yang, Yan Wang, Luonan Chen, and Chunman Zuo. stmixer for scalable mosaic integration and label transfer in spatial histology and multi-omics. *Advanced Science*, pp. e75905, 2026.

Amit Zeisel, Hannah Hochgerner, Peter Lönnerberg, Anna Johnsson, Fatima Memic, Job Van Der Zwan, Martin Häring, Emelie Braun, Lars E Borm, Gioele La Manno, et al. Molecular architecture of the mouse nervous system. *Cell*, 174(4):999–1014, 2018.

Di Zhang, Leslie A Rubio Rodríguez-Kirby, Yingxin Lin, Wenqi Wang, Mengyi Song, Li Wang, Lijun Wang, Shigeaki Kanatani, Tony Jimenez-Beristain, Yonglong Dang, et al. Spatial dynamics of brain development and neuroinflammation. *Nature*, 647(8088):213–227, 2025.

Peimeng Zhen, Han Shu, Bingtao Wang, Yongtian Wang, Jialu Hu, Jiajie Peng, Xuequn Shang, Tao Wang, and Jing Chen. Mosaic integration of spatial multiomic data based on hierarchical graph contrastive learning with spatialmosi. *Genome Research*, 2026.

Xubin Zheng, Xinlei Huang, Xiang Zhou, Dian Meng, Ziyue Qiao, Zhiyuan Yuan, and Lixin Cheng. Spamode: A broadly applicable framework for deciphering spatial multi-omics using multimodal mixture of disentangled experts. *Advanced Science*, pp. e75478, 2026.

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

AI USE DISCLOSURE

Generative AI tools were used to assist with language editing and improve the clarity and readability of the manuscript. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final content of the manuscript.

Appendix

Table of Contents

A Related Work **15**

B Additional Implementation Details **17**

B.1 Multi-Relational Graph Construction 17

B.2 Optimization Strategy 17

C Additional Experimental Materials **18**

C.1 Real Data Details 18

C.2 Data Preprocessing 19

C.3 Evaluation Metrics 20

C.4 Additional Results with RNA Available in All Sections 20

C.5 Additional Results for the RNA-Missing Section 22

C.6 Additional Sensitivity Analysis 23

A Related Work

Spatial deconvolution. Spatial transcriptomic measurements obtained at multicellular resolution represent mixtures of signals from multiple cell populations, motivating a large body of work on estimating cell-type compositions from spatial gene expression. Existing methods span probabilistic generative models, matrix factorization and regression approaches, graph-based methods, and single-cell-to-spatial mapping strategies (Li et al., 2023; Yan & Sun, 2023). Among reference-guided probabilistic approaches, RCTD models spatial transcriptomic counts using cell-type-specific expression profiles while accounting for platform effects (Cable et al., 2022), whereas cell2location uses a hierarchical Bayesian model to estimate cell-type abundances while accounting for technical variation (Kleshchevnikov et al., 2022). Stereoscope similarly adopts a probabilistic count model to infer cell-type mixtures from spatial transcriptomic measurements (Andersson et al., 2020), and DestVI extends probabilistic deconvolution to capture continuous within-cell-type transcriptional variation (Lopez et al., 2022). Other representative methods include SPOTlight (Elosua-Bayes et al., 2021), SpatialDWLS (Dong & Yuan, 2021), and CARD (Ma & Zhou, 2024), which exploit matrix factorization, regression, or spatial dependence to improve composition estimation. Tangram instead learns a mapping between single-cell and spatial transcriptomic measurements (Biancalani et al., 2021), while STitch3D jointly analyzes multiple spatial transcriptomic sections together with a single-cell reference to recover three-dimensional tissue structure (Wang et al., 2023).

Despite substantial methodological differences, most spatial deconvolution methods are designed for spatial transcriptomic data and therefore require RNA measurements in the target section. This assumption becomes restrictive in mosaic spatial multi-omics, where some sections may contain protein, chromatin, or other molecular measurements without RNA. ARCADE aims to solve the deconvolution objective in this setting by transferring information across modalities and sections while retaining explicit reference-guided inference of cell-type compositions.

Spatial multi-omics integration. The increasing availability of spatially resolved transcriptomic, proteomic, and epigenomic measurements has motivated methods for integrating multiple molecular views while preserving spatial organization. SpatialGlue, for example, uses graph neural networks and dual attention to integrate spatial location with multiple omics modalities for spatial-domain identification (Long et al., 2024). Such approaches primarily address vertical integration, where multiple modalities are observed within the same tissue section. Related methods for multi-section spatial transcriptomics instead focus on horizontal integration across sections, including PRECAST (Liu et al., 2023), STAligner (Zhou et al., 2023) and STitch3D (Wang et al., 2023). These developments establish spatially aware representation learning as an important computational paradigm for integrating heterogeneous spatial datasets.

A more challenging setting arises when modality availability itself varies across sections, leading to *mosaic spatial multi-omics*. Recent methods have been developed specifically for this regime. SpaMosaic combines graph neural networks with contrastive learning to construct a modality-agnostic, batch-corrected latent space from sections with partially overlapping modalities, supporting spatial-domain identification and missing-modality imputation (Yan et al., 2026). SpaMode provides a more general framework for vertical, horizontal, and mosaic spatial multi-omics integration by disentangling modality-invariant and modality-specific information and translating missing modalities across sections (Zheng et al., 2026). SpaMIE addresses systematic missing modalities through a two-stage strategy that first performs spatially informed cross-modal imputation and subsequently integrates measured and imputed profiles into a unified representation (Liu et al., 2026). stMixer further considers scalable mosaic integration and label transfer across heterogeneous spatial histology and multi-omics datasets through cross-attention, multimodal metric learning, and graph-guided label propagation (Yang et al., 2026). More recently, SpatialMOSI employs hierarchical graph contrastive learning to jointly address cross-modality and cross-section variation in mosaic spatial multi-omics (Zhen et al., 2026).

These methods demonstrate that heterogeneous modality coverage can be effectively addressed through cross-modal alignment, missing-modality translation, and spatially informed representation learning. Importantly, therefore, the distinction between ARCADE and spatial mosaic integration is *not* whether heterogeneous modality configurations can be handled. Rather, the methods address different primary inference objectives. Existing mosaic integration approaches principally

seek integrated representations, spatial domains, missing-modality reconstruction, or label transfer. ARCADE instead targets *cell-type composition inference*: it couples cross-section and cross-modal representation learning with a reference-guided latent mixture model, allowing information transferred through multimodal bridge sections to directly constrain cell-type compositions, including in sections where RNA is unavailable.

Incomplete multi-view learning. From a machine-learning perspective, mosaic spatial multi-omics is closely related to incomplete multi-view learning, in which individual observations are represented by multiple views but some views are unavailable for a subset of observations. A substantial literature has addressed this problem through matrix factorization, subspace learning, graph learning, view completion, and deep representation learning; recent surveys provide systematic reviews of these formulations (Tang et al., 2024). Early incomplete multi-view methods typically sought a shared latent representation while accounting for missing observations. For example, partial-view clustering learns a common subspace from incomplete paired views (Li et al., 2014), while DAIMC uses weighted semi-nonnegative matrix factorization and alignment regularization to learn a common latent representation from more than two incomplete views (Hu & Chen, 2019).

More recent approaches combine representation learning with explicit cross-view prediction or contrastive objectives. COMPLETER jointly learns consistent representations and recovers missing views by combining cross-view contrastive learning with dual prediction (Lin et al., 2021). Related contrastive incomplete multi-view methods further exploit agreement and complementary information across available views to obtain robust representations under missing-view patterns (Chao et al., 2024). These developments are conceptually related to ARCADE because they use partially overlapping observations to transfer information between views rather than requiring every view to be observed for every sample.

Nevertheless, mosaic spatial deconvolution differs from conventional incomplete multi-view learning in several important respects. First, spatial observations are not independent: neighboring spots exhibit structured relationships that provide additional information for representation and composition inference. Second, modality missingness can occur systematically at the section level rather than independently across observations. Third, and most importantly, the target is not merely a shared representation or a cluster assignment. Each spatial spot is a mixture of latent cell populations, and the desired output is a simplex-valued vector of cell-type compositions anchored to external single-cell reference information. ARCADE therefore connects incomplete multi-view representation learning with structured latent mixture inference, using spatial and cross-section relationships to transfer information while retaining an explicit biological interpretation of the inferred compositions.

Cross-modal contrastive and graph representation learning. ARCADE is also related to broader developments in multimodal contrastive learning and graph representation learning. Contrastive multiview coding demonstrates that representations can be learned by maximizing agreement between different views of the same underlying observation (Tian et al., 2020). This principle has subsequently become a common mechanism for aligning heterogeneous modalities and is particularly relevant when different data views contain complementary information. In spatial omics, graph neural networks provide an additional mechanism for incorporating local tissue organization, since spots can be connected according to spatial proximity, molecular similarity, or cross-section correspondence. SpatialGlue (Long et al., 2024), SpaMosaic (Yan et al., 2026), SpaMIE (Liu et al., 2026), and related spatial integration methods illustrate the utility of combining graph-based propagation with multimodal alignment.

Existing cross-modal representation learning methods primarily optimize representation-level objectives, leaving open how transferred information can be translated into cell-type compositions under mosaic modality coverage. ARCADE addresses this gap by jointly coupling cross-modal and cross-section representation learning with reference-guided latent mixture inference. Importantly, ARCADE introduces a directional–magnitude factorization to accommodate the distinct information requirements of these two objectives: the directional component supports cross-modal alignment, while the magnitude component preserves complementary abundance-related information for composition inference. This joint formulation enables transferred multimodal information to directly inform simplex-valued cell-type composition estimation, including in sections where RNA is unavailable.

B Additional Implementation Details

B.1 Multi-Relational Graph Construction

In the graph construction, for each modality m, we construct a global multi-relational graph $\mathcal{G}^{(m)} = (\mathcal{V}^{(m)}, \mathcal{E}^{(m)})$ over all spots measured by modality m across all sections. Specifically, $\mathcal{V}^{(m)} = \bigcup_{b:m\in\mathcal{H}_b} \mathcal{V}_b^{(m)}$, where nodes represent spots profiled by modality m, and edges encode three complementary types of biological relationships. Specifically, the adjacency matrix is defined as

$$\mathbf{A}^{(m)} = \mathbf{A}^{(m)}_{\text{spa}} + w_f \mathbf{A}^{(m)}_{\text{feat}} + w_g \mathbf{A}^{(m)}_{\text{cross}},$$

where the three components encode within-section spatial dependency, within-section molecular similarity, and cross-section correspondence, respectively.

The spatial component $\mathbf{A}^{(m)}_{\text{spa}}$ preserves local tissue organization, as neighboring spots often exhibit related biological states. For each section b, we construct a density-adaptive spatial kNN graph. Specifically, the neighborhood size of spot s in section b is determined as

$$k_{bs} = \text{clip}\left(k_{\text{base}} \left(\frac{\rho_{bs}}{\text{median}(\boldsymbol{\rho}_b)}\right)^{\alpha}, k_{\min}, k_{\max}\right).$$

where ρ_{bs} denotes the local spot density estimated from inter-spot distances within section b, $\boldsymbol{\rho}_b$ collects these density estimates, and α controls the strength of density adaptation. The parameters k_{base}, $k_{\min}$, and $k_{\max}$ denote the baseline, minimum, and maximum neighborhood sizes, respectively. The spatial graphs of sections where modality m is observed are assembled into the block-diagonal matrix $\mathbf{A}^{(m)}_{\text{spa}}$.

The molecular similarity component $\mathbf{A}^{(m)}_{\text{feat}}$ captures biological relationships beyond physical proximity. Specifically, we project $X_b^{(m)}$ into a low-dimensional PCA space and construct a kNN graph, where each spot is connected to its k_{feat} nearest neighbors with similar molecular profiles. The resulting block-diagonal adjacency matrix is denoted as $\mathbf{A}^{(m)}_{\text{feat}}$.

The cross-section correspondence component $\mathbf{A}^{(m)}_{\text{cross}}$ is critical for information transfer under mosaic modality configurations. For each modality m, we construct a mutual nearest neighbor (MNN) graph among spots from different sections sharing the same modality. Specifically, candidate neighbors are identified based on the k_{mnn} nearest neighbors in the PCA feature space using approximate nearest-neighbor search, and mutual pairs are retained to form cross-section edges. These cross-section edges connect biologically corresponding locations across sections, enabling knowledge transfer from bridge sections to sections with missing modalities. The resulting adjacency matrix is denoted as $\mathbf{A}^{(m)}_{\text{cross}}$.

B.2 Optimization Strategy

The model is optimized using the Adam optimizer with separate parameter groups for the graph encoder and the deconvolution module to account for their distinct learning dynamics.

Specifically, the graph encoder parameters $\boldsymbol{\Theta}_{\text{GNN}}$ are updated with a higher learning rate to facilitate rapid structural representation learning, while the deconvolution-related parameters $\boldsymbol{\Theta}_{\text{NB}}$ are trained with a smaller learning rate to ensure stable estimation of the negative binomial likelihood.

To improve training stability, we employ a two-tier optimization scheme:

$$\boldsymbol{\Theta}_{\text{GNN}} \leftarrow \boldsymbol{\Theta}_{\text{GNN}} - \eta_{\text{GNN}} \nabla \mathcal{L}, \quad \boldsymbol{\Theta}_{\text{NB}} \leftarrow \boldsymbol{\Theta}_{\text{NB}} - \eta_{\text{NB}} \nabla \mathcal{L},$$

with $\eta_{\text{GNN}} > \eta_{\text{NB}}$. In the reported experiments, hyperparameters are specified at the dataset level and shared across sections rather than tuned separately for individual tissue slices. The model is trained for 200 epochs with $\eta_{\text{GNN}} = 0.01$, $\eta_{\text{NB}} = 0.001$, $\tau = 0.01$, $w_{\text{con}} = 1$, and $w_{\text{spa}} = 0.01$. For computational efficiency, contrastive learning is performed using mini-batches for large bridge sections, while the negative binomial likelihood is evaluated on full sections with chunked gene processing.

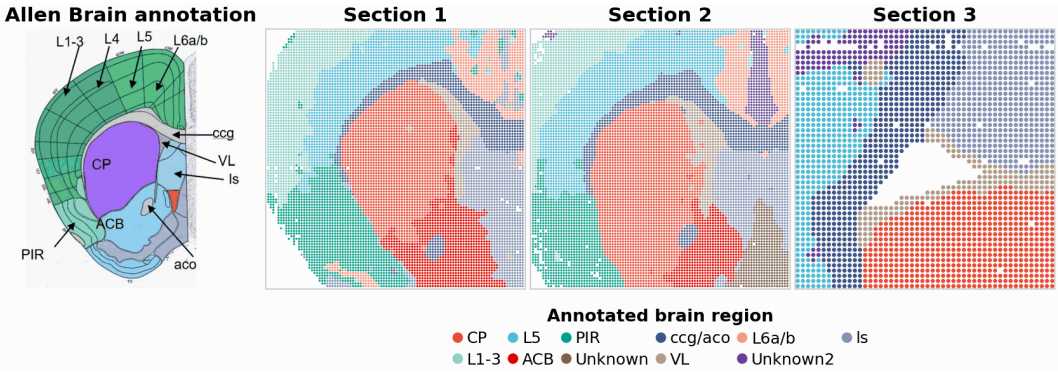


Figure S1: Reference atlas and manual spatial-domain annotations for the mouse brain dataset. The leftmost panel shows a coronal section from the Allen Mouse Brain Atlas (P56) with annotated brain regions. The remaining panels display the manually curated spatial-domain annotations for the three sections, obtained by matching spatial patterns to the corresponding reference regions.

C Additional Experimental Materials

C.1 Real Data Details

Mouse brain dataset. Three spatial sections were profiled, with each section containing RNA paired with a distinct epigenomic modality: H3K27ac, H3K27me3, or ATAC, respectively. All sections contain RNA measurements, making this dataset suitable for evaluating the additional benefit of complementary epigenomic modalities beyond transcriptomic information. The three sections contain 9,363, 9,692, and 2,370 spots, respectively, with the third section containing substantially fewer spatial spots. The single-cell reference atlas contains 12 annotated cell types. Cells labeled as "Unknown" or "Low Quality" were excluded, resulting in 10 cell types retained for downstream inference. The spatial omics data are publicly available from GEO (accession: GSE205055), and the corresponding single-cell RNA-seq reference data are available from ArrayExpress (accession: E-MTAB-11115). Spatial domain annotations were obtained by first identifying spatial domains using SpatialGlue for each section and subsequently refining the domain assignments based on anatomical structures in the Allen Mouse Brain Atlas (P56). The final annotations were manually reviewed to ensure consistency with known brain regions. The reference atlas and annotated spatial domains for the three sections are shown in Figure S1.

Human tonsil dataset. Four spatial sections were profiled, with the first two sections originating from one donor and the remaining two from another donor. The sections exhibit heterogeneous modality configurations: sections 1 and 3 contain paired RNA and protein measurements and serve as bridge sections that enable information transfer between RNA- and protein-only sections, whereas section 2 contains RNA measurements only and section 4 contains protein measurements only. Section 4 provides the primary RNA-missing evaluation setting, where cell-type compositions must be inferred without direct transcriptomic observations. After quality control, the four sections contain 4,908, 4,194, 4,337, and 4,519 spots, respectively. The processed data include 7,924 genes and 30 proteins. The single-cell RNA reference atlas contains 23 annotated cell types; cell types with fewer than 10 reference cells are excluded, resulting in 19 cell types used for inference. The four spatial multi-omics sections were obtained from multiple public sources, including 10x Genomics and the NCBI Gene Expression Omnibus (GEO). The 10x Genomics datasets are available at `https://www.10xgenomics.com/datasets/visium-cytassist-gene-and-protein-expression-library-of-human-tonsil-with-add-on-antibodies-h-e-6-5-mm-ffpe-2-standard` and `https://www.10xgenomics.com/datasets/gene-protein-expression-library-of-human-tonsil-cytassist-ffpe-2-standard`, while the remaining sections are available from GEO under accession numbers GSM8195495 and GSM8195497. The corresponding single-cell RNA reference atlas is available at Zenodo: `https://zenodo.org/records/8373756`. Spatial domain annotations were obtained independently for each section by running SpatialGlue on the paired RNA

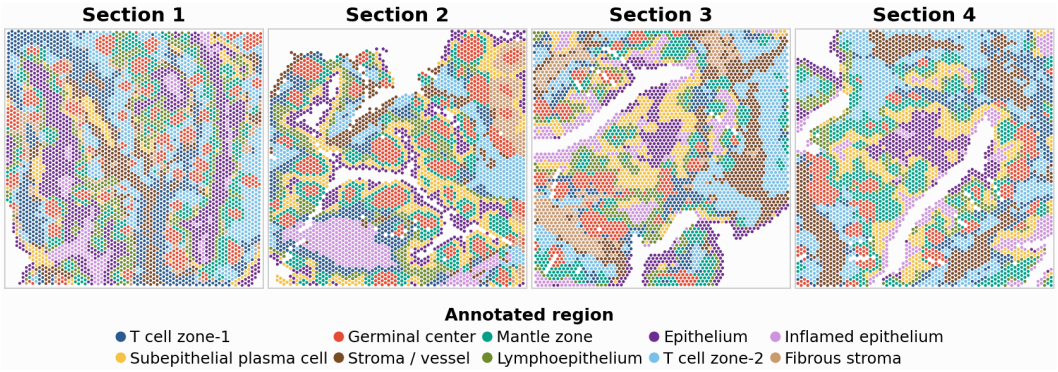


Figure S2: Manual spatial-domain annotations for the four human tonsil sections.

and protein measurements, followed by Gaussian mixture clustering of the fused embedding. This annotation procedure uses both modalities for all four sections and is therefore independent of the modality configuration presented to the deconvolution methods. The resulting domain assignments were subsequently annotated according to known tonsil tissue architecture (Massoni-Badosa et al., 2024; Crowell et al., 2026), including germinal centers, lymphoid follicles, T-cell-enriched regions, epithelial regions, and stromal areas, and supported by characteristic spatial patterns of canonical marker genes and proteins. The final annotations were manually reviewed to ensure consistency with expected spatial organization. The annotated spatial domains for the four sections are shown in Figure S2.

C.2 Data Preprocessing

RNA modality preprocessing Since RNA serves as the anchoring modality, both spatial transcriptomic data and scRNA-seq references were preprocessed before model training. Following (Ma & Zhou, 2024), scRNA-seq cell types with fewer than 10 cells were excluded. Cell-type-specific informative genes were identified using differential expression analysis: genes with log-fold change larger than 1.25 for each cell type were retained as candidate markers, followed by variance-to-mean ratio filtering to remove highly unstable genes. Genes above the 99th percentile of the averaged variance-to-mean ratio were excluded. The remaining genes were intersected with those available in all spatial transcriptomic datasets, and the resulting scRNA-seq expression profiles were averaged within each cell type and normalized to construct the reference expression matrix.

For spatial transcriptomic data, spots with fewer than 10 detected genes were removed for quality control. RNA measurements from all sections were concatenated, log-normalized, and processed using 5,000 highly variable genes while accounting for section-specific variation. PCA was performed on the selected genes, retaining the top 50 principal components, followed by Harmony (Korsunsky et al., 2019) to mitigate cross-section technical variation. The resulting batch-corrected representation was used for RNA-specific graph construction.

Protein modality preprocessing Protein expression data were processed using sequential quality control and normalization. Specifically, spots with fewer than 10 detected proteins were removed using *Scanpy*. To ensure a consistent feature space across sections, we retained only proteins detected in all sections, resulting in 30 shared proteins. The retained protein expression matrix was normalized using centered log-ratio (CLR) transformation to account for compositional bias and variation in total antibody abundance. Since protein measurements were available across multiple sections, the normalized protein matrix was concatenated across sections and reduced by principal component analysis (PCA) to 29 dimensions. Harmony was then applied to the PCA embedding to mitigate section-specific technical variation. The resulting batch-corrected protein representation was used for protein-specific graph construction.

Epigenomic modality preprocessing For the mouse brain dataset, epigenomic modalities, including ATAC-seq, H3K27ac, and H3K27me3 profiles, were processed using the R package *ArchR* (Granja et al., 2021). Specifically, raw fragment files were converted into ArchR Arrow

files using the mm10 reference genome. For each epigenomic modality, gene activity score matrices were generated using ArchR's GeneScoreMatrix, which summarizes signals around gene bodies and nearby regulatory regions to obtain gene-level regulatory activity profiles. The resulting gene activity matrices were log-normalized, followed by selection of 5,000 highly variable genes and PCA retaining the top 50 dimensions. The resulting PC embeddings were used for epigenomic-specific graph construction.

C.3 Evaluation Metrics

Standard evaluation metrics. We use the local inverse Simpson's index (LISI) (Korsunsky et al., 2019) and silhouette coefficient (SC) (Řezanková, 2018) to evaluate local separation and domain purity, and the percentage of abnormal spots (PAS) and CHAOS (Shang & Zhou, 2022) to evaluate spatial coherence. Adjusted Rand index (ARI) and normalized mutual information (NMI) assess global domain recovery after clustering the inferred composition profiles into the same number of domains as the annotations. Because PAS and CHAOS primarily measure spatial smoothness and may be artificially improved by excessive smoothing, we interpret them jointly with SMS, BEAR, ARI, and NMI, which assess whether the inferred composition patterns preserve biologically meaningful spatial structure.

Spatial map similarity Spatial map similarity (SMS) quantifies how well the inferred dominant cell-type distributions recapitulate the spatial organization of annotated domains. To reduce the influence of spot-level variability, we first construct smoothed spatial representations for both inferred and reference maps by applying a Gaussian kernel over neighboring spots. We then compare the resulting spatial patterns using normalized cosine similarity, which measures their directional agreement while being insensitive to differences in magnitude.

Since the inferred cell-type components and annotated spatial domains may differ in their numbers and identities, we establish correspondences between them through an optimal bipartite matching procedure based on the Hungarian algorithm (Wright, 1990). SMS is calculated as the mean cosine similarity over the matched pairs. A larger SMS value indicates a stronger agreement between the inferred cell-type spatial organization and the annotated biological structures.

Balanced edge agreement ratio Balanced edge agreement ratio (BEAR) evaluates whether the local spatial organization of inferred cell-type compositions agrees with manually curated spatial domains. Let $\mathcal{C}$ denote the set of annotated spatial domain labels, and let $\mathcal{K}$ denote the dominant cell types inferred by a deconvolution method. A spatial k-nearest-neighbor graph is constructed over spots, and BEAR compares the consistency between annotated domain relationships and inferred dominant cell type relationships.

Specifically, edges in the spatial graph are divided into two groups: edges connecting spots with the same annotated domain in $\mathcal{C}$ and edges connecting spots with different annotated domains. For within-domain edges, BEAR measures the proportion of edges whose two endpoints are assigned to the same inferred dominant cell type in $\mathcal{K}$. For between-domain edges, it measures the proportion of edges whose endpoints are assigned to different inferred dominant cell types. The final BEAR score is the average of these two quantities, giving equal weight to spatial continuity within domains and separation between domains.

This edge-based formulation is analogous to balanced accuracy and avoids favoring methods that exploit the dominance of within-domain edges in spatial datasets (Brodersen et al., 2010). Notably, assigning all spots to a single dominant cell type yields a BEAR score of 0.5, indicating that values above this baseline reflect recovery of meaningful spatial organization and domain boundaries.

C.4 Additional Results with RNA Available in All Sections

Figures S3–S5 compare the cell-type compositions estimated by ARCADE and five competing methods with manual annotations for six representative cell types across the three mouse brain sections. Overall, ARCADE more accurately recovers cell-type compositions that align with the manually annotated spatial structures, whereas the competing methods exhibit missed detections, abundance underestimation, or spatial misassignment.

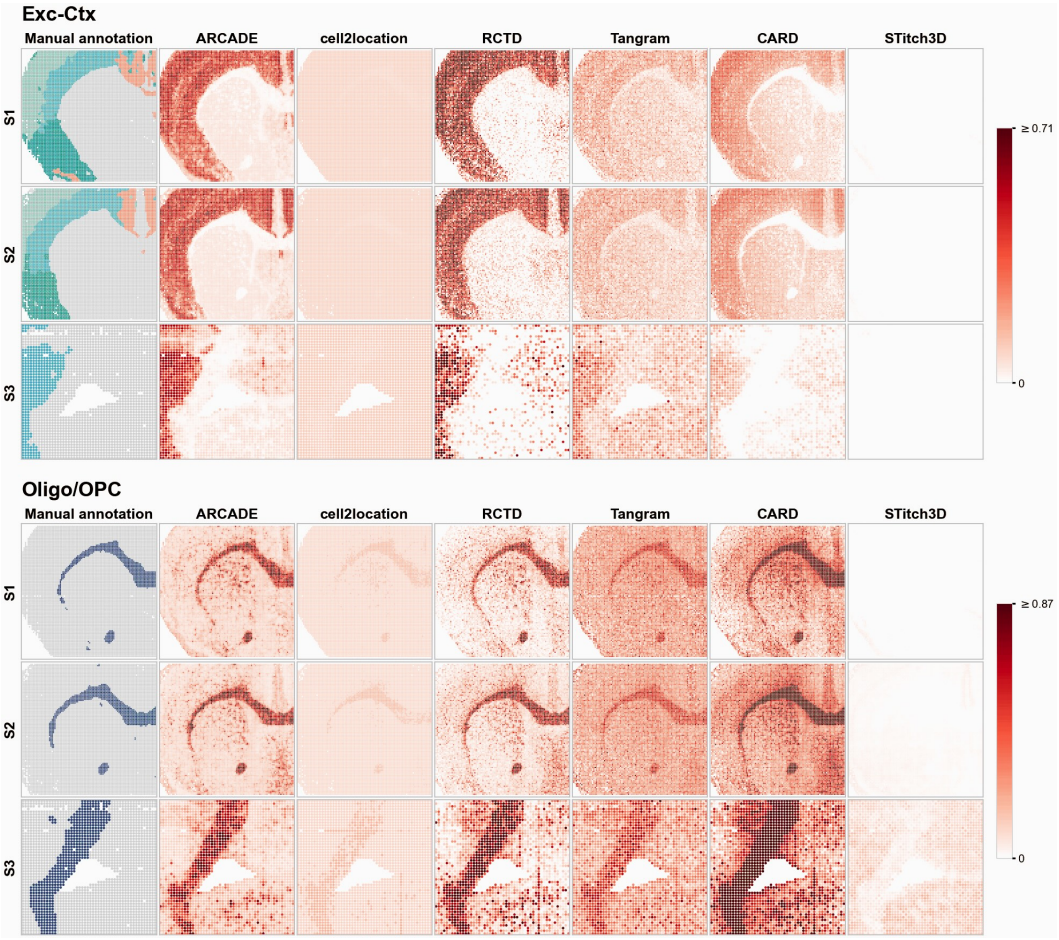


Figure S3: Estimated compositions of Exc-Ctx (excitatory cortical neurons; top panel) and Oligo/OPC (oligodendrocytes and oligodendrocyte precursor cells; bottom panel) for three mouse brain sections. Within each panel, rows are the three sections and columns are the manual annotation followed by ARCADE and the five competing methods. In the annotation column only the domains corresponding to that cell type are coloured (Exc-Ctx: L1–3, L5, L6a/b, PIR; Oligo/OPC: ccg/aco); all other domains are grey. The upper panel shows that cell2location and STitch3D fail to detect Exc-Ctx cells, while Tangram and CARD tend to underestimate their abundance. The lower panel shows that STitch3D fails to detect Oligo/OPC cells, whereas cell2location tends to underestimate their abundance.

Specifically, in Figure S3, cell2location and STitch3D fail to detect Exc-Ctx cells, while Tangram and CARD tend to underestimate their abundance. For Oligo/OPC cells, STitch3D fails to detect the population, whereas cell2location underestimates its abundance. In Figure S4, cell2location fails to detect Inh-Meis2 cells, while STitch3D incorrectly assigns Oligo/OPC cells to Inh-Meis2 in Section 1 and fails to detect Inh-Meis2 cells in the last two sections. Tangram and CARD tend to underestimate Inh-Meis2 abundance. For Nb cells, cell2location fails to detect the population, CARD and STitch3D incorrectly assign other cell types to Nb, Tangram underestimates its abundance, and RCTD produces noisy spatial predictions. In Figure S5, cell2location and CARD fail to identify Inh-General cells, while STitch3D incorrectly assigns other cell types to Inh-General; Tangram and RCTD produce noisy and poorly localized spatial patterns. Finally, for Endo cells, cell2location, RCTD, and Tangram fail to detect the population, whereas CARD and STitch3D incorrectly assign other cell types to Endo.

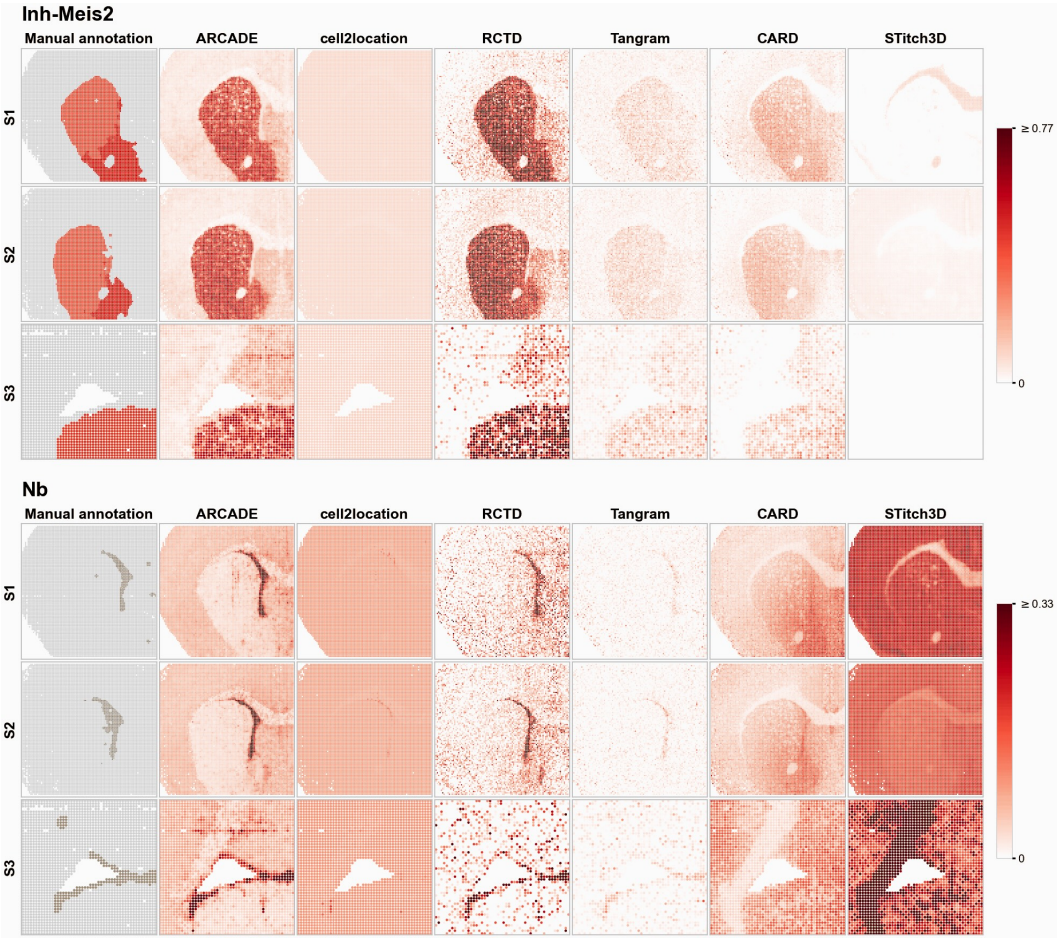


Figure S4: Estimated compositions of Inh-Meis2 (Meis2-positive inhibitory neurons, corresponding to annotated domains CP and ACB; top panel) and Nb (neuroblasts corresponding to annotated domain VL; bottom panel) on the mouse brain. Conventions as in Figure S3. The top panel shows that cell2location fails to detect Inh-Meis2 cells. STitch3D incorrectly identifies Oligo/OPC cells as Inh-Meis2 in Section 1 and fails to detect Inh-Meis2 cells in the last two sections, while Tangram and CARD tend to underestimate their abundance. The bottom panel shows that cell2location fails to detect Nb cells, whereas CARD and STitch3D incorrectly identify other cell types as Nb cells. Tangram underestimates Nb cell abundance, while RCTD produces numerous noisy predictions.

C.5 Additional Results for the RNA-Missing Section

Per-section results on the human tonsil. The main text reports the average performance across the three RNA-containing sections (Table 2). Here, we further provide section-wise results under the same evaluation protocol. As shown in Table S1, ARCADE achieves the best performance on at least six of the eight metrics in each of the three sections, including seven metrics in Section 1, while ranking among the top three methods on the remaining metrics, demonstrating that the observed improvement is consistent across sections rather than attributable to averaging effects.

Results of other mosaic representation transfer-based strategy To investigate that representation-only transfer is not specific to a particular representation learner or prediction model, we replace SpaMosaic with another spatial mosaic integration method SpaMode, and replace kNN transfer with neural network or random forest predictors. The results are summarized in Tables S2 and S3. Across both representations, ARCADE remains the best on seven of the eight metrics under SpaMosaic and on all eight under SpaMode. The strongest transfer variant reaches an ARI of 0.151 with SpaMosaic and 0.039 with SpaMode, against 0.299 for joint inference. No predictor is con-

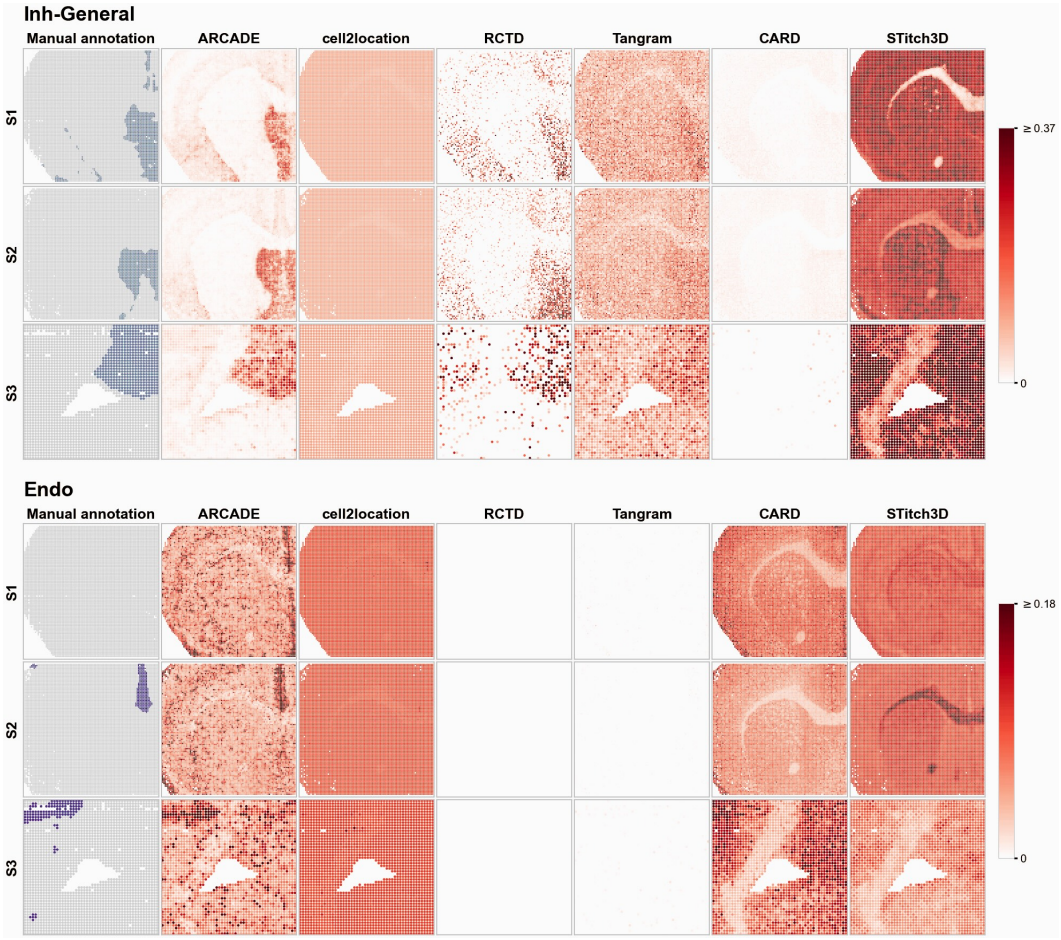


Figure S5: Estimated compositions of Inh-General (general inhibitory neurons corresponding to annotated domain ls; top panel) and Endo (endothelial cells; bottom panel) for three mouse brain sections. Conventions as in Figure S3. The top panel shows that cell2location and CARD fail to identify Inh-General cells, while STitch3D incorrectly identifies other cell types as Inh-General. Tangram and RCTD produce noisy and poorly localized spatial patterns. The bottom panel shows that cell2location, RCTD, and Tangram fail to detect Endo cells, whereas CARD and STitch3D incorrectly identify other cell types as Endo cells.

sistently better than the others. The random-forest variants attain the smoothest spatial scores while their ARI stays near zero, indicating that smoothness alone does not recover the annotated domains.

C.6 Additional Sensitivity Analysis

We analyse four graph-construction hyperparameters on the human tonsil dataset: the neighbourhood sizes k_{feat} and k_{mnn}, and the fusion weights w_f and w_g. We first evaluate the sensitivity to neighbourhood sizes by sweeping $k_{\text{feat}} \in \{6, 10, 15, 20, 30\}$ and $k_{\text{mnn}} \in \{3, 5, 10, 15, 20\}$ (Table S4). The performance variation across these settings is consistently small for all evaluation metrics, indicating that ARCADE is robust to moderate changes in neighbourhood selection. We therefore use fixed neighbourhood sizes across datasets.

In previous experiments, we used $(w_f, w_g) = (0.2, 0.2)$ for the mouse brain dataset and $(w_f, w_g) = (0.4, 0.2)$ for the human tonsil dataset. The following sensitivity analysis investigates how these predefined settings relate to the overall parameter landscape and examines the relative contributions of the two graph components. Specifically, we investigate the sensitivity of ARCADE to the choices of (w_f, w_g) by performing a grid search over $\{0, 0.2, 0.4, 0.6, 0.8\} \times \{0, 0.2, 0.4, 0.6, 0.8\}$ for each

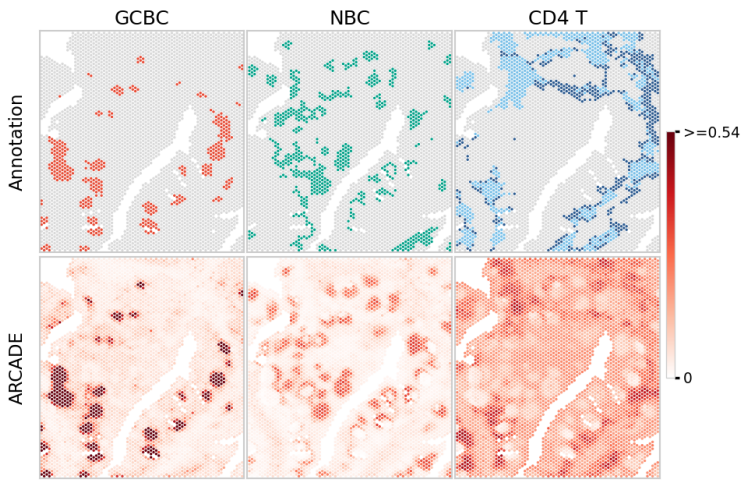


Figure S6: Spatial distributions of germinal center B cells (GCBC), naive B cells (NBC), and CD4 T cells in Section 4 of the human tonsil dataset. The upper panels show manual annotations, and the lower panels show ARCADE-estimated cell-type compositions.

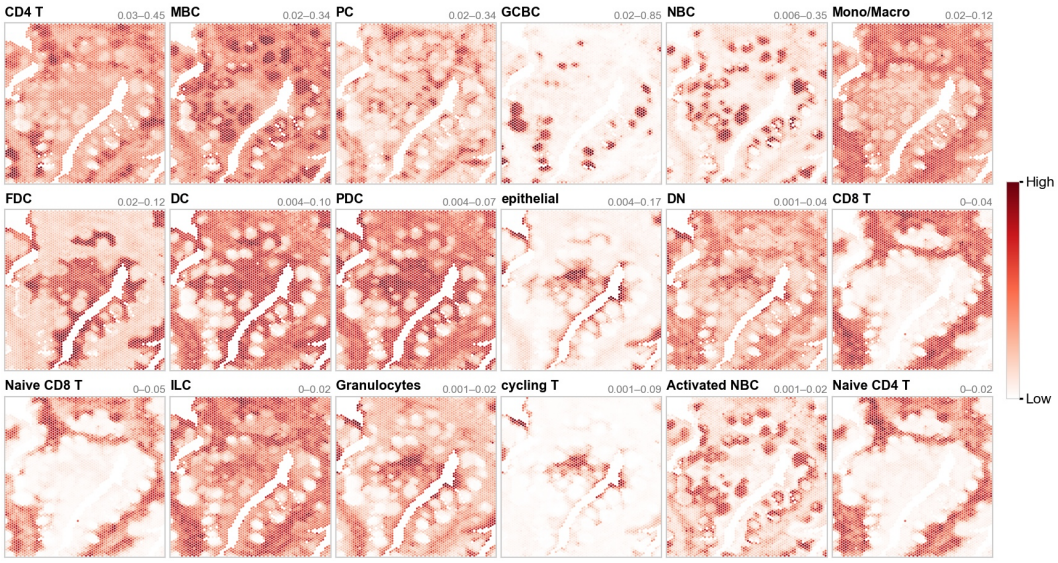


Figure S7: Spatial distribution of cell-type compositions estimated by ARCADE for Section 4 of the human tonsil dataset, where RNA measurements are unavailable. Eighteen of the nineteen cell types are shown, ordered by their mean estimated proportions; NK cells, the least abundant cell type, are omitted. To better visualize the spatial distribution of each cell type, each panel is independently scaled to the full range of estimated proportions for that cell type, as indicated above the panel. Consequently, color intensities are not directly comparable across panels.

dataset. All 25 configurations are trained from scratch, and their performance is evaluated using eight metrics. Figures S10 and S11 show the resulting metric surfaces for the mouse brain and human tonsil datasets, respectively. This analysis aims to characterize the influence of the graph construction weights in ARCADE and to understand the relative contributions of feature similarity and cross-section correspondence. Due to the distinct biological organizations of the mouse brain and human tonsil tissues, the two weights exhibit different effects across these datasets.

On the mouse brain dataset, the performance of ARCADE is primarily influenced by w_f, whereas w_g has a relatively limited effect. For example, when fixing $w_g = 0.2$, the LISI score changes from 1.617 at $w_f = 0.2$ to 1.832 at $w_f = 0.8$; in contrast, when fixing $w_f = 0.2$, the LISI score varies

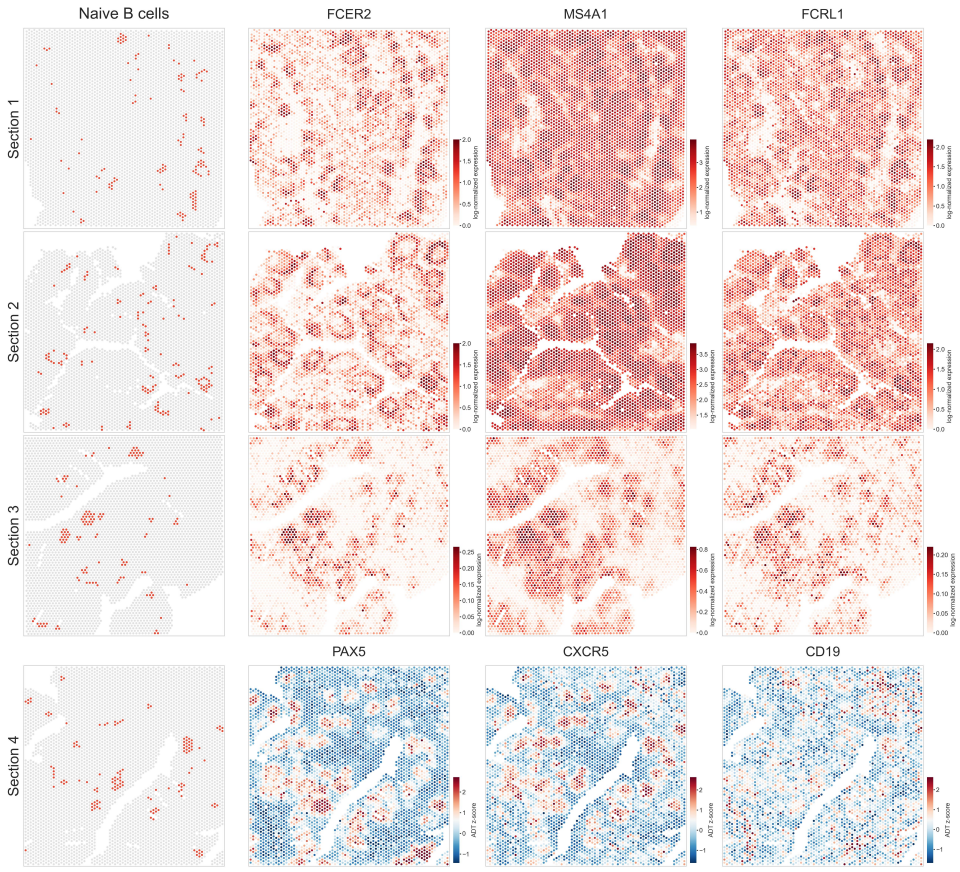


Figure S8: Spatial marker support for ARCADE-inferred naive B cell (NBC)-dominant spots of four human tonsil sections. Rows correspond to Sections 1–4. The left column highlights spots for which NBC has the largest ARCADE-inferred cell-type composition in red, with all remaining spots shown in grey. For the RNA-containing Sections 1–3, the other columns show RNA expression levels of genes *FCER2*, *MS4A1*, and *FCRL1*. For RNA-missing Section 4, the other columns show ADT z-scores for proteins PAX5, CXCR5, and CD19. The spatial correspondence between NBC-dominant spots and these naive-B-associated or B-cell identity markers provides additional molecular support for the inferred NBC regions.

only slightly from 1.619 to 1.639 across different values of w_g. These results suggest that within-section feature similarity plays a more important role than cross-section correspondence for the mouse brain dataset. On the human tonsil dataset, both weights have comparable effects, resulting in relatively flatter metric surfaces. The only clearly suboptimal setting is $(w_f, w_g) = (0, 0)$, which removes both the within-section feature graph and the cross-section correspondence graph, leading to degraded performance. This result highlights the importance of incorporating both molecular feature similarity and cross-section correspondence for this dataset. Overall, these findings indicate that cross-section correspondence and within-section molecular similarity play complementary and comparably important roles in the human tonsil dataset.

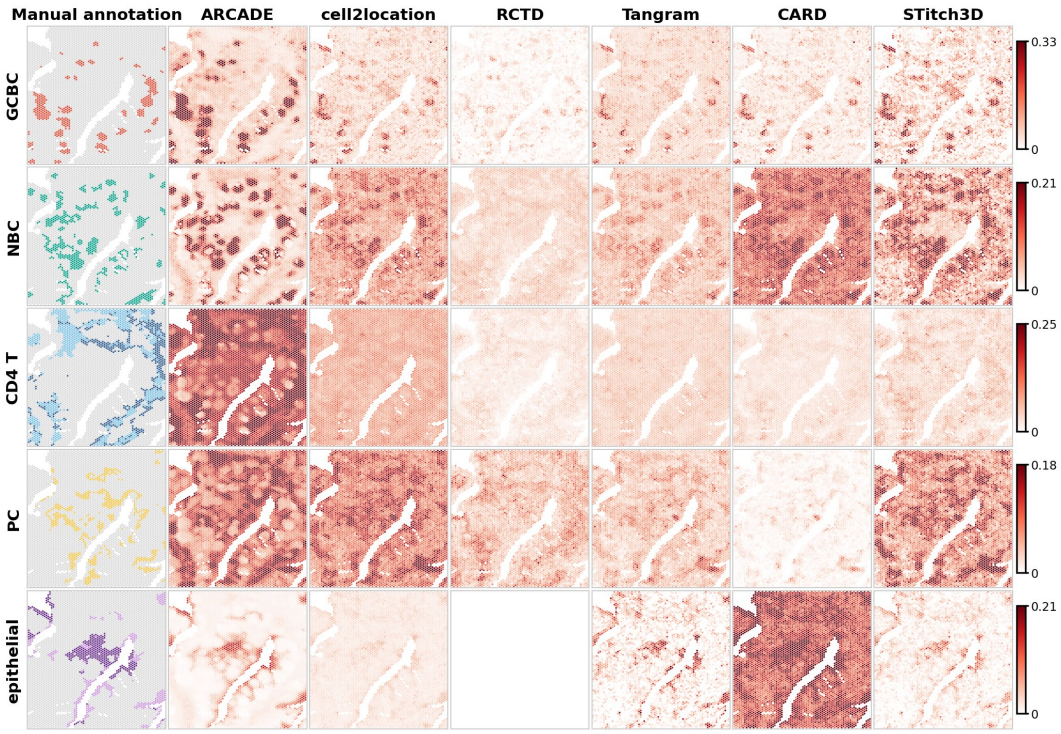


Figure S9: Estimated compositions of five representative cell types on section 4 of the human tonsil, the section for which RNA is unavailable. Rows are cell types and columns are the manual annotation followed by ARCADE and the five transfer-based baselines. In the annotation column only the domains assigned to that cell type are coloured (GCBC: germinal centre; NBC: mantle zone; CD4 T: the two T-cell zones; PC: subepithelial plasma cell; epithelial: epithelium and inflamed epithelium); all other domains are grey, and domain colours follow Figure S2. All method panels within a row share one colour scale, whose upper limit is the 99th percentile of the pooled estimates scaled by 0.75 and is shown at the right; scales are therefore not comparable across rows. RCTD does not report an epithelial component and its panel in the bottom row is left blank.

Table S1: Per-section performance across the three RNA-containing sections of the human tonsil dataset. Best values are shown in bold and second-best values are underlined.

Section	Method	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
S1	Tangram	2.767	0.599	0.343	0.652	0.628	−0.012	0.186	0.305
	RCTD	1.903	0.389	0.226	0.703	0.623	0.043	0.238	0.430
	cell2location	1.887	0.358	0.221	0.630	0.654	0.059	0.281	0.448
	CARD	2.069	0.417	0.251	0.669	0.619	0.007	0.234	0.413
	STitch3D	1.818	0.357	0.225	**0.708**	0.641	0.066	0.306	0.463
	ARCADE	**1.740**	**0.306**	**0.199**	0.675	**0.667**	**0.089**	**0.339**	**0.504**
S2	Tangram	2.873	0.520	0.290	0.631	0.645	−0.005	0.187	0.293
	RCTD	2.074	0.344	0.203	**0.675**	0.639	0.086	0.273	0.429
	cell2location	2.036	0.323	0.200	0.607	**0.689**	0.083	0.286	0.450
	CARD	2.245	0.381	0.238	0.628	0.601	0.046	0.260	0.418
	STitch3D	1.990	0.305	0.194	0.668	0.680	0.098	0.287	0.427
	ARCADE	**1.933**	**0.281**	**0.183**	0.642	0.669	**0.111**	**0.336**	**0.487**
S3	Tangram	3.454	0.660	0.391	0.651	0.616	−0.003	0.157	0.259
	RCTD	2.211	0.403	0.248	0.683	0.611	0.033	0.273	0.414
	cell2location	2.084	0.378	0.251	0.616	**0.634**	0.022	0.212	0.390
	CARD	2.427	0.450	0.272	0.650	0.574	0.016	0.229	0.380
	STitch3D	2.217	0.398	0.262	**0.684**	0.623	0.041	0.278	0.400
	ARCADE	**1.917**	**0.329**	**0.220**	0.675	0.632	**0.074**	**0.287**	**0.472**

Table S2: Representation-only transfer on section 4 of the human tonsil, which carries protein only. SpaMosaic learns a shared cross-section representation; three predictors (kNN, MLP, random forest) are then trained on the composition labels of sections 1–3 and applied to section 4. Labels are taken in turn from the five deconvolution baselines, are used without prior normalisation, and the transferred predictions are row-normalised afterwards. All three predictors receive the raw embedding, so that the only difference between them is the predictor itself. Arrows indicate the favourable direction. Best values in bold, second best underlined.

Labels	Predictor	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
ARCADE(joint)		**1.850**	**0.316**	**0.210**	**0.709**	0.639	**0.089**	**0.299**	**0.477**
Tangram	kNN	3.758	0.607	0.330	0.565	0.634	−0.020	0.044	0.134
	MLP	3.745	0.605	0.327	0.532	0.619	−0.041	0.040	0.094
	RF	2.359	0.419	0.250	0.568	**0.647**	0.002	−0.000	0.068
RCTD	kNN	3.254	0.547	0.316	0.600	0.519	−0.019	0.087	0.196
	MLP	4.219	0.638	0.370	0.548	0.541	−0.063	0.037	0.105
	RF	2.499	0.435	0.270	0.600	0.528	0.005	0.073	0.227
cell2location	kNN	3.400	0.558	0.341	0.532	0.608	−0.033	0.151	0.224
	MLP	4.223	0.621	0.368	0.508	0.592	−0.087	0.061	0.124
	RF	2.718	0.460	0.293	0.527	0.617	−0.004	0.110	0.243
CARD	kNN	3.555	0.593	0.353	0.572	0.545	−0.051	0.081	0.174
	MLP	4.418	0.664	0.387	0.530	0.532	−0.069	0.027	0.091
	RF	2.713	0.467	0.285	0.566	0.537	−0.034	0.068	0.200
STitch3D	kNN	3.170	0.544	0.324	0.615	0.566	−0.027	0.039	0.190
	MLP	4.049	0.639	0.374	0.571	0.569	−0.070	0.027	0.127
	RF	2.448	0.432	0.266	0.610	0.565	0.001	0.028	0.198

Table S3: Representation-only transfer on section 4 of the human tonsil using SpaMode as the mosaic integration method. The setup is identical to Table S2 except that the shared cross-section representation comes from SpaMode instead of SpaMosaic. Arrows indicate the favourable direction. Best values in bold, second best underlined.

Labels	Predictor	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
ARCADE(joint)		**1.850**	**0.316**	**0.210**	**0.709**	**0.639**	**0.089**	**0.299**	**0.477**
Tangram	kNN	5.196	0.782	0.475	0.529	0.603	−0.035	0.012	0.089
	MLP	5.626	0.815	0.482	0.510	0.596	−0.045	0.012	0.031
	RF	5.059	0.775	0.452	0.477	0.536	−0.074	−0.000	0.005
RCTD	kNN	5.084	0.772	0.468	0.568	0.518	−0.061	0.035	0.124
	MLP	5.597	0.842	0.510	0.523	0.521	−0.065	0.015	0.052
	RF	4.660	0.743	0.443	0.496	0.494	−0.091	0.003	0.042
cell2location	kNN	5.135	0.775	0.472	0.504	0.602	−0.075	0.034	0.124
	MLP	5.556	0.870	0.547	0.476	0.585	−0.093	0.018	0.043
	RF	4.830	0.755	0.471	0.463	0.573	−0.113	0.005	0.058
CARD	kNN	5.139	0.745	0.458	0.528	0.544	−0.072	0.039	0.119
	MLP	6.026	0.850	0.520	0.471	0.535	−0.094	0.006	0.033
	RF	4.944	0.697	0.401	0.472	0.521	−0.137	0.003	0.039
STitch3D	kNN	5.004	0.768	0.476	0.552	0.553	−0.062	0.026	0.139
	MLP	5.555	0.830	0.511	0.508	0.544	−0.069	0.015	0.052
	RF	4.782	0.743	0.439	0.484	0.495	−0.105	0.002	0.038

Table S4: Sensitivity to the neighbourhood sizes used in graph construction, evaluated on the human tonsil. Each value is trained from scratch; metrics are computed per section and averaged over the four sections. The "Range" row gives the spread of each metric across the five settings. Defaults are $k_{\text{feat}} = 15$ and $k_{\text{mnn}} = 10$.

	LISI ↓	PAS ↓	CHAOS ↓	SMS ↑	BEAR ↑	SC ↑	ARI ↑	NMI ↑
Feature graph, k_{feat}								
6	1.861	0.307	0.203	0.675	0.645	0.092	0.312	0.482
10	1.856	0.304	0.203	0.675	0.649	0.092	0.315	0.484
15 (DEFAULT)	1.860	0.308	0.202	0.675	0.652	0.091	0.315	0.485
20	1.865	0.308	0.202	0.675	0.652	0.090	0.316	0.484
30	1.873	0.310	0.204	0.675	0.652	0.089	0.324	0.501
Range	0.016	0.006	0.002	0.001	0.007	0.003	0.012	0.019
Cross-section MNN graph, k_{mnn}								
3	1.851	0.306	0.202	0.675	0.651	0.093	0.313	0.484
5	1.854	0.306	0.202	0.675	0.652	0.092	0.315	0.485
10 (DEFAULT)	1.860	0.308	0.202	0.675	0.652	0.091	0.315	0.485
15	1.867	0.307	0.203	0.675	0.652	0.090	0.315	0.484
20	1.875	0.308	0.205	0.675	0.652	0.089	0.316	0.484
Range	0.025	0.002	0.003	0.000	0.001	0.004	0.003	0.001

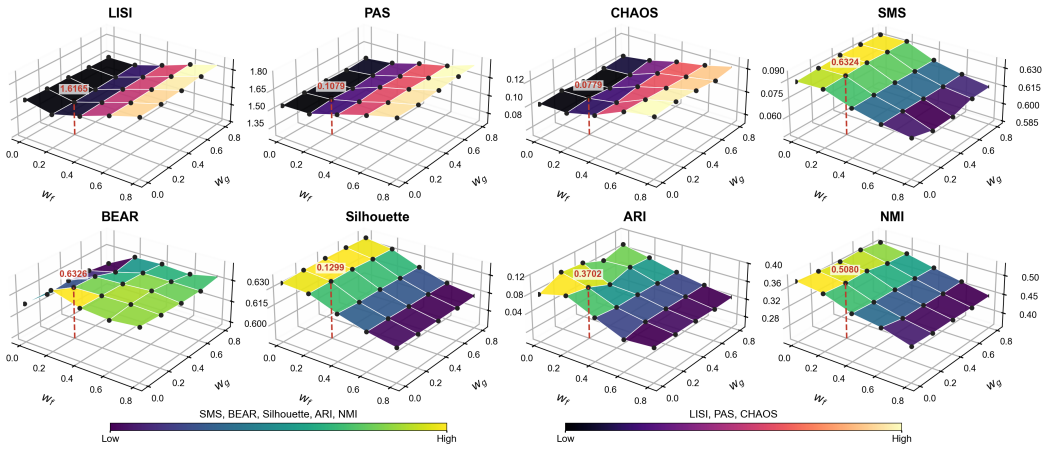


Figure S10: All eight evaluation metrics over the 5×5 grid of fusion weights (w_f, w_g) on the mouse brain. Black dots are the 25 measured configurations and the red marker $(w_f, w_g) = (0.2, 0.2)$ is the setting used for ARCADE in the preceding experiments. Note that lower values indicate better performance for LISI, PAS, and CHAOS, whereas higher values indicate better performance for SMS, BEAR, SC, ARI, and NMI.

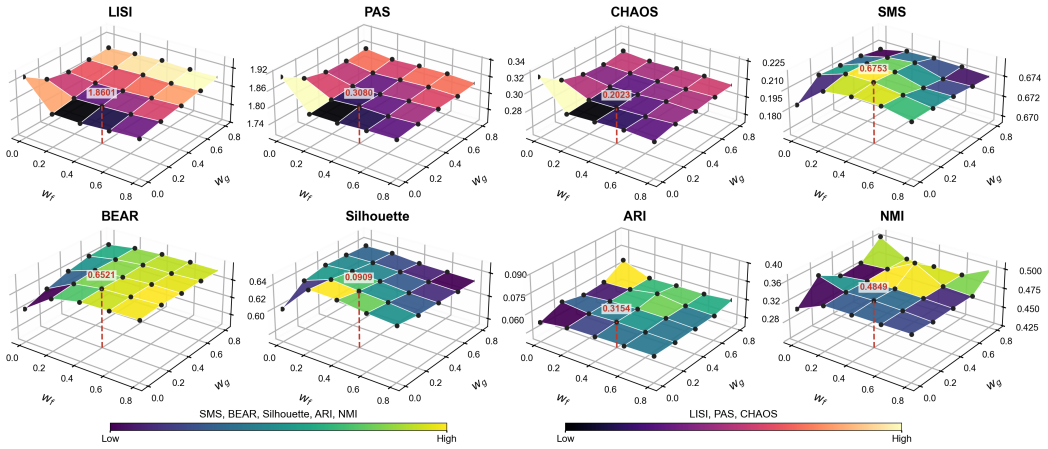


Figure S11: All eight evaluation metrics over the 5×5 grid of fusion weights (w_f, w_g) on the human tonsil. Black dots are the 25 measured configurations and the red marker $(w_f, w_g) = (0.4, 0.2)$ is the setting used for ARCADE in the preceding experiments. Note that lower values indicate better performance for LISI, PAS, and CHAOS, whereas higher values indicate better performance for SMS, BEAR, SC, ARI, and NMI.