DESPOTX: IDENTIFIABILITY-BASED DECONTAMINATION FOR SPATIAL TRANSCRIPTOMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial transcriptomics (ST) at single-cell resolution profiles gene expression in its native spatial context, but a substantial fraction of transcripts contaminate neighboring cells, compromising downstream biological analyses. Existing decontamination methods either ignore the spatial structure of contamination or aggregate over neighbors without constraining how contamination is separated from native expression, leaving the decomposition ambiguous. To resolve this ambiguity, we introduce DeSpotX, a deep generative model that uses anchor genes, defined as genes not natively expressed in a given cell cluster, to constrain the contamination decomposition. DeSpotX further uses spatial information to estimate contamination locally through a cluster-masked, distance-weighted average over neighboring cells. On spike-in simulations across five datasets and four ST platforms, DeSpotX attains the highest AUROC in 27 of 30 benchmark settings, exceeding the best baseline by up to 0.12, and remains robust to inaccuracies in the cell-cluster annotation and in anchor gene construction. Using matched scRNA-seq as an independent reference for cell-type-specific expression, DeSpotX retains 94% of within-type signal while improving marker-gene specificity. Decontaminated expression also shows greater spatial coherence and yields cell-cell communication networks consistent with known biology. Finally, iterating between decontamination and cell-cluster annotation refines these outcomes, reassigning ligand-receptor signaling to expected source cells in mouse brain and breast cancer tissues.

1 INTRODUCTION

Single-cell-resolution spatial transcriptomics (ST) platforms, such as Xenium (Janesick et al., 2023), MERFISH (Chen et al., 2015; Zhang et al., 2023), CosMx (He et al., 2022), and Stereo-seq (Chen et al., 2022), now profile millions of cells in tissue, enabling the study of cellular organization, cell-cell communication, and disease microenvironments at single-cell resolution (Moffitt et al., 2022; Palla et al., 2022). However, these technologies suffer from a systematic contamination artifact, where transcripts from one cell are frequently assigned to neighboring cells through ambient diffusion, segmentation errors, and vertical overlap of cells in the tissue slice (Ergen & Yosef, 2025; Kwok et al., 2025; Chen et al., 2026). The contamination is widespread, with 20-40% of cells showing substantial contamination and 10-20% of cells expressing markers of mutually exclusive cell types according to recent studies (Ergen & Yosef, 2025; Kwok et al., 2025). These artifacts compromise downstream biological analyses, causing misannotation of cell types, blurring the spatial organization of gene expression, and producing spurious cell-cell communication networks that can mislead the discovery of therapeutic targets.

Decontamination of ST data poses two fundamental challenges. First, the native and contamination components cannot be uniquely separated from observed counts. The same observed count distribution can arise from multiple combinations of native expression, contamination profile, and contamination fraction, so methods fail to identify which combination is correct without additional constraints on the decomposition (Khemakhem et al., 2020). Second, contamination in ST is spatially structured, since it diffuses locally from nearby cells and the contaminating signal reflects the cellular composition of each cell's local environment. A single global ambient profile, as used by methods designed for dissociated single-cell RNA-seq (Yang et al., 2020; Young & Behjati, 2020; Fleming et al., 2023), fails to capture this local structure.

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.

To address these challenges, we present DeSpotX, an identifiability-based deep generative model for decontaminating single-cell-resolution ST data. For the *non-identifiability* challenge, DeSpotX introduces anchor genes, defined as genes that are not natively expressed in a given cell cluster and inferred automatically from per-cluster expression rates, and uses them to constrain the decomposition. Identifiability holds when the contamination profile is known; when it is estimated, we bound the error from within-cluster contamination. For the *spatial structure* challenge, DeSpotX estimates the local contamination profile as a cluster-masked, distance-weighted average over cross-cluster spatial neighbors, separating spatially diffused contamination from each cell's native expression.

Across five spatial transcriptomics datasets spanning four platforms and four tissue contexts, DeSpotX outperforms all leading decontamination baselines and maintains strong performance even when cell-cluster annotations or anchor gene construction are imperfect. Downstream analyses on the decontaminated counts yield cleaner cluster separation, more spatially coherent marker expression, and more biologically coherent cell-cell communication networks. We further show that iterating between decontamination and cell-cluster annotation progressively improves these outcomes.

2 Related work

Decontamination of single-cell RNA-seq. Several methods correct ambient RNA contamination in dissociated single-cell RNA-seq. SoupX (Young & Behjati, 2020) estimates a global ambient profile from empty droplets and uses non-expressed genes, selected manually or automatically from cluster markers, to estimate a contamination fraction that is typically shared across cells. DecontX (Yang et al., 2020) fits a Bayesian mixture model that decomposes each cell's expression into its native cluster profile and a global contamination profile aggregated across clusters, recovering per-cell contamination fractions by variational inference. CellBender (Fleming et al., 2023) models raw barcode-level counts with a learned ambient prior estimated from empty droplets, separating ambient transcripts from cell-derived expression. These methods rely on a single global ambient profile and do not incorporate spatial information, making them unable to capture the spatial structure of contamination in single-cell-resolution ST.

Decontamination of spatial transcriptomics. Recent methods adapt count correction to ST data by incorporating spatial information into the contamination estimate. SpaceBender (Chen et al., 2026) replaces CellBender's global ambient profile with a spatially local profile computed by averaging expression over k-nearest-neighbor or distance-based spatial neighbors. DenoIST (Kwok et al., 2025) fits a Poisson mixture that classifies each gene count as endogenous or contamination based on the cell's count relative to its local neighborhood. ResolVI (Ergen & Yosef, 2025) decomposes observed counts into native expression, neighbor-derived diffusion, and non-specific background. While these methods leverage spatial structure, none constrain the separation of spatially diffused contamination from native expression using genes not natively expressed in each cell's cluster. Moreover, all of these methods regularize the decomposition through Bayesian priors, latent-variable architectures, or mixture-model assumptions, none of which resolve the underlying non-identifiability. These regularizers can produce decompositions that fit the observed counts, but the recovered decompositions can deviate from the true profile. A detailed comparison between DeSpotX and existing methods is provided in Appendix A.

Identifiability of latent variable models. Identifiability is a long-standing concern in latent variable models, where multiple parameter combinations can produce the same observed data distribution (Khemakhem et al., 2020). Two main strategies resolve non-identifiability. The first uses auxiliary variables to make latent dimensions conditionally independent, as in identifiable VAEs (Khemakhem et al., 2020), but no such auxiliary variable is available in the decontamination setting. The second imposes structural constraints on the model, such as anchor-word identifiability in topic models (Arora et al., 2013) and separability in non-negative matrix factorization (Donoho & Stodden, 2003), and fits the decontamination problem naturally, since each cell cluster expresses only a subset of genes. DeSpotX builds on this strategy, proving that the contamination decomposition is non-identifiable from observed counts alone and deriving an anchor-gene constraint that restores identifiability in the idealized setting, with anchors used as per-cluster constraints on each cell's decomposition rather than only to estimate a shared contamination fraction as in SoupX.

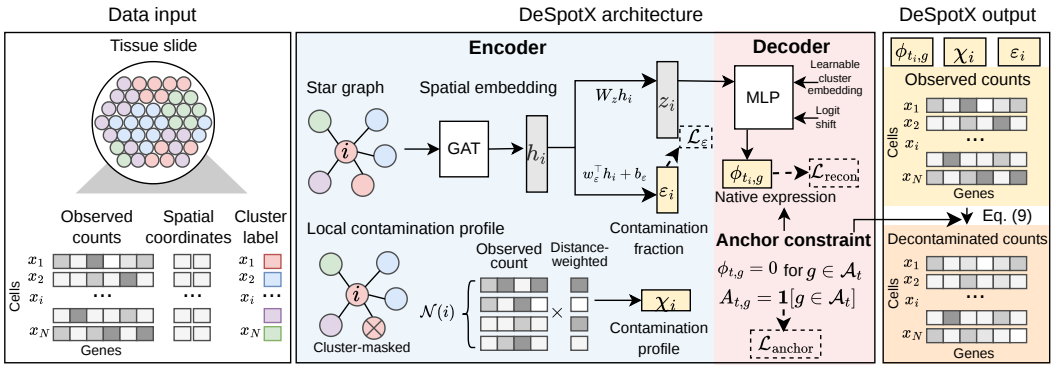

Figure 1: Overview of DeSpotX for decontaminating spatial transcriptomics data.

3 METHODS

3.1 PROBLEM FORMULATION

Consider a single-cell-resolution spatial transcriptomics dataset of N cells indexed by $i \in \{1, \dots, N\}$. For each cell i, we observe a count vector $x_i \in \mathbb{Z}_{\geq 0}^{G}$ over G genes, a discrete cluster label $t_i \in \{1, \dots, T\}$, and a spatial coordinate $s_i \in \mathbb{R}^2$. Let $\mathcal{N}(i) \subset \{1, \dots, N\}$ with $|\mathcal{N}(i)| = K$ denote the K nearest spatial neighbors of cell i under the Euclidean metric. We model the observed counts as a mixture of the cell's native transcriptional output and contamination from its local spatial environment:

$$x_{i,g} \;\sim\; \mathrm{NB}(d_i\, m_{i,g},\, \theta_g)\,, \qquad m_i \;=\; (1-\varepsilon_i)\,\phi_{t_i}(z_i) + \varepsilon_i\,\chi_i, \tag{1}$$

where $d_i = \sum_g x_{i,g}$ is the library size, $\theta_g \in \mathbb{R}_{>0}$ is a gene-wise dispersion parameter, $\phi_{t_i}(z_i) \in \Delta^{G-1}$ is the cell's native expression profile over genes conditioned on its cluster index t_i and a latent state $z_i \in \mathbb{R}^{d_z}$, $\chi_i \in \Delta^{G-1}$ is the local contamination profile at location s_i, and $\varepsilon_i \in [0, 1]$ is the per-cell contamination fraction. We adopt the mean-dispersion parameterization throughout, where $\mathrm{NB}(\mu, \theta)$ denotes the negative binomial with mean μ and variance $\mu + \mu^2/\theta$.

The inference task is to recover, for each cell, the native profile $\phi_{t_i}(z_i)$, the local contamination profile χ_i, and the contamination fraction ε_i from the observation x_i, the cluster index t_i, and the spatial context $\{(x_j, t_j, \|s_j - s_i\|) : j \in \mathcal{N}(i)\}$. The likelihood in Eq. 1 depends on $(\varepsilon_i, \phi_{t_i}, \chi_i)$ only through the mixture m_i, making the decomposition non-identifiable.

Lemma 1 (Non-identifiability of the native-contamination decomposition)**.** *For any observed mean $m \in \Delta^{G-1}$, the set of parameters consistent with m,*

$$\mathcal{S}(m) \;=\; \big\{(\varepsilon, \phi, \chi) \in [0,1] \times \Delta^{G-1} \times \Delta^{G-1} : (1-\varepsilon)\phi + \varepsilon\chi = m\big\},$$

contains a continuum of solutions on which the likelihood in Eq. 1 is constant.

A proof is provided in Appendix B.1. Consequently, any decontamination method must impose external constraints on at least one of ϕ, χ, or ε to recover a unique decomposition. DeSpotX imposes such constraints through anchor positions in ϕ and a spatial estimator for χ.

3.2 DESPOTX ARCHITECTURE

DeSpotX instantiates Eq. 1 as a deep generative model with three components (Figure 1). A **spatial graph encoder** maps each cell's spatial context to a latent encoding z_i of its native expression state and an estimate of the contamination fraction $\bar{\varepsilon}_i$, alongside a local contamination profile χ_i aggregated from cross-cluster spatial neighbors. A **cluster-conditioned decoder** maps (z_i, t_i) to the native profile $\phi_{t_i}(z_i)$. **Anchor constraints**, inferred automatically from per-cluster expression rates in the data, promote identifiability of the decomposition by fixing $\phi_{t,g}$ to zero at (t, g) pairs where gene g is not natively expressed in cluster t.

Spatial graph encoder. For each cell i we construct a star graph $\mathcal{G}_i$ in which a single center node, cell i, is connected to K leaf nodes corresponding to the cells in $\mathcal{N}(i)$. Node features are log-transformed counts concatenated with a cluster embedding; edge features are kernel weights $w_{ij} = \exp(-\|s_j - s_i\|/\rho_i)$ with adaptive bandwidth ρ_i set to the median distance from cell i to its K nearest neighbors. A GATv2 layer (Brody et al., 2022) aggregates neighbor information into the center, producing a spatial-context embedding $h_i \in \mathbb{R}^{d_h}$. From h_i, the encoder produces a latent state z_i and a contamination-fraction estimate $\bar{\varepsilon}_i$:

$$z_i = W_z h_i, \qquad \bar{\varepsilon}_i = \sigma\big(w_\varepsilon^\top h_i + b_\varepsilon\big), \tag{2}$$

where W_z, w_ε, and b_ε are learned parameters, and the sigmoid ensures $\bar{\varepsilon}_i \in (0, 1)$. Implementation details are provided in Appendix C.1.

We estimate the local contamination profile χ_i as a cluster-masked, distance-weighted average over $\mathcal{N}(i)$:

$$\chi_i \;=\; \frac{1}{Z_i} \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot \mathbf{1}[t_j \neq t_i] \cdot \frac{x_j}{d_j}, \qquad Z_i = \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot \mathbf{1}[t_j \neq t_i], \tag{3}$$

with fallback to a dataset-level cross-cluster ambient profile when no other-cluster neighbor is available. The mask $\mathbf{1}[t_j \neq t_i]$ restricts χ_i to a cross-cluster aggregation of the local spatial environment, isolating shared contamination from the center cell's own native signal. This masking excludes within-cluster contamination; the resulting error in $\hat{\phi}_i$ is bounded by $\varepsilon_i \cdot \sigma_i^{\text{within}}$, where σ_i^{within} measures within-cluster expression heterogeneity (Appendix B.2, examined in Appendix K).

Cluster-conditioned decoder. The decoder maps the latent state and cluster index to the native expression profile:

$$\phi_{t_i}(z_i) \;=\; \mathrm{softmax}\big(f_\theta(z_i, e_{t_i}) + b_{t_i}\big), \tag{4}$$

where f_θ is a multilayer perceptron, e_{t_i} is a learned cluster embedding, and $b_{t_i} \in \mathbb{R}^G$ is a cluster-specific logit shift initialized from the log-mean expression of cluster t_i. Implementation details are provided in Appendix C.2.

Anchor constraints. By Lemma 1, identifiability requires external constraints. We introduce *anchor genes*: for each cluster t, let $\mathcal{A}_t \subseteq \{1, \ldots, G\}$ denote a set of genes not natively expressed in cluster t, and let $A_{t,g} = \mathbf{1}[g \in \mathcal{A}_t]$. The set $\mathcal{A}_t$ is identified from per-cluster expression rates (Appendix C.3).

Lemma 2 (Identifiability of the decomposition under anchor constraints)**.** *Fix a cell i with cluster index t_i, and assume χ_i is constructed without using x_i. Suppose $\phi_{t,g} = 0$ for all $g \in \mathcal{A}_t$, and that $\chi_{i,g} > 0$ for at least one $g \in \mathcal{A}_{t_i}$. Then ε_i is identifiable from the observations at anchor positions, and ϕ_{t_i} is identifiable on non-anchor positions whenever $\varepsilon_i < 1$.*

A proof is provided in Appendix B.3. Note that the condition requires only a single anchor gene $g \in \mathcal{A}_{t_i}$ with $\chi_{i,g} > 0$.

Training objective. The three components are trained together under

$$\mathcal{L} \;=\; \mathcal{L}_{\text{recon}} + \lambda_A \, \mathcal{L}_{\text{anchor}} + \lambda_E \, \mathcal{L}_\varepsilon, \tag{5}$$

where the reconstruction loss is the per-cell negative log-likelihood under Eq. 1,

$$\mathcal{L}_{\text{recon}} \;=\; -\frac{1}{N} \sum_i \sum_g \log p_{\text{NB}}\big(x_{i,g};\, d_i \cdot m_{i,g},\, \theta_g\big), \quad m_i = (1 - \bar{\varepsilon}_i)\phi_{t_i}(z_i) + \bar{\varepsilon}_i \chi_i, \tag{6}$$

the anchor penalty

$$\mathcal{L}_{\text{anchor}} \;=\; \frac{1}{N} \sum_i \sum_g A_{t_i,g} \cdot (1 - \bar{\varepsilon}_i)\phi_{t_i,g}(z_i) \cdot d_i \tag{7}$$

penalizes nonzero $\phi_{t_i,g}$ at anchor positions, and the contamination-fraction regularizer

$$\mathcal{L}_\varepsilon \;=\; -\frac{1}{N} \sum_i \log \mathrm{Beta}\big(\bar{\varepsilon}_i;\, \mu_\varepsilon \nu,\, (1 - \mu_\varepsilon)\nu\big) \tag{8}$$

pulls $\bar{\varepsilon}_i$ toward a learnable target μ_ε with fixed concentration ν. The anchor weight λ_A is linearly increased from zero over the first E_{warm} epochs, letting the decoder establish reasonable profiles before the constraint is enforced fully. Optimization details are provided in Appendix C.4.

Practical considerations. Lemma 2 assumes a known χ_i and exact anchor zeros. Lemma 4 (Appendix B.4) bounds the recovery error in terms of the errors of χ_i, the anchor zeros, and the model fit, and DeSpotX remains accurate with a single anchor per cluster, with misspecified anchors, and with estimated χ_i (Figure 2; Appendix I).

3.3 Inference

With all components fixed after training, denoised counts are produced as follows. The star graph encoder yields the spatial-context embedding h_i, the latent state z_i, the contamination fraction $\bar{\varepsilon}_i$, and the local contamination profile χ_i. The decoder then maps (z_i, t_i) to the native profile $\phi_{t_i}(z_i)$, and each observation is reweighted by the native component's contribution to the mixture,

$$\hat{x}_{i,g} \;=\; x_{i,g} \cdot \frac{(1-\bar{\varepsilon}_i)\,\phi_{t_i,g}(z_i)}{(1-\bar{\varepsilon}_i)\,\phi_{t_i,g}(z_i) + \bar{\varepsilon}_i \chi_{i,g}}. \tag{9}$$

4 Experiments

4.1 Datasets

Table 1: Single-cell-resolution spatial transcriptomics datasets.

Dataset name	Technology	Tissue	# cells	# genes	# cell clusters
`Xenium_Breast`	Xenium	Breast cancer	118,370	313	20
`MERFISH_Brain`	MERFISH	Mouse brain	115,779	550	29
`CosMx_NSCLC`	CosMx	Non-small cell lung cancer	99,185	980	38
`Xenium5k_Breast`	Xenium 5K Prime	Breast cancer	577,258	5,101	7
`Stereo-seq_E14`	Stereo-seq	Mouse embryo (E14.5)	92,928	18,582	16

We evaluate DeSpotX and baseline methods on five publicly available single-cell-resolution spatial transcriptomics datasets (Table 1). The five datasets span the major imaging- and sequencing-based platforms (Xenium (Janesick et al., 2023), Xenium 5K Prime (10x Genomics, 2024), CosMx (He et al., 2022), MERFISH (Zhang et al., 2023), and Stereo-seq (Chen et al., 2022)), four tissue contexts, and panel sizes ranging from 313 to 18,582 genes. Cell-cluster annotations are taken from the original studies where available; for Xenium5k_Breast, annotations are obtained by CellTypist (Domínguez Conde et al., 2022) label transfer using a published human breast atlas as reference (Reed et al., 2024). All datasets are used as published, without additional quality control or filtering; raw integer counts are provided to all methods, with spatial centroids and cell-cluster annotations additionally provided to methods that use them.

4.2 Spike-in simulations

Spatial transcriptomics data lack ground truth for ambient contamination, so we construct a spike-in benchmark from real datasets.

4.2.1 Simulation design

To reduce overlap with DeSpotX's modeling assumptions, we sample contamination counts from a Poisson likelihood (rather than DeSpotX's negative-binomial) and draw per-cell contamination rates from a Beta distribution (rather than DeSpotX's regularized encoder estimate). For each of the five datasets in Table 1, we sample $40{,}000$ cells and inject synthetic ambient counts at six noise levels $\bar{\varepsilon} \in \{0, 5, 10, 20, 30, 40\}\%$ under six contamination mechanisms. The *local cross-cluster* mechanism matches DeSpotX's assumptions, and the remaining five (*global ambient*, *segmentation spillover*, *long-range diffusion*, *anisotropic diffusion*, and *vertical overlap*) are off-model. The spiked counts $s_{i,g} = n_{i,g} + c_{i,g}$ are provided to each method as input, while the native counts $n_{i,g}$ and per-gene contamination labels $y_{i,g} = \mathbf{1}[c_{i,g} > 0]$ are retained as ground truth. Details are provided in Appendix D.

Table 2: Benchmark comparison on Xenium_Breast and MERFISH_Brain across six contamination mechanisms (one on-model, five off-model). AUROC (mean ± std across the five non-zero contamination levels). Best value per row in bold.

Dataset	Mechanism	**DeSpotX**	SoupX	DecontX	ResolVI	SpaceBender
Xenium_Breast	Local cross-cluster	**0.955 ± 0.008**	0.606 ± 0.069	0.677 ± 0.014	0.851 ± 0.002	0.805 ± 0.009
	Global ambient	**0.897 ± 0.005**	0.804 ± 0.049	0.796 ± 0.037	0.837 ± 0.006	0.720 ± 0.008
	Segmentation spillover	**0.904 ± 0.010**	0.572 ± 0.051	0.633 ± 0.011	0.826 ± 0.004	0.788 ± 0.036
	Long-range diffusion	**0.929 ± 0.004**	0.692 ± 0.019	0.690 ± 0.014	0.840 ± 0.003	0.801 ± 0.006
	Anisotropic diffusion	**0.954 ± 0.007**	0.704 ± 0.021	0.677 ± 0.017	0.829 ± 0.002	0.831 ± 0.007
	Vertical overlap	**0.912 ± 0.006**	0.604 ± 0.075	0.636 ± 0.011	0.827 ± 0.005	0.850 ± 0.002
MERFISH_Brain	Local cross-cluster	**0.948 ± 0.009**	0.631 ± 0.008	0.721 ± 0.013	0.812 ± 0.009	0.824 ± 0.013
	Global ambient	**0.876 ± 0.014**	0.846 ± 0.048	0.805 ± 0.044	0.851 ± 0.024	0.747 ± 0.015
	Segmentation spillover	**0.902 ± 0.010**	0.600 ± 0.004	0.724 ± 0.013	0.850 ± 0.009	0.808 ± 0.036
	Long-range diffusion	**0.921 ± 0.009**	0.637 ± 0.017	0.720 ± 0.015	0.860 ± 0.015	0.842 ± 0.008
	Anisotropic diffusion	**0.947 ± 0.009**	0.635 ± 0.007	0.723 ± 0.013	0.852 ± 0.012	0.854 ± 0.007
	Vertical overlap	**0.906 ± 0.006**	0.613 ± 0.010	0.702 ± 0.013	0.847 ± 0.008	0.859 ± 0.010

4.2.2 BENCHMARK EVALUATION

Baselines. We compare DeSpotX against four published decontamination baselines, including SoupX (Young & Behjati, 2020), DecontX (Yang et al., 2020), ResolVI (Ergen & Yosef, 2025), and SpaceBender (Chen et al., 2026), all run with default settings. CellBender (Fleming et al., 2023) is excluded because it requires empty-droplet observations that are unavailable for the ST platforms, and DenoIST (Kwok et al., 2025) requires transcript coordinates, which the count-level spike-in does not produce, so we include it only in the transcript-level simulation described below. All methods receive the same spiked counts $s_{i,g}$ as input; spatial coordinates and cell-cluster labels are additionally provided to methods that use them. Each method outputs decontaminated counts $\hat{n}_{i,g}$.

Benchmark performance. Table 2 reports AUROC on Xenium_Breast and MERFISH_Brain (full results across all five datasets in Appendix E), measuring how well each method's removed counts predict the binary contamination labels, averaged across the five non-zero noise levels. DeSpotX attains the highest mean AUROC in 27 of 30 benchmark settings and remains above 0.87 throughout; in the other three it trails the best baseline by 0.001 to 0.002. On the local cross-cluster scenario, DeSpotX is uniformly above 0.94, while the next-best method, ResolVI, averages 0.892 across datasets. On the five off-model mechanisms, it outperforms the best baseline by up to 0.12, with the three exceptions differing by at most 0.002, confirming that performance does not depend on the simulator matching DeSpotX's spatial assumptions. We further examine performance separately at anchor and non-anchor positions. On non-anchor entries, AUROC remains 0.841-0.932 across six contamination mechanisms on both datasets (Appendix G, Table 10), supporting the model's ability to detect contamination outside anchor positions. Beyond AUROC, DeSpotX also recovers the magnitude of contamination most accurately, with the lowest relative mean absolute error (relMAE) in 27 of 30 settings (Appendix E). DeSpotX also performs best in a label-free, transcript-level diffusion simulation, with 0.052-0.065 higher AUROC than the best baseline and 0.225-0.545 lower relMAE than applying Lemma 2 directly without the learned model (Appendix F).

Ablation study. We assess each architectural component by removing it from the full model on Xenium_Breast and MERFISH_Brain (full results in Appendix H). Removing the identifiability constraints drops AUROC below 0.90 on MERFISH_Brain and makes it vary from 0.83 to 0.95 across noise levels on Xenium_Breast. Removing spatial information, the GAT layer, or the cluster mask each reduces AUROC, confirming that every component contributes to DeSpotX's accuracy.

Anchor validation. We validate the inferred anchors against scRNA-seq from the same tissue and show that DeSpotX is robust to anchor misspecification. In independent scRNA-seq from the same Xenium_Breast tissue (Janesick et al., 2023), genes marked as anchors have a median detection of only 1.5% in their assigned cell type, compared with 20% for non-anchors (Figure 2a). Additionally, the decomposition of the remaining genes is robust to anchor misspecification. First, adding 20 actively expressed genes (detected in over 15% of cells) per cluster as false anchors removes those genes from the affected clusters but changes AUROC on the remaining genes by less than 0.002, with all-gene AUROC dropping by less than 0.003 (Figure 2b). Second, reducing the anchor set to a single

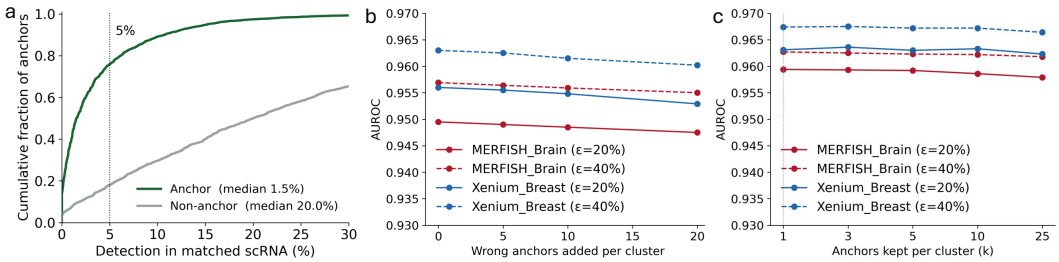

Figure 2: Validation of anchor genes. **a.** Cumulative detection rate of anchor and non-anchor genes in the matched Xenium_Breast scRNA-seq reference. **b.** AUROC after adding deliberately wrong anchors per cluster. **c.** AUROC when retaining only k anchors per cluster.

gene per cluster barely affects AUROC (Figure 2c). This is consistent with Lemma 2, which requires only one valid anchor with nonzero contamination per cluster for identifiability.

Robustness and consistency. DeSpotX is further robust to random perturbation of the cell-cluster labels, the number of spatial neighbors K, and the cell-cluster annotation method (Appendix I). The performance of DeSpotX is consistent across contamination levels and random seeds (Appendix J). The within-cluster contamination bound is examined on spike-in simulations (Appendix K).

4.3 Application to Real Spatial Transcriptomics Data

We apply DeSpotX to the full five datasets in Table 1; it runs much faster than other ST methods (Appendix L). We then assess whether decontamination improves downstream analyses.

4.3.1 Cell-Cluster Identity

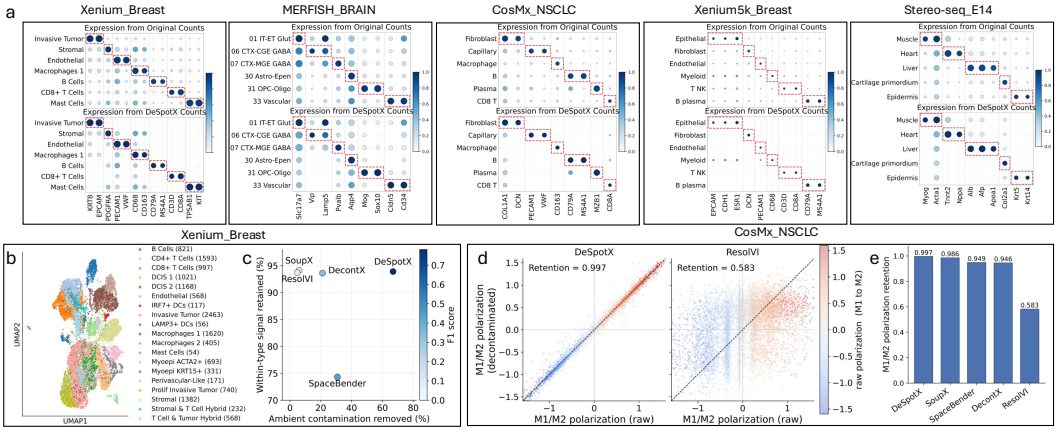

Figure 3: DeSpotX improves marker-gene specificity while preserving native signal and within-cluster cell states. **a.** Marker-gene dotplots for raw and DeSpotX-decontaminated counts across cell clusters in each dataset, with red boxes marking canonical marker-cluster pairs. **b.** UMAP embedding of the matched scRNA-seq reference for Xenium_Breast. **c.** Within-type signal retained vs. contamination removed per method. **d.** Per-cell M1/M2 macrophage subtype score in raw vs. decontaminated counts for DeSpotX and ResolVI in CosMx_NSCLC. **e.** M1/M2 score retention across methods.

We first examine whether decontamination tightens the localization of marker-gene expression to canonical clusters. Figure 3a shows marker-gene dotplots for the raw counts and the DeSpotX-decontaminated counts across cell clusters in each dataset. After decontamination, expression of canonical markers, defined from established biology independently of DeSpotX, is reduced in non-expressing clusters while remaining concentrated in their expected clusters (red boxes). This pattern is consistent with the removal of ambient contamination from non-expressing cells, with native

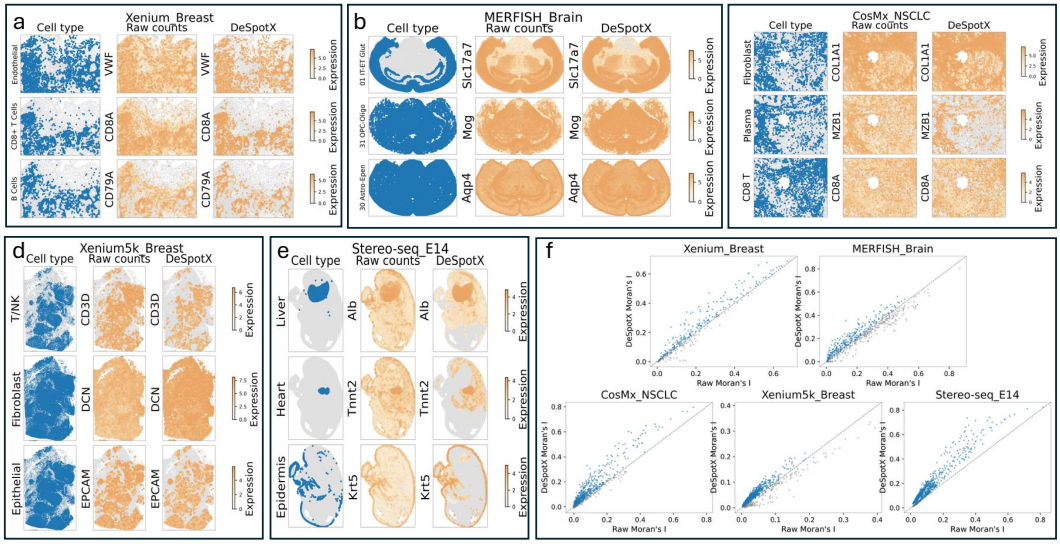

Figure 4: DeSpotX-decontaminated expression is spatially coherent. (a-e) Spatial maps of canonical cell-cluster markers in each dataset, comparing raw and DeSpotX-decontaminated counts. (f) Per-gene Moran's I for raw and DeSpotX-decontaminated counts across the five datasets.

expression preserved in expressing cells. The same comparison for the four baseline methods is provided in Appendix M. UMAP embeddings and kNN purity for all methods are provided in Appendix N. Negative controls with split and permuted labels show that DeSpotX does not create label-driven structure in each cluster's own genes or in kNN purity (Appendix O).

We test whether DeSpotX preserves native signal and within-cluster cell states using two label-independent analyses. First, we use matched scRNA-seq from the same Xenium_Breast tissue (Janesick et al., 2023) to label genes as present or absent per cell type (Figure 3b). DeSpotX retains 94% of within-type signal, indicating limited over-correction and achieving the highest F1 (Figure 3c). Second, we test whether decontamination preserves within-cluster cell states on CosMx_NSCLC macrophages, which share a single cluster label but span the M1/M2 continuum. DeSpotX preserves subtype structure with retention of 0.997, comparable to SoupX, DecontX, and SpaceBender, while ResolVI distorts it substantially with retention of 0.583 (Figure 3d,e).

4.3.2 Spatial coherence

We next assess whether DeSpotX decontamination preserves the spatial structure of marker expression. Figure 4a-e shows the spatial distribution of canonical cell-cluster markers in the raw and DeSpotX-decontaminated counts. In the raw counts, marker expression is detected throughout the tissue, including in regions where the canonical expressing cluster is absent. After DeSpotX, marker expression is concentrated within the canonical clusters and the surrounding tissue is largely cleared, indicating that DeSpotX removes ambient signal in non-expressing regions while preserving expression in regions where the marker is natively transcribed. Spatial maps of the removed counts further show that DeSpotX's removal is spatially structured, concentrating near canonical expressing clusters, consistent with the assumption that contamination spreads locally from source cells (Appendix P).

Figure 4f reports per-gene Moran's I for the raw and DeSpotX-decontaminated counts across the five datasets. Moran's I measures whether gene expression is spatially autocorrelated, with higher values indicating that cells with similar expression are spatially close. Most genes show higher Moran's I after DeSpotX, with the largest gains in genes that already exhibit some spatial structure in the raw data, indicating that decontamination preserves and strengthens biologically meaningful spatial patterns. The same comparison for baseline methods is provided in Appendix Q.

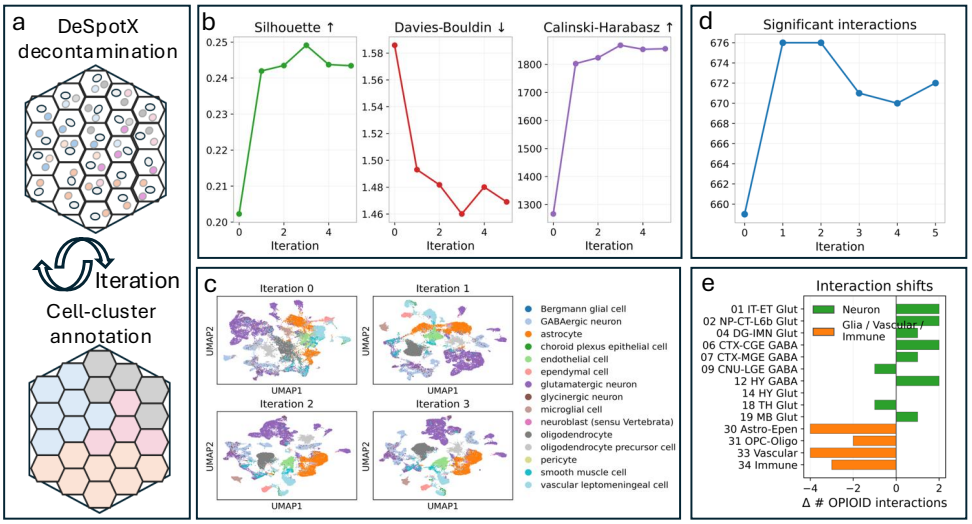

Figure 5: Iterative DeSpotX decontamination and cell-cluster annotation on `MERFISH_Brain`. (a) Schematic of the iterative procedure. (b) Cluster metrics across iterations. (c) UMAP embeddings of the decontaminated counts at iterations 0-3. (d) Total number of significant cell-cell interactions inferred by CellChat across iterations. (e) Change in the number of significant OPIOID-pathway interactions per sender cell-cluster between iteration 0 and iteration 5.

4.3.3 Iterative decontamination and cell-cluster annotation

DeSpotX requires a cell-cluster annotation as input, but the quality of this annotation depends on the data, which itself contains contamination. We therefore evaluate whether iterating between decontamination and cell-cluster annotation produces progressively better results. At each iteration, DeSpotX decontaminates the raw counts using the previous iteration's annotation, and the decontaminated counts are re-annotated by a CellTypist classifier (Domínguez Conde et al., 2022) trained on an external reference atlas (Yao et al., 2023), which provides labels independent of DeSpotX (Figure 5a).

We apply this procedure to `MERFISH_Brain` for five iterations. Cluster-validity metrics for the current annotation improve and stabilize by iteration 3 (Figure 5b), with Silhouette rising from 0.20 to 0.24, Davies-Bouldin falling from 1.59 to 1.47, and Calinski-Harabasz increasing from 1,267 to 1,856. The UMAP embeddings show progressively cleaner separation of the brain cell types (Figure 5c). To test whether these improvements reflect genuine biological signal, we run CellChat (Jin et al., 2021) on inferred ligand-receptor signaling networks at each iteration. The number of significant interactions stabilizes by iteration 2-3 (Figure 5d). Within the OPIOID pathway, the sender distribution shifts (Figure 5e), moving away from glial, vascular, and immune senders toward neuronal populations, consistent with the neuronal origin of opioid peptides (Le Merrer et al., 2009; Yao et al., 2023). The same iterative procedure applied to `Xenium_Breast` also shows improvements in downstream analyses (Appendix R).

5 Discussion

We have presented DeSpotX, which constrains the contamination decomposition with automatically inferred anchor genes and estimates contamination locally from cross-cluster neighbors. DeSpotX attains the highest AUROC in 27 of 30 benchmark settings, remains the best method in a label-free transcript-level simulation, and produces more biologically coherent outputs on real tissue.

We outline two limitations along with corresponding future work. First, DeSpotX is designed for single-cell-resolution ST platforms; spot-resolution platforms such as Visium or Slide-seq aggregate multiple cells per spot, and extending DeSpotX would require modeling cell-cluster mixtures. Second, the identifiability guarantee depends on the anchor genes within each cluster, which may be sparse in highly homogeneous tissues; future work could explore softer anchor formulations.

AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on experimental design, help identify related work, and check code and mathematical arguments for potential errors. We also used these tools to improve the grammar and wording of text written by the authors and to clarify the presentation of mathematical arguments. We did not use these tools to design the proposed method, formulate mathematical claims, generate proofs, or interpret results. The authors reviewed and verified all AI-assisted suggestions, including those related to code and mathematics, and made the final decisions about the experiments and manuscript. We take responsibility for the final content of this work.

ETHICS STATEMENT

This study uses only publicly available, de-identified spatial transcriptomics and single-cell RNA-seq datasets, and no new human or animal data were collected. The method is intended for research use and has not been validated for clinical decision-making.

REPRODUCIBILITY STATEMENT

The problem formulation, model, and inference procedure are described in Section 3, with implementation details and hyperparameters in Appendix C. Proofs of Lemmas 1-4 are provided in Appendix B. Datasets are listed in Table 1, and the spike-in simulation design is described in Section 4.2.1 and Appendix D. Hardware and runtime are reported in Appendices C.4 and L. We will publicly release the source code upon acceptance.

REFERENCES

10x Genomics. FFPE Human Breast Cancer with 5K Human Pan Tissue and Pathways Panel plus 100 Custom Genes. Public dataset, Xenium Onboard Analysis v3.0.0, October 2024. URL https://www.10xgenomics.com/datasets/xenium-prime-ffpe-human-breast-cancer. Accessed: 2026-04-22.

Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu. A practical algorithm for topic modeling with provable guarantees. In *International conference on machine learning*, pp. 280–288. PMLR, 2013.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations*, 2022.

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

Daniel Guorui Chen, Antoni Ribas, and Katie M Campbell. SpaceBender: Denoising Spatial Transcriptomics Data to Enhance Biological Signals. *bioRxiv*, pp. 2026–04, 2026.

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

César Domínguez Conde, Chao Xu, Louie B Jarvis, Daniel B Rainbow, Sara B Wells, Tamir Gomes, SK Howlett, O Suchanek, K Polanski, HW King, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594):eabl5197, 2022.

David Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? *Advances in neural information processing systems*, 16, 2003.

Can Ergen and Nir Yosef. ResolVI-addressing noise and bias in spatial transcriptomics. *bioRxiv*, pp. 2025–01, 2025.

Stephen J Fleming, Mark D Chaffin, Alessandro Arduini, Amer-Denis Akkad, Eric Banks, John C Marioni, Anthony A Philippakis, Patrick T Ellinor, and Mehrtash Babadi. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nature methods*, 20(9):1323–1335, 2023.

Shanshan He, Ruchir Bhatt, Carl Brown, Emily A Brown, Derek L Buhr, Kan Chantranuvatana, Patrick Danaher, Dwayne Dunaway, Ryan G Garrison, Gary Geiss, et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature biotechnology*, 40(12):1794–1806, 2022.

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

Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, and Qing Nie. Inference and analysis of cell-cell communication using CellChat. *Nature communications*, 12(1):1088, 2021.

Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ica: A unifying framework. In *International conference on artificial intelligence and statistics*, pp. 2207–2217. PMLR, 2020.

Aaron Wing Cheung Kwok, Annika Vannan, Nicholas E Banovich, Jonathan A Kropski, Heejung Shim, and Davis J McCarthy. Denoising image-based spatial transcriptomics data with DenoIST. *bioRxiv*, pp. 2025–11, 2025.

Julie Le Merrer, Jérôme AJ Becker, Katia Befort, and Brigitte L Kieffer. Reward processing by the opioid system in the brain. *Physiological reviews*, 2009.

Jeffrey R Moffitt, Emma Lundberg, and Holger Heyn. The emerging landscape of spatial profiling technologies. *Nature Reviews Genetics*, 23(12):741–759, 2022.

Giovanni Palla, David S Fischer, Aviv Regev, and Fabian J Theis. Spatial components of molecular tissue biology. *Nature biotechnology*, 40(3):308–318, 2022.

Austin D Reed, Sara Pensa, Adi Steif, Jack Stenning, Daniel J Kunz, Linsey J Porter, Kui Hua, Peng He, Alecia-Jane Twigger, Abigail JQ Siu, et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nature genetics*, 56(4):652–662, 2024.

Shiyi Yang, Sean E Corbett, Yusuke Koga, Zhe Wang, W Evan Johnson, Masanao Yajima, and Joshua D Campbell. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome biology*, 21(1):57, 2020.

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

Matthew D Young and Sam Behjati. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*, 9(12):giaa151, 2020.

Meng Zhang, Xingjie Pan, Won Jung, Aaron R Halpern, Stephen W Eichhorn, Zhiyun Lei, Limor Cohen, Kimberly A Smith, Bosiljka Tasic, Zizhen Yao, et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature*, 624(7991):343–354, 2023.

A Comparison of decontamination methods

Table 3 summarizes the key differences between DeSpotX and existing decontamination methods for scRNA-seq and ST data. The methods differ in the data type, the model architecture, the contamination estimator, the use of cluster annotations, and the constraint used to ensure identifiability. SoupX uses non-expressed genes, specified manually or selected automatically from cluster markers, to estimate a contamination fraction against a global ambient profile, but does not use them to constrain the per-cell decomposition. DeSpotX uses anchor genes as a structural constraint on each cell's decomposition within a spatial generative model, with a spatially varying contamination profile.

Table 3: Comparison of decontamination methods for scRNA-seq and ST data.

Method	Data type	Architecture	Contamination estimator	Cluster-aware	Identifiability constraint
SoupX	scRNA-seq	Linear subtraction	Global ambient profile from empty droplets	Yes	Non-expressed genes (global fraction)
DecontX	scRNA-seq	Bayesian mixture	Global ambient profile, cluster-aggregated	Yes	None
CellBender	scRNA-seq	Deep generative model	Learned ambient prior from empty droplets	No	None
SpaceBender	ST	Deep generative model	Local ambient profile from spatial neighbors	No	None
DenoIST	ST	Mixture model	Local Poisson mixture per cell	No	None
ResolVI	ST	Deep generative model	Latent decomposition with neighbor diffusion	No	None
DeSpotX	**ST**	**Deep generative model with anchor constraints**	**Cluster-isolated spatial contamination**	**Yes**	**Anchor genes (per-cell decomposition)**

B PROOFS

B.1 PROOF OF LEMMA 1

We want to show two things: (i) the set $\mathcal{S}(m)$ contains infinitely many triples, and (ii) the likelihood in Eq. 1 is the same at all of them.

Fix $m \in \Delta^{G-1}$. For any $\varepsilon \in [0, 1)$ and any $\chi \in \Delta^{G-1}$ satisfying $\varepsilon\chi_g \leq m_g$ for all g, define

$$\phi = \frac{m - \varepsilon\chi}{1 - \varepsilon}.$$

We check that the triple $(\varepsilon, \phi, \chi)$ lies in $\mathcal{S}(m)$, i.e., that it satisfies the conditions in the Lemma's definition. First, ϕ is a valid probability vector; each coordinate is non-negative because $\varepsilon\chi_g \leq m_g$, and the coordinates sum to one,

$$\sum_g \phi_g = \frac{\sum_g m_g - \varepsilon \sum_g \chi_g}{1 - \varepsilon} = \frac{1 - \varepsilon}{1 - \varepsilon} = 1,$$

using $\sum_g m_g = \sum_g \chi_g = 1$.

Second, ϕ satisfies the mixture equation: substituting the definition of ϕ gives

$$(1 - \varepsilon)\phi + \varepsilon\chi = (1 - \varepsilon) \cdot \frac{m - \varepsilon\chi}{1 - \varepsilon} + \varepsilon\chi = m - \varepsilon\chi + \varepsilon\chi = m.$$

Hence $(\varepsilon, \phi, \chi) \in \mathcal{S}(m)$.

Since ε and χ can be chosen from a continuum of admissible values, and each choice produces a distinct triple, $\mathcal{S}(m)$ is a continuum, proving (i). For (ii), the likelihood in Eq. 1 depends on $(\varepsilon, \phi, \chi)$ only through the mixture m, so it is constant across $\mathcal{S}(m)$.

B.2 ERROR UNDER WITHIN-CLUSTER CONTAMINATION

DeSpotX's cross-cluster mask (Eq. 3) excludes same-cluster neighbors when estimating contamination, leaving a within-cluster residual in $\hat{\phi}_i$. We show this residual contributes a native-profile error bounded by the contamination level times the within-cluster heterogeneity, while $\hat{\varepsilon}_i$ recovers the cross-cluster contamination rate.

Decompose the true contamination profile at cell i as

$$\chi_i^* \;=\; \alpha_i\, \chi_i^{\text{cross}} \;+\; (1 - \alpha_i)\, \chi_i^{\text{within}}, \qquad \alpha_i \in [0, 1], \tag{10}$$

where α_i is the cross-cluster fraction. Substituting into Eq. 1 gives

$$m_i \;=\; (1-\varepsilon_i^*)\,\phi_{t_i}(z_i) \;+\; \varepsilon_i^*\alpha_i\,\chi_i^{\text{cross}} \;+\; \varepsilon_i^*(1-\alpha_i)\,\chi_i^{\text{within}}, \tag{11}$$

while DeSpotX fits

$$m_i \;=\; (1-\hat{\varepsilon}_i)\,\hat{\phi}_i \;+\; \hat{\varepsilon}_i\,\chi_i^{\text{cross}}. \tag{12}$$

Let

$$\sigma_i^{\text{within}} \;=\; \left\|\chi_i^{\text{within}} - \phi_{t_i}(z_i)\right\|_\infty \tag{13}$$

denote the within-cluster heterogeneity at cell i, the deviation between cell i's native expression and the average expression of its same-cluster spatial neighbors.

Lemma 3 (Error from within-cluster contamination)**.** *Under the conditions of Lemma 2, with $\varepsilon_i^* < 1$, define the scaling factor*

$$\beta_i \;=\; \frac{\varepsilon_i^*(1-\alpha_i)}{1-\alpha_i\varepsilon_i^*}. \tag{14}$$

The unique fit of Eq. 12 to the true mixture in Eq. 11 satisfies $\hat{\varepsilon}_i = \alpha_i\varepsilon_i^*$ and

$$\left\|\hat{\phi}_i - \phi_{t_i}(z_i)\right\| \;=\; \beta_i \cdot \sigma_i^{\text{within}} \;\le\; \varepsilon_i^* \cdot \sigma_i^{\text{within}}. \tag{15}$$

Proof. Equating Eqs. 11 and 12 gives

$$(1-\hat{\varepsilon}_i)\hat{\phi}_i + \hat{\varepsilon}_i\chi_i^{\text{cross}} \;=\; (1-\varepsilon_i^*)\phi_{t_i}(z_i) + \varepsilon_i^*\alpha_i\chi_i^{\text{cross}} + \varepsilon_i^*(1-\alpha_i)\chi_i^{\text{within}}.$$

At any anchor position $g \in \mathcal{A}_{t_i}$, $\phi_{t_i,g}(z_i) = 0$ by definition. Since χ_i^{within} averages the native expression of the cluster-t_i cells, which is zero at anchor positions, $\chi_{i,g}^{\text{within}} = 0$ as well. The anchor constraint forces $\hat{\phi}_{i,g} = 0$, so the equation at g reduces to $\hat{\varepsilon}_i\chi_{i,g}^{\text{cross}} = \varepsilon_i^*\alpha_i\chi_{i,g}^{\text{cross}}$. Since at least one anchor satisfies $\chi_{i,g}^{\text{cross}} > 0$, $\hat{\varepsilon}_i = \alpha_i\varepsilon_i^*$. Substituting back,

$$(1-\alpha_i\varepsilon_i^*)\hat{\phi}_i \;=\; (1-\varepsilon_i^*)\phi_{t_i}(z_i) + \varepsilon_i^*(1-\alpha_i)\chi_i^{\text{within}}.$$

Dividing by $1-\alpha_i\varepsilon_i^*$ and subtracting $\phi_{t_i}(z_i)$ gives $\hat{\phi}_i - \phi_{t_i}(z_i) = \beta_i[\chi_i^{\text{within}} - \phi_{t_i}(z_i)]$, and Eq. 15 follows by taking norms.

Remarks. The error in Eq. 15 is the product of two small quantities. The scaling factor β_i is bounded by ε_i^* (for example, $\beta_i = 0.07$ at $\varepsilon_i^* = 0.2$ and $\alpha_i = 0.7$). The within-cluster heterogeneity σ_i^{within} is small for well-defined clusters.

The recovered contamination fraction $\hat{\varepsilon}_i = \alpha_i\varepsilon_i^*$ is the cross-cluster rate, which drives the downstream artifacts that motivate decontamination, such as cell-type misannotation, smeared marker maps, and spurious cell-cell communication. The within-cluster portion absorbed into $\hat{\phi}_i$ does not cause these artifacts, since the recipient cell expresses similar genes natively.

We defer validation of Lemma 3 on spike-in simulations to Appendix K.

B.3 Proof of Lemma 2

We show two things: (i) ε_i is identifiable from observations at anchor positions, and (ii) ϕ_{t_i} is identifiable at all non-anchor positions. Both follow under the preconditions of the Lemma: χ_i given as input, constructed without using x_i, and at least one anchor with $\chi_{i,g} > 0$.

Recall that $m_i = (1-\varepsilon_i)\phi_{t_i} + \varepsilon_i\chi_i$ denotes the mixture mean satisfying $\mathbb{E}[x_i/d_i] = m_i$. Because χ_i is not a function of x_i, the identification of ε_i and ϕ_{t_i} from x_i is not circular.

For (i), consider any anchor position $g \in \mathcal{A}_{t_i}$. By the anchor assumption $\phi_{t_i,g} = 0$, the mixture equation reduces to

$$m_{i,g} = \varepsilon_i\chi_{i,g}.$$

Since by assumption there exists $g \in \mathcal{A}_{t_i}$ with $\chi_{i,g} > 0$, solving gives

$$\varepsilon_i = \frac{m_{i,g}}{\chi_{i,g}},$$

which is a function of observable quantities and the (independently obtained) χ_i.

For (ii), fix ε_i and χ_i as above. For any non-anchor position $g \notin \mathcal{A}_{t_i}$, the mixture equation gives

$$\phi_{t_i,g} = \frac{m_{i,g} - \varepsilon_i \chi_{i,g}}{1 - \varepsilon_i},$$

which expresses $\phi_{t_i,g}$ in terms of observables and parameters identified in (i).

Hence both ε_i and ϕ_{t_i} are identifiable.

B.4 Stability under approximate inputs

Lemma 2 assumes an exact contamination profile, exact anchor zeros, and an exact fit. We give a conditional bound when all three are approximate. Fix a cell and drop the index i. Let $\varepsilon, \hat{\varepsilon} \in [0,1]$ and $\phi, \hat{\phi}, \chi, \hat{\chi} \in \Delta^{G-1}$, with true mean $m = (1-\varepsilon)\phi + \varepsilon\chi$, fitted mean $\hat{m} = (1-\hat{\varepsilon})\hat{\phi} + \hat{\varepsilon}\hat{\chi}$, and $\Delta = \hat{\chi} - \chi$. For the anchor set $\mathcal{A} = \mathcal{A}_{t_i}$, let $\eta = \sum_{g\in\mathcal{A}} \phi_g$, $\delta = \sum_{g\in\mathcal{A}} \hat{\phi}_g$, $\rho = \|\hat{m} - m\|_1$, and $\hat{\chi}_{\mathcal{A}} = \sum_{g\in\mathcal{A}} \hat{\chi}_g > 0$.

Lemma 4 (Stability of the decomposition)**.** *Under the definitions above,*

$$|\hat{\varepsilon} - \varepsilon| \;\le\; \frac{\rho + (1-\varepsilon)\eta + (1-\hat{\varepsilon})\delta + \varepsilon\|\Delta\|_1}{\hat{\chi}_{\mathcal{A}}}, \tag{16}$$

and, whenever $\hat{\varepsilon} < 1$,

$$\|\hat{\phi} - \phi\|_\infty \;\le\; \frac{\rho + |\hat{\varepsilon} - \varepsilon| + \|\Delta\|_\infty}{1 - \hat{\varepsilon}}. \tag{17}$$

If $\Delta = 0$, $\eta = \delta = \rho = 0$, and $\varepsilon < 1$, then $\hat{\varepsilon} = \varepsilon$ and $\hat{\phi} = \phi$, recovering Lemma 2.

Proof. Summing both mixtures over $g \in \mathcal{A}$ gives $m_{\mathcal{A}} = (1-\varepsilon)\eta + \varepsilon\chi_{\mathcal{A}}$ and $\hat{m}_{\mathcal{A}} = (1-\hat{\varepsilon})\delta + \hat{\varepsilon}\hat{\chi}_{\mathcal{A}}$. Substituting $\chi_{\mathcal{A}} = \hat{\chi}_{\mathcal{A}} - \Delta_{\mathcal{A}}$ and subtracting,

$$(\hat{\varepsilon} - \varepsilon)\,\hat{\chi}_{\mathcal{A}} = (\hat{m}_{\mathcal{A}} - m_{\mathcal{A}}) - (1-\hat{\varepsilon})\delta + (1-\varepsilon)\eta - \varepsilon\Delta_{\mathcal{A}},$$

and the first bound follows from $|\hat{m}_{\mathcal{A}} - m_{\mathcal{A}}| \le \rho$ and $|\Delta_{\mathcal{A}}| \le \|\Delta\|_1$. Solving both mixtures for the native profile and subtracting gives

$$\hat{\phi} - \phi = \frac{(\hat{m} - m) + (\hat{\varepsilon} - \varepsilon)(\phi - \chi) - \hat{\varepsilon}\Delta}{1 - \hat{\varepsilon}},$$

and the second bound follows from $\|\phi - \chi\|_\infty \le 1$ and $\hat{\varepsilon} \le 1$.

Remarks. The bound is conditional and is small when the approximation errors are small and $\hat{\chi}_{\mathcal{A}}$ and $1 - \hat{\varepsilon}$ are bounded away from zero, so anchors with little contamination mass give weaker stability. The checks in Figure 2a and Appendix I relate to η and Δ but do not quantify them, and the training loss does not directly estimate ρ.

C Model details

C.1 Spatial graph encoder: implementation details

Node features. For each node j in the star graph $\mathcal{G}_i$, the input feature is

$$\mathbf{n}_j = \left[\, \log(1 + x_j) \,\|\, \mathbf{e}_{t_j} \,\right] \in \mathbb{R}^{G+d_e},$$

where $\log(1+\cdot)$ is applied elementwise to the count vector, $\mathbf{e}_{t_j} \in \mathbb{R}^{d_e}$ is a learned cluster embedding, and $\|$ denotes concatenation. The cluster embedding is initialized randomly and trained jointly with the rest of the model. A linear projection $W_{\text{in}} : \mathbb{R}^{G+d_e} \to \mathbb{R}^{d_h}$ maps $\mathbf{n}_j$ to a hidden representation $\mathbf{n}_j^{(0)} = W_{\text{in}}\mathbf{n}_j$ used as input to the GAT layer.

Edge features. The edge from leaf j to center i carries the kernel weight $w_{ij} = \exp(-\|s_j - s_i\|/\rho_i) \in (0,1]$ as a single scalar feature, with ρ_i set to the median Euclidean distance from cell i to its K nearest neighbors. The exponential form approximates the spatial decay of free ambient RNA, and the per-cell adaptive bandwidth ρ_i accommodates variation in cell density across the dataset.

GATv2 attention. GATv2 (Brody et al., 2022) computes attention coefficients of the form

$$\alpha_{ji} \;\propto\; \exp\Big(\mathbf{a}^\top \mathrm{LeakyReLU}\big(W_q \mathbf{n}_i^{(0)} + W_k \mathbf{n}_j^{(0)} + W_e w_{ji}\big)\Big),$$

where W_q, W_k are query and key projections, W_e embeds the scalar edge weight w_{ji}, and $\mathbf{a}$ is a learned scoring vector. We use H attention heads with averaging across heads and per-head output dimension d_h, so the output dimension is d_h. For batched training, B star graphs are concatenated into a single disjoint graph and the center nodes are recovered by index after the forward pass.

Readouts from h_i. Two linear heads applied to the center's spatial-context embedding h_i produce the latent state and the contamination-fraction estimate in Eq. 2:

$$z_i = W_z h_i, \tag{18}$$

$$\bar{\varepsilon}_i = \sigma\big(w_\varepsilon^\top h_i + b_\varepsilon\big), \tag{19}$$

where W_z, w_ε, and b_ε are learned parameters. We initialize $b_\varepsilon = \mathrm{logit}(0.20)$.

Hyperparameters. Default values used in our experiments: GAT hidden dimension $d_h = 256$, $H = 4$ attention heads, cluster embedding dimension $d_e = 32$, $K = 12$ spatial neighbors, and dropout 0.1.

C.2 Cluster-conditioned decoder: implementation details

Architecture. The function f_θ in Eq. 4 is a two-layer MLP with hidden dimension d_{dec} and Softplus activations:

$$f_\theta(z_i, e_{t_i}) \;=\; W_2\,\mathrm{Softplus}\big(W_1[z_i \,\|\, e_{t_i}] + b_1\big) + b_2,$$

where $[\cdot \,\|\, \cdot]$ denotes concatenation, $W_1 \in \mathbb{R}^{d_{\mathrm{dec}} \times (d_z + d_e)}$, $W_2 \in \mathbb{R}^{G \times d_{\mathrm{dec}}}$, $b_1 \in \mathbb{R}^{d_{\mathrm{dec}}}$, $b_2 \in \mathbb{R}^G$, and $e_{t_i} \in \mathbb{R}^{d_e}$ is a learned cluster embedding. We initialize W_2 with Xavier-uniform values scaled by 0.1 and $b_2 = 0$.

Cluster-specific logit shift. The vector $b_{t_i} \in \mathbb{R}^G$ in Eq. 4 is initialized to $b_{t_i,g} = \log \bar{x}_{t_i,g} - \frac{1}{G}\sum_{g'} \log \bar{x}_{t_i,g'}$, where $\bar{x}_{t_i,g}$ is the mean count of gene g over cluster-t_i training cells. After initialization, b_{t_i} is fixed; cluster-specific learnable signal is carried by e_{t_i} and f_θ.

Hyperparameters. Default values used in our experiments: $d_{\mathrm{dec}} = 256$, $d_e = 32$.

C.3 Anchor mask construction: implementation details

Per-cluster expression statistics. Let $r_{t,g}$ denote the fraction of cluster-t cells with nonzero counts for gene g in the training data. We construct the binary anchor mask

$$A_{t,g} \;=\; \mathbf{1}\big[r_{t,g} < \tau_g\big], \tag{20}$$

with adaptive per-gene threshold

$$\tau_g \;=\; \mathrm{clip}\big(\kappa \cdot \max_t r_{t,g},\; \tau_{\min},\; \tau_{\max}\big),$$

which scales with each gene's maximum cluster-level expression rate. Default value used in our experiments: $\kappa = 0.3$.

C.4 Training: implementation details

Loss weights. Default values used in our experiments: $\lambda_A = 50$ and $\lambda_E = 1$. Although λ_A appears large relative to λ_E, the anchor penalty vanishes once $\phi_{t_i,g}$ is driven to zero at anchor positions, so λ_A only governs the convergence speed of this collapse.

Regularizer on $\bar{\varepsilon}_i$. We fix $\nu = 20$ and learn μ_ε as a global scalar, initialized at 0.20 and clamped to $[0.05, 0.95]$. The clamp keeps both Beta shape parameters at least one, so the negative log-density penalty is bounded below and minimized at the Beta mode $(\mu_\varepsilon \nu - 1)/(\nu - 2)$ (≈ 0.167 at $\mu_\varepsilon = 0.2$).

Anchor warm-up. The anchor weight λ_A is linearly increased from 0 to 50 over the first $E_{\text{warm}} = 3$ epochs and held at 50 thereafter.

Optimizer. All parameters, including the encoder, decoder, gene-wise dispersion θ, and μ_ε, are optimized with Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$). We use learning rate 10^{-3}, batch size 128, and gradient clipping at norm 10. Training runs for 10 epochs.

Compute. All experiments were run on a single NVIDIA Tesla T4 GPU (16 GB), with training taking 5-20 minutes per dataset.

D SPIKE-IN SIMULATION DETAILS

Per-cell contamination rates. For each cell i with native counts $\mathbf{n}_i \in \mathbb{N}^G$, library size $d_i = \sum_g n_{i,g}$, and cell-cluster label t_i, we draw a per-cell contamination rate from a Beta distribution centered at the noise level $\bar{\varepsilon}$,

$$\varepsilon_i \sim \text{Beta}\big(4,\, 4(1-\bar{\varepsilon})/\bar{\varepsilon}\big), \qquad \mathbb{E}[\varepsilon_i] = \bar{\varepsilon}, \tag{21}$$

with shape parameter fixed at 4 to give moderate per-cell variance. The Beta concentration $4/\bar{\varepsilon}$ coincides with DeSpotX's fixed $\nu = 20$ only at $\bar{\varepsilon} = 20\%$; AUROC is stable across all five noise levels (Figure 13), suggesting that performance is insensitive to this match. Unless otherwise noted, contamination counts are sampled independently per gene as

$$c_{i,g} \sim \text{Poisson}\big(\varepsilon_i\, d_i\, a_{i,g}\big), \tag{22}$$

and the spiked counts delivered to each method are $s_{i,g} = n_{i,g} + c_{i,g}$. The six mechanisms differ in how the ambient profile $\mathbf{a}_i$ is constructed.

Contamination mechanisms. The local cross-cluster mechanism matches DeSpotX's assumptions, and the other five are off-model, representing a global ambient pool, segmentation errors, long-range diffusion, directional diffusion along a tissue axis, and vertical cell stacking.

- *Local cross-cluster.* $\mathbf{a}_i = Z_i^{-1} \sum_{j \in \mathcal{N}_K^{\neq}(i)} w_{ij} \cdot \mathbf{n}_j / d_j$ with $w_{ij} = (\|\mathbf{s}_i - \mathbf{s}_j\| + \epsilon_0)^{-1}$, where $\mathcal{N}_K^{\neq}(i)$ is the set of $K = 12$ Euclidean nearest neighbors of cell i restricted to cells with $t_j \neq t_i$, falling back to the unrestricted neighbor set when no other-cluster neighbor exists; $\epsilon_0 = 10^{-6}$ is a numerical stabilizer; and $Z_i = \sum_{j \in \mathcal{N}_K^{\neq}(i)} w_{ij}$ normalizes $\mathbf{a}_i$ to the simplex.
- *Global ambient.* A single dataset-wide profile $a_g = \sum_i x_{i,g} / \sum_{i,g} x_{i,g}$ is used for all cells, independent of spatial location.
- *Segmentation spillover.* Each transcript in cell i is independently transferred to cell i's nearest spatial neighbor with probability ε_i, regardless of cell type. Contamination counts are the integer transcripts received rather than Poisson samples.
- *Long-range diffusion.* The ambient profile is computed from K=50 cross-cluster neighbors with power-law weights $w_{ij} = (\|\mathbf{s}_j - \mathbf{s}_i\| + \epsilon_0)^{-\alpha}$, $\alpha = 1.5$.
- *Anisotropic diffusion.* The ambient profile uses direction-dependent weights $w_{ij} = \exp(-\|\mathbf{s}_j - \mathbf{s}_i\|/\rho_i) \cdot \exp(\gamma \cdot |\cos\theta_{ij}|)$ with $\gamma = 3.0$, where θ_{ij} is the angle to a per-dataset random axis $\mathbf{u}$ and ρ_i is the median K-nearest-neighbor distance.
- *Vertical overlap.* The ambient profile is the full transcriptome of cell i's single nearest 2D neighbor, L_1-normalized, regardless of cell type.

Each off-model mechanism targets a different structural assumption of DeSpotX's estimator, spanning spatial locality, decay shape, isotropy, cross-cluster masking, and diffusive mixing.

E FULL SPIKE-IN SIMULATION RESULTS

Table 4 extends the AUROC comparison of Table 2 to CosMx_NSCLC, Xenium5k_Breast, and Stereo-seq_E14. DeSpotX attains the highest AUROC in 14 of the 18 settings and ties ResolVI on

Table 4: Benchmark comparison on CosMx_NSCLC, Xenium5k_Breast, and Stereo-seq_E14 across six contamination mechanisms (one on-model, five off-model). AUROC (mean ± std across the five non-zero contamination levels). Best value per row in bold; values within 0.005 of the best are also bolded. Results for Xenium_Breast and MERFISH_Brain are in Table 2.

Dataset	Mechanism	**DeSpotX**	SoupX	DecontX	ResolVI	SpaceBender
CosMx_NSCLC	Local cross-cluster	**0.967 ± 0.003**	0.696 ± 0.011	0.683 ± 0.002	0.886 ± 0.005	0.700 ± 0.013
	Global ambient	**0.903 ± 0.002**	0.768 ± 0.070	0.761 ± 0.031	0.891 ± 0.009	0.618 ± 0.012
	Segmentation spillover	**0.907 ± 0.001**	0.571 ± 0.016	0.673 ± 0.006	0.885 ± 0.011	0.669 ± 0.035
	Long-range diffusion	**0.935 ± 0.001**	0.598 ± 0.019	0.688 ± 0.003	0.890 ± 0.011	0.684 ± 0.006
	Anisotropic diffusion	**0.966 ± 0.003**	0.596 ± 0.027	0.685 ± 0.003	0.884 ± 0.008	0.717 ± 0.006
	Vertical overlap	**0.929 ± 0.003**	0.582 ± 0.018	0.667 ± 0.003	0.884 ± 0.008	0.742 ± 0.010
Xenium5k_Breast	Local cross-cluster	**0.997 ± 0.000**	0.705 ± 0.024	0.732 ± 0.006	0.980 ± 0.000	0.604 ± 0.020
	Global ambient	**0.949 ± 0.009**	0.941 ± 0.000	0.696 ± 0.012	**0.949 ± 0.001**	0.521 ± 0.001
	Segmentation spillover	**0.946 ± 0.001**	0.928 ± 0.006	0.689 ± 0.005	**0.944 ± 0.000**	0.539 ± 0.007
	Long-range diffusion	**0.968 ± 0.001**	0.938 ± 0.001	0.672 ± 0.015	**0.969 ± 0.002**	0.582 ± 0.005
	Anisotropic diffusion	**0.992 ± 0.000**	0.937 ± 0.004	0.677 ± 0.008	0.979 ± 0.001	0.608 ± 0.009
	Vertical overlap	**0.982 ± 0.001**	0.874 ± 0.084	0.667 ± 0.005	0.976 ± 0.000	0.654 ± 0.012
Stereo-seq_E14	Local cross-cluster	**0.984 ± 0.001**	0.719 ± 0.051	0.721 ± 0.002	0.929 ± 0.004	0.904 ± 0.010
	Global ambient	**0.910 ± 0.004**	0.844 ± 0.023	0.767 ± 0.012	**0.908 ± 0.001**	0.809 ± 0.000
	Segmentation spillover	**0.907 ± 0.001**	0.731 ± 0.007	0.707 ± 0.002	**0.908 ± 0.000**	0.808 ± 0.000
	Long-range diffusion	**0.942 ± 0.000**	0.732 ± 0.004	0.725 ± 0.003	0.936 ± 0.000	0.863 ± 0.000
	Anisotropic diffusion	**0.984 ± 0.000**	0.726 ± 0.007	0.720 ± 0.002	0.937 ± 0.002	0.909 ± 0.000
	Vertical overlap	**0.920 ± 0.000**	0.752 ± 0.023	0.725 ± 0.005	**0.922 ± 0.002**	0.914 ± 0.000

Xenium5k_Breast global ambient; in the remaining three, ResolVI leads by 0.001 to 0.002. The margin is largest on CosMx_NSCLC, where DeSpotX exceeds the best baseline by 0.012 to 0.082.

Table 5 evaluates removal magnitude using the relative mean absolute error on contaminated entries, $\mathrm{relMAE} = \sum_{C_{ig}>0} |R_{ig} - C_{ig}| / \sum_{C_{ig}>0} C_{ig}$, where R_{ig} is the removed count and C_{ig} the injected contamination. DeSpotX attains the lowest relMAE in 27 of 30 settings, and in the other three it is within 0.004 of the best baseline. The gains are largest under local cross-cluster and anisotropic diffusion, reaching 0.20 on MERFISH_Brain, and smallest under global ambient, where SoupX, whose model assumes a global ambient profile, performs comparably.

F Transcript-level diffusion simulation

The spike-in benchmark injects contamination at the count level. We complement it with a transcript-level simulation in which contamination arises purely from tissue geometry. Individual transcripts diffuse between cells, no cell-cluster label enters the simulator, transcripts are transferred rather than added, and the labels given to each method are inferred from the contaminated data.

F.1 Simulation design

Data. We use the Xenium_Breast transcript table and segmented cell boundaries (Janesick et al., 2023) and select a spatially contiguous region of 33,640 cells. We retain the 6,363,187 transcripts from the 313 gene-targeting probes (excluding negative-control probes, negative-control codewords, antisense and blank codewords) with quality value at least 20 that are assigned to a cell.

Diffusion. Each transcript is displaced from its recorded position by an isotropic Gaussian kernel with per-axis standard deviation $\sigma \in \{0.25, 0.5, 1, 2\} \times D$, where $D = 13.4\,\mu$m is the median cell diameter derived from the segmentation boundaries. We additionally use a heavy-tailed Student-t kernel with $\nu = 3$ degrees of freedom (scaled by σ). A displaced transcript is assigned to the cell whose segmentation polygon contains its new position and discarded if it lands outside all cells, so source cells lose the transcripts they contribute.

Ground truth. Let $\mathcal{T}_i$ be the transcripts assigned to cell i after diffusion, $\mathcal{R}_i \subseteq \mathcal{T}_i$ those originating in another cell, and $\mathcal{R}_i^{\times} \subseteq \mathcal{R}_i$ those originating in a cell of a different type under the published Xenium annotation of Janesick et al. (2023), which is used only for evaluation and is never provided

Table 5: Removal-magnitude accuracy on five datasets across six contamination mechanisms (one on-model, five off-model). relMAE (mean ± std across the non-zero contamination levels). Lower is better; best value per row in bold; values within 0.005 of the best are also bolded.

Dataset	Mechanism	**DeSpotX**	SoupX	DecontX	ResolVI	SpaceBender
Xenium_Breast	Local cross-cluster	**0.514 ± 0.075**	0.847 ± 0.096	0.826 ± 0.071	0.870 ± 0.218	0.694 ± 0.145
	Global ambient	**0.663 ± 0.033**	**0.664 ± 0.073**	0.696 ± 0.122	0.779 ± 0.096	0.772 ± 0.103
	Segmentation spillover	**0.705 ± 0.016**	0.899 ± 0.088	0.850 ± 0.038	0.731 ± 0.110	**0.707 ± 0.091**
	Long-range diffusion	**0.567 ± 0.042**	0.682 ± 0.043	0.800 ± 0.075	0.796 ± 0.132	0.701 ± 0.144
	Anisotropic diffusion	**0.527 ± 0.064**	0.679 ± 0.026	0.822 ± 0.068	0.789 ± 0.140	0.679 ± 0.146
	Vertical overlap	**0.677 ± 0.026**	0.870 ± 0.072	0.890 ± 0.048	0.803 ± 0.201	0.688 ± 0.153
MERFISH_Brain	Local cross-cluster	**0.640 ± 0.230**	0.842 ± 0.048	0.834 ± 0.106	1.268 ± 0.471	1.082 ± 0.369
	Global ambient	**0.759 ± 0.060**	**0.756 ± 0.077**	0.772 ± 0.158	1.280 ± 0.468	1.087 ± 0.476
	Segmentation spillover	**0.678 ± 0.097**	0.900 ± 0.038	0.750 ± 0.100	0.995 ± 0.480	0.860 ± 0.366
	Long-range diffusion	**0.648 ± 0.192**	0.824 ± 0.060	0.834 ± 0.113	1.378 ± 0.576	1.138 ± 0.567
	Anisotropic diffusion	**0.639 ± 0.226**	0.841 ± 0.045	0.839 ± 0.105	1.290 ± 0.523	1.059 ± 0.530
	Vertical overlap	**0.749 ± 0.129**	0.883 ± 0.050	0.859 ± 0.092	1.324 ± 0.605	0.862 ± 0.345
CosMx_NSCLC	Local cross-cluster	**0.568 ± 0.165**	0.683 ± 0.020	0.834 ± 0.115	0.834 ± 0.240	0.705 ± 0.058
	Global ambient	**0.677 ± 0.088**	**0.679 ± 0.110**	0.693 ± 0.152	0.686 ± 0.129	0.823 ± 0.035
	Segmentation spillover	**0.629 ± 0.074**	0.896 ± 0.040	0.794 ± 0.072	0.637 ± 0.124	0.739 ± 0.047
	Long-range diffusion	**0.630 ± 0.168**	0.833 ± 0.050	0.814 ± 0.110	0.711 ± 0.157	0.725 ± 0.056
	Anisotropic diffusion	**0.599 ± 0.183**	0.845 ± 0.055	0.825 ± 0.108	0.696 ± 0.152	0.692 ± 0.062
	Vertical overlap	**0.653 ± 0.161**	0.886 ± 0.050	0.863 ± 0.088	0.704 ± 0.212	0.693 ± 0.039
Xenium5k_Breast	Local cross-cluster	**0.472 ± 0.079**	0.623 ± 0.039	0.658 ± 0.029	0.478 ± 0.022	0.810 ± 0.039
	Global ambient	**0.380 ± 0.006**	0.395 ± 0.011	0.653 ± 0.032	**0.385 ± 0.001**	0.959 ± 0.002
	Segmentation spillover	**0.325 ± 0.010**	0.359 ± 0.021	0.665 ± 0.018	**0.322 ± 0.007**	0.924 ± 0.014
	Long-range diffusion	**0.311 ± 0.006**	0.317 ± 0.017	0.716 ± 0.020	0.325 ± 0.018	0.844 ± 0.006
	Anisotropic diffusion	**0.270 ± 0.013**	0.313 ± 0.010	0.715 ± 0.019	0.283 ± 0.011	0.796 ± 0.016
	Vertical overlap	**0.286 ± 0.020**	0.429 ± 0.135	0.735 ± 0.006	0.299 ± 0.021	0.744 ± 0.007
Stereo-seq_E14	Local cross-cluster	**0.782 ± 0.130**	0.882 ± 0.126	0.998 ± 0.363	0.841 ± 0.309	1.590 ± 1.028
	Global ambient	**0.830 ± 0.077**	**0.826 ± 0.139**	0.978 ± 0.232	0.879 ± 0.218	1.599 ± 0.000
	Segmentation spillover	**0.728 ± 0.054**	0.735 ± 0.011	0.816 ± 0.124	**0.729 ± 0.000**	2.869 ± 0.000
	Long-range diffusion	**0.772 ± 0.119**	0.783 ± 0.049	0.981 ± 0.205	1.060 ± 0.000	2.545 ± 0.000
	Anisotropic diffusion	**0.790 ± 0.116**	0.840 ± 0.046	0.984 ± 0.189	0.885 ± 0.235	2.481 ± 0.000
	Vertical overlap	**0.827 ± 0.136**	0.892 ± 0.056	0.975 ± 0.195	0.980 ± 0.000	2.372 ± 0.000

to any method. We define

$$\varepsilon_i^* = \frac{|\mathcal{R}_i|}{|\mathcal{T}_i|}, \qquad \alpha_i = \frac{|\mathcal{R}_i^{\times}|}{|\mathcal{R}_i|}. \tag{23}$$

Methods. We evaluate DeSpotX, the four baselines, and DenoIST (Kwok et al., 2025), which requires transcript coordinates and can therefore be run in this setting. To isolate the contribution of the learned model, we also evaluate a plug-in estimator of Lemma 2 that uses χ_i from Eq. 3, estimates the contamination fraction from anchor counts,

$$\hat{\varepsilon}_i^{\text{plug}} = \min\Big(1,\ \frac{\sum_{g\in\mathcal{A}_{t_i}} x_{i,g}}{d_i \sum_{g\in\mathcal{A}_{t_i}} \chi_{i,g}}\Big), \tag{24}$$

and reconstructs the native counts by the Lemma 2(ii) inversion, $\hat{\phi}_i = \big[(x_i/d_i - \hat{\varepsilon}_i^{\text{plug}}\chi_i)/(1 - \hat{\varepsilon}_i^{\text{plug}})\big]_+$, equivalently removing the expected contamination $\hat{n}_i = \big[x_i - d_i\,\hat{\varepsilon}_i^{\text{plug}}\chi_i\big]_+$. Two variants stabilize this estimate by averaging it over the $K = 12$ spatial neighbors of cell i or by shrinking it toward the dataset mean, $0.3\,\hat{\varepsilon}_i^{\text{plug}} + 0.7\,\langle\hat{\varepsilon}^{\text{plug}}\rangle$. Methods that use cell-cluster labels receive Leiden clusters computed on the contaminated counts (19-21 clusters per configuration).

F.2 Results

Table 6 summarizes the contamination produced at each diffusion scale. Mean contamination ranges from 7.5% to 17.9%, of which 42-51% is cross-type.

Detection. DeSpotX attains the highest cross-type AUROC at every diffusion scale and under the heavy-tailed kernel (Table 7), exceeding the best baseline (ResolVI) by 0.052-0.065 and staying between 0.922 and 0.935 as σ grows from $0.25D$ to $2D$; SpaceBender declines to 0.805 at $2D$ (0.798 under Student-t) and DenoIST reaches 0.691-0.741. DeSpotX exceeds the plug-in by 0.004-0.007

Table 6: Contamination statistics of the transcript-level diffusion simulation on Xenium_Breast (mean across cells; $\langle\alpha\rangle$ weighted by per-cell contamination, cross-type defined by the published annotation).

Kernel	σ	$\langle\varepsilon^*\rangle$	$\langle\alpha\rangle$
Gaussian	$0.25D$	0.075	0.416
	$0.5D$	0.126	0.429
	$1D$	0.164	0.452
	$2D$	0.179	0.490
Student-t	$2D$	0.179	0.506

Table 7: Cross-type AUROC in the transcript-level diffusion simulation on Xenium_Breast, with cross-type contamination defined by the independent published annotation. Higher is better; best value per row in bold.

Kernel	σ	**DeSpotX**	Plug-in	SoupX	DecontX	ResolVI	SpaceBender	DenoIST
Gaussian	$0.25D$	**0.922**	0.916	0.516	0.639	0.868	0.832	0.691
	$0.5D$	**0.927**	0.923	0.541	0.649	0.870	0.836	0.709
	$1D$	**0.933**	0.926	0.523	0.665	0.868	0.817	0.729
	$2D$	**0.935**	0.928	0.542	0.682	0.879	0.805	0.741
Student-t	$2D$	**0.934**	0.929	0.535	0.706	0.882	0.798	0.741

Table 8: Removal-magnitude accuracy in the transcript-level diffusion simulation on Xenium_Breast. relMAE on all contaminated entries, including within-type contamination. Lower is better; best value per row in bold; values within 0.005 of the best are also bolded.

Kernel	σ	**DeSpotX**	SoupX	DecontX	ResolVI	SpaceBender	DenoIST
Gaussian	$0.25D$	**0.735**	0.969	0.987	1.226	1.005	1.216
	$0.5D$	**0.674**	0.929	0.915	0.951	0.796	1.108
	$1D$	**0.639**	0.964	0.872	0.811	0.685	1.033
	$2D$	**0.613**	0.936	0.821	0.788	0.669	0.988
Student-t	$2D$	**0.599**	0.947	0.816	0.778	0.675	0.982

Table 9: relMAE of DeSpotX and the network-free plug-in estimator of Lemma 2 (inversion in part (ii)), without stabilization and with the contamination fraction smoothed over spatial neighbors or shrunk toward the dataset mean. Lower is better; best value per row in bold; values within 0.005 of the best are also bolded.

Kernel	σ	**DeSpotX**	Plug-in	Plug-in (smooth)	Plug-in (shrink)
Gaussian	$0.25D$	**0.735**	1.321	1.305	1.280
	$0.5D$	**0.674**	1.098	1.083	1.051
	$1D$	**0.639**	0.932	0.912	0.891
	$2D$	**0.613**	0.879	0.858	0.838
Student-t	$2D$	**0.599**	0.887	0.873	0.848

AUROC. As Lemma 2 predicts, χ_i with the anchor constraints suffices to *detect* which counts are contamination.

Removal magnitude. DeSpotX attains the lowest relMAE in every setting (Table 8). The plug-in reaches 0.88-1.32, and even its best variant, with shrinkage, stays 0.225-0.545 above DeSpotX (Table 9). Accurate removal magnitude therefore requires the learned model, not only the anchor-based estimate.

G ANCHOR-STRATIFIED EVALUATION

Anchor positions are (t, g) pairs with low native expression, so an observed count there is likely contamination. If most contaminated entries fell on anchor positions, a high AUROC could come from the anchor mask alone rather than from the learned decomposition. We test whether this is the case by computing AUROC separately on anchor and non-anchor positions, and by measuring the fraction of contaminated entries and of removed counts that fall on anchor positions (Table 10); if the anchor mask drove the benchmark, both fractions would be large.

Contribution of the anchor mask. Anchor positions cover 43% of the matrix on `Xenium_Breast` and 37% on `MERFISH_Brain`, yet carry only 9.5-17.6% of the positive labels and account for 16.2-39.4% of the removed count mass in all twelve settings (Table 10). The remaining 82-91% of contaminated entries lie on non-anchor positions, where DeSpotX detects contamination without the anchor mask. The benchmark performance is therefore driven mainly by the learned decomposition rather than the anchor mask.

Restricted to non-anchor positions, DeSpotX attains AUROC of 0.841-0.932, 0.016-0.051 below its overall AUROC but still above the best baseline's overall AUROC in Table 2 in 11 of 12 settings.

Table 10: Anchor-stratified evaluation on Xenium_Breast and MERFISH_Brain across six contamination mechanisms. AUROC is computed separately on anchor positions ($A_{t_i,g} = 1$) and non-anchor positions ($A_{t_i,g} = 0$), together with the share of positive contamination labels and the share of removed count mass that fall on anchor positions (mean across the five non-zero contamination levels). Non-anchor AUROC, which excludes the positions where the anchor mask can contribute, is in bold.

		AUROC ↑			Anchor share	
Dataset	Mechanism	all	anchor	non-anchor	(i) labels	(ii) mass
Xenium_Breast	Local cross-cluster	0.955	0.979	**0.930**	0.170	0.281
	Global ambient	0.897	0.976	**0.846**	0.173	0.345
	Segmentation spillover	0.904	0.980	**0.853**	0.150	0.394
	Long-range diffusion	0.929	0.978	**0.895**	0.176	0.307
	Anisotropic diffusion	0.954	0.979	**0.928**	0.170	0.284
	Vertical overlap	0.912	0.980	**0.869**	0.126	0.347
MERFISH_Brain	Local cross-cluster	0.948	0.966	**0.932**	0.111	0.162
	Global ambient	0.876	0.961	**0.841**	0.095	0.210
	Segmentation spillover	0.902	0.969	**0.858**	0.168	0.276
	Long-range diffusion	0.921	0.964	**0.900**	0.105	0.182
	Anisotropic diffusion	0.947	0.966	**0.930**	0.111	0.165
	Vertical overlap	0.906	0.968	**0.871**	0.117	0.230

Variation across mechanisms. Anchor AUROC is nearly constant across mechanisms (0.976-0.980 and 0.961-0.969), while non-anchor AUROC spans 0.846-0.930 and 0.841-0.932, so most of the variation in the aggregate metric comes from the non-anchor positions. On `Xenium_Breast`, moving from the on-model mechanism to global ambient costs 0.058 in aggregate AUROC ($0.955 \rightarrow 0.897$) but 0.084 on non-anchor positions ($0.930 \rightarrow 0.846$); the corresponding figures on `MERFISH_Brain` are 0.072 and 0.091. This degradation follows spatial locality. Performance is highest under local cross-cluster and anisotropic diffusion, intermediate under long-range diffusion and vertical overlap, and lowest under segmentation spillover and global ambient, consistent with the estimator in Eq. 3. Anchor positions thus behave as a mechanism-independent component, and the non-anchor positions isolate the contribution of the spatial contamination estimator.

H ABLATION STUDIES

We perform ablation studies on `Xenium_Breast` and `MERFISH_Brain` to assess the contribution of each architectural component. We compare the full DeSpotX model against four ablations, each of which removes one component while keeping the rest of the architecture intact.

Ablation conditions. Each ablation removes a single component from the full model.

- *w/o spatial*: spatial information is removed. The GAT layer is replaced by an MLP applied to each cell's own counts, and χ_i is replaced by a single dataset-level ambient profile shared across cells.
- *w/o GAT*: the GAT layer is replaced by an MLP applied to each cell's own counts, while χ_i is still computed from cross-cluster spatial neighbors (Eq. 3).
- *w/o cluster-mask*: cluster conditioning is removed. The cross-cluster mask in χ_i (Eq. 3) is dropped so all neighbors contribute regardless of cluster identity, and the cluster-specific decoder bias b_{t_i} is no longer warm-started from per-cluster expression statistics.
- *w/o anchor*: the anchor penalty is removed. With the constraint removed, the negative-binomial likelihood and the simplex constraint on ϕ remain, corresponding to the unconstrained setting of Lemma 1.

Results. Figure 6 shows that the full model achieves the highest AUROC on both datasets, confirming that each architectural component contributes to DeSpotX's performance. Removing spatial information produces the largest median drop on Xenium_Breast and the second largest on MERFISH_Brain, confirming that the spatial graph encoder and the local contamination estimator are central to DeSpotX's accuracy. Replacing the GAT layer alone also degrades AUROC on both datasets, though less than removing all spatial information, since χ_i still captures the local neighborhood. Removing the identifiability constraints drops AUROC below 0.90 on MERFISH_Brain and makes it vary widely across noise levels on Xenium_Breast, consistent with Lemma 1, which states that without external constraints the $(\phi, \varepsilon, \chi)$ decomposition is not uniquely determined. Removing the cluster-mask also degrades AUROC on both datasets, with a smaller but consistent effect, since the mask supplies cell-cluster information that helps separate native from contaminating signal.

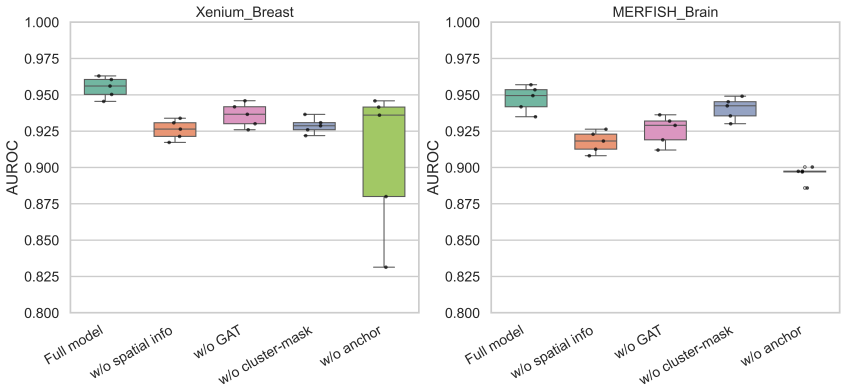

Figure 6: Ablation studies on `Xenium_Breast` (left) and `MERFISH_Brain` (right), reporting AUROC across the five non-zero noise levels. Each box summarizes the distribution of values across noise levels.

I ROBUSTNESS STUDIES

We assess DeSpotX's robustness to perturbations of the adaptive-anchor threshold κ, random perturbation of the anchor mask, noise in the cell-cluster labels, the number of spatial neighbors K, the cell-cluster annotation method, and the source of the contamination profile χ_i. For each factor, we

run the spike-in benchmark on `Xenium_Breast` and `MERFISH_Brain` with the perturbed input while keeping the rest of the configuration at its default.

Robustness conditions. Each condition perturbs a single input from the default configuration.

- *Adaptive-anchor threshold*: κ is varied across $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. Larger κ produces a more permissive threshold and anchors more genes per cluster.
- *Anchor mask perturbation*: a uniform fraction $\{5\%, 10\%, 20\%\}$ of entries in the default $\kappa = 0.3$ anchor mask is randomly flipped (both $0 \to 1$ and $1 \to 0$).
- *False anchor injection*: for each cluster, the 20 genes with highest detection rate among genes detected in over 15% of cells are added to the anchor set. The experiment is run at noise levels $\bar{\varepsilon} \in \{20\%, 40\%\}$ on Xenium_Breast and MERFISH_Brain.
- *Anchor reduction*: the anchor set is reduced to $k \in \{1, 3, 5, 10, 25\}$ genes per cluster, selected as the k genes with lowest detection rate. The experiment is run at noise levels $\bar{\varepsilon} \in \{20\%, 40\%\}$ on Xenium_Breast and MERFISH_Brain.
- *Cell-cluster label perturbation*: the published cell-cluster annotation is perturbed by reassigning a random fraction $\{5\%, 10\%, 20\%, 30\%\}$ of cells to a different cluster, sampled uniformly. Both the cross-cluster mask in χ_i (Eq. 3) and the anchor mask (Eq. 20) are derived from the cluster annotation.
- *Number of spatial neighbors*: K is varied across $\{8, 12, 24\}$. K sets the number of spatial neighbors used by the GAT graph and the contamination estimator χ_i (Eq. 3).
- *Cell-cluster annotation method*: the published cell-cluster annotation is replaced by Leiden clustering at three target cluster counts $N \in \{15, 25, 30\}$. The Leiden resolution is binary-searched per dataset to hit each target.
- *Contamination profile source*: χ_i is estimated from neighbors' observed counts (estimated, the default), from their uncontaminated counts (clean), or replaced by the profile injected by the simulator (true), under the local cross-cluster mechanism.

Results. DeSpotX is robust to perturbations of the anchor mask. Across $\kappa \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, AUROC moves by at most 0.011 on either dataset (Figure 7). When up to 20% of mask entries are flipped, AUROC decreases by less than 0.01 on either dataset (Figure 8). Adding 20 actively expressed genes per cluster as false anchors removes these genes from the clusters where they are anchored (Eq. 7). On the remaining genes, AUROC changes by less than 0.002 and native retention decreases by 2 to 6 percentage points on both datasets. Reducing the anchor set to a single gene per cluster keeps AUROC within 0.007 of the full set (Figure 2b,c). These results confirm that identifiability is maintained by a small number of correct anchors, consistent with Lemma 2.

Figure 9 shows that AUROC values remain close to the default values as the cell-cluster label perturbation rate increases from 0% to 30%. AUROC drops by less than 0.005 on both datasets, indicating that DeSpotX is not sensitive to label noise. Figure 10 shows that all AUROC values remain close to the default values across $K \in \{8, 12, 24\}$ on both datasets, indicating that DeSpotX is robust to the choice of K in this range.

When the published cell-cluster annotation is replaced by Leiden clustering, AUROC drops by less than 0.03 on both datasets (Figure 11). The performance is consistent across the three Leiden cluster counts $N \in \{15, 25, 30\}$, indicating that DeSpotX is not sensitive to the specific cluster count chosen by Leiden.

Replacing the estimated χ_i with the clean or true profile increases AUROC by less than 0.008 on both datasets (Figure 12), indicating that DeSpotX is robust to errors in estimating χ_i from contaminated neighbors.

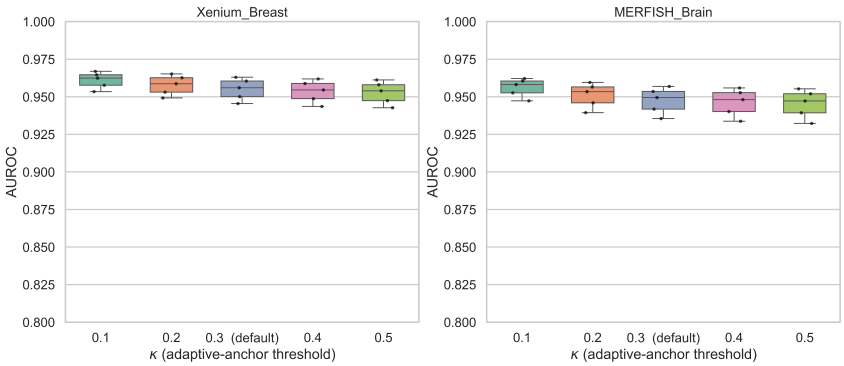

Figure 7: Robustness to the adaptive-anchor threshold κ. Each box summarizes the metric distribution across the five non-zero noise levels.

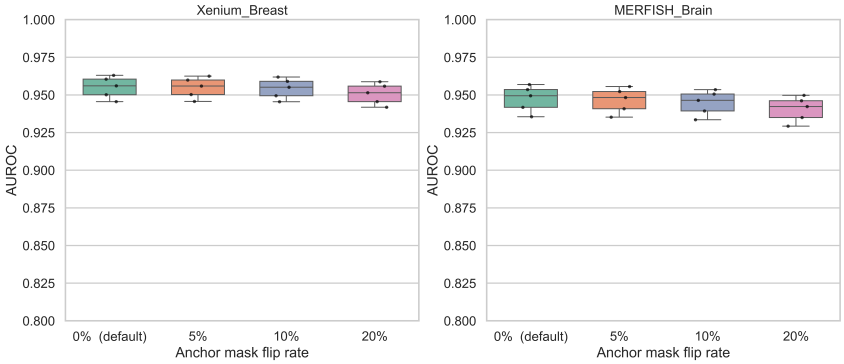

Figure 8: Robustness to random perturbation of the anchor mask. Each box summarizes the metric distribution across the five non-zero noise levels.

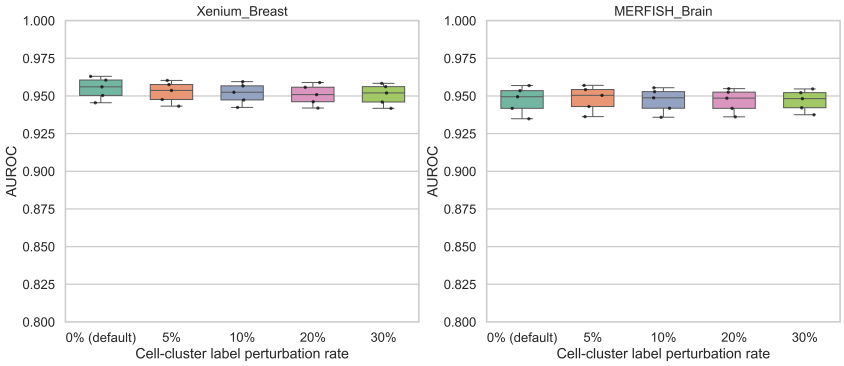

Figure 9: Robustness to cell-cluster label perturbation. Each box summarizes the metric distribution across the five non-zero noise levels.

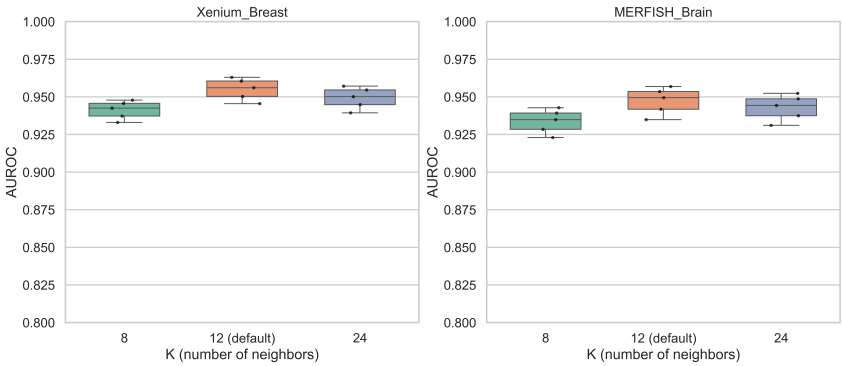

Figure 10: Robustness to the number of spatial neighbors K. Each box summarizes the metric distribution across the five non-zero noise levels.

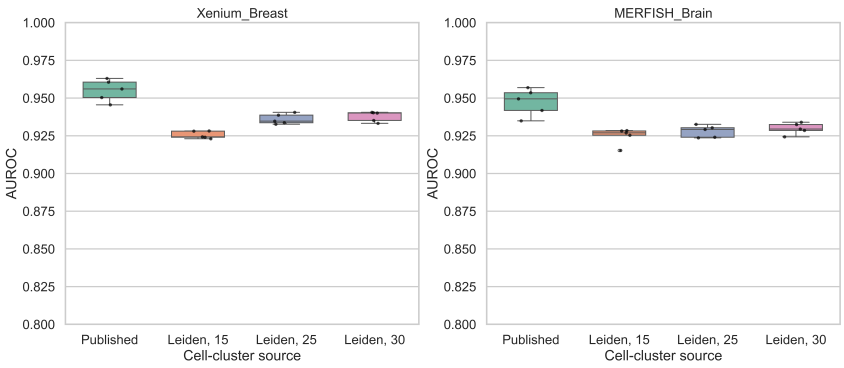

Figure 11: Robustness to the cell-cluster annotation method. Each box summarizes the metric distribution across the five non-zero noise levels.

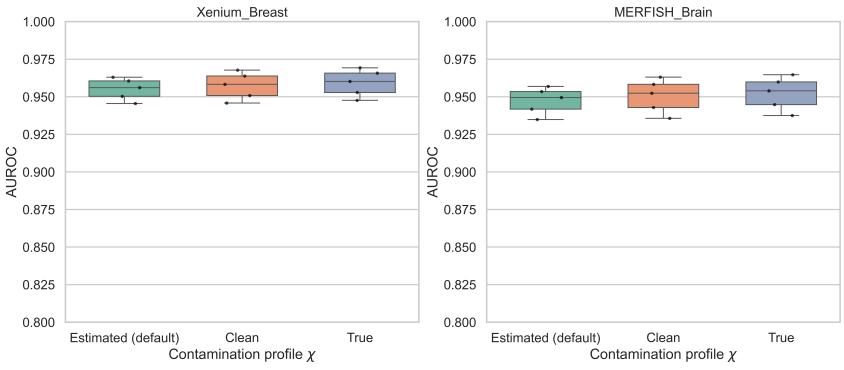

Figure 12: Robustness to the source of the contamination profile χ_i. Each box summarizes AUROC across the five non-zero contamination levels.

J CONSISTENCY ACROSS CONTAMINATION LEVELS AND RANDOM SEEDS

We examine how DeSpotX's performance varies with the noise level ε on each of the five spike-in benchmark datasets.

AUROC of DeSpotX across noise levels. Figure 13 shows DeSpotX's AUROC at each noise level on each dataset. Across all five datasets, AUROC varies by at most 0.025 between $\bar{\varepsilon} = 5\%$

and $\bar{\varepsilon} = 40\%$. On `Xenium_Breast` and `MERFISH_Brain`, AUROC increases slightly with $\bar{\varepsilon}$, since stronger contamination produces clearer signals that are easier to identify against the native expression background. DeSpotX therefore performs consistently across contamination levels.

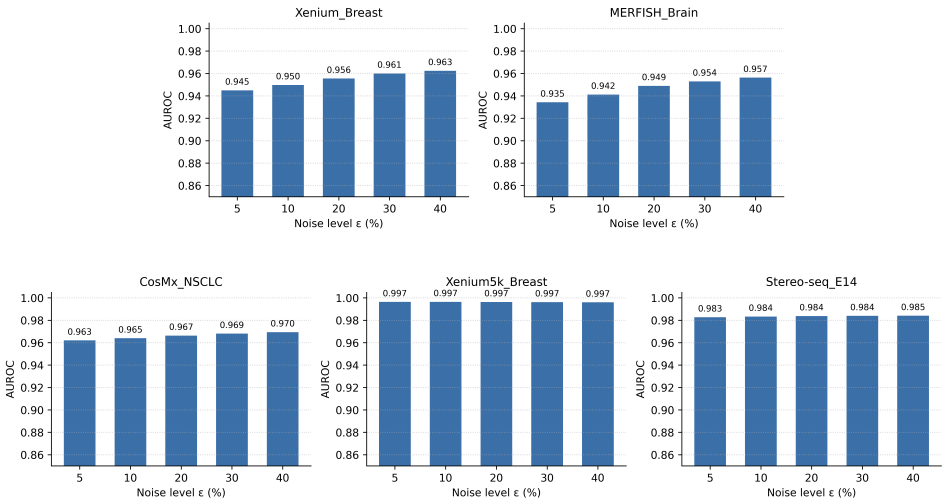

Figure 13: DeSpotX's AUROC on each of the five spike-in benchmark datasets at five contamination levels $\varepsilon \in \{5, 10, 20, 30, 40\}\%$.

We retrained DeSpotX with three random seeds on Xenium_Breast and MERFISH_Brain at $\bar{\varepsilon} = 20\%$ for all six mechanisms (Table 11). The standard deviation across seeds is at most 0.0007 in AUROC and 0.018 in relMAE, showing that its advantage over the baselines is not an artifact of a particular seed.

Table 11: DeSpotX across three random seeds on Xenium_Breast and MERFISH_Brain at $\bar{\varepsilon} = 20\%$ (mean $\pm$ std across seeds).

Dataset	Mechanism	AUROC	relMAE
Xenium_Breast	Local cross-cluster	0.9565 ± 0.0003	0.494 ± 0.007
	Global ambient	0.8975 ± 0.0004	0.638 ± 0.006
	Segmentation spillover	0.9048 ± 0.0005	0.698 ± 0.003
	Long-range diffusion	0.9295 ± 0.0001	0.548 ± 0.007
	Anisotropic diffusion	0.9557 ± 0.0003	0.502 ± 0.004
	Vertical overlap	0.9131 ± 0.0005	0.668 ± 0.004
MERFISH_Brain	Local cross-cluster	0.9496 ± 0.0002	0.555 ± 0.018
	Global ambient	0.8766 ± 0.0003	0.719 ± 0.002
	Segmentation spillover	0.9014 ± 0.0005	0.628 ± 0.003
	Long-range diffusion	0.9218 ± 0.0007	0.583 ± 0.005
	Anisotropic diffusion	0.9484 ± 0.0003	0.554 ± 0.018
	Vertical overlap	0.9052 ± 0.0007	0.695 ± 0.004

K VALIDATION OF THE WITHIN-CLUSTER BOUND

We test Lemma 3 (Appendix B.2) on a spike-in simulation in which each cell receives contamination from a mixture of its cross-cluster and same-cluster spatial neighbors, weighted α and $1 - \alpha$ respectively, with per-cell rate sampled from a Beta distribution centered at $\bar{\varepsilon} = 0.2$. We vary $\alpha \in \{0.3, 0.5, 0.7, 0.9, 1.0\}$ on `Xenium_Breast` and `MERFISH_Brain`.

AUROC on cross-cluster contamination labels remains at or above 0.93 across all α on both datasets (Figure 14a), confirming $\hat{\varepsilon}_i$ tracks the cross-cluster rate. The per-cell native-profile error grows linearly with the scaling factor β_i, with slopes 0.015 on `Xenium_Breast` and 0.010 on `MERFISH_Brain` (Figure 14b), well below the empirical $\langle \sigma^{\text{within}} \rangle$ of 0.057 and 0.063 on the same datasets (Figure 14c). Because $\langle \sigma^{\text{within}} \rangle$ in Figure 14c is estimated from observed counts as $\|x_i/d_i - \hat{\phi}_i\|_\infty$, it includes

sampling noise and is expected to exceed the heterogeneity in Eq. 13, so the fitted slopes are consistent with Eq. 15. The intercept (≈ 0.022) reflects α-independent error sources unrelated to within-cluster contamination, including DeSpotX's removal of natural platform contamination from the unspiked reference.

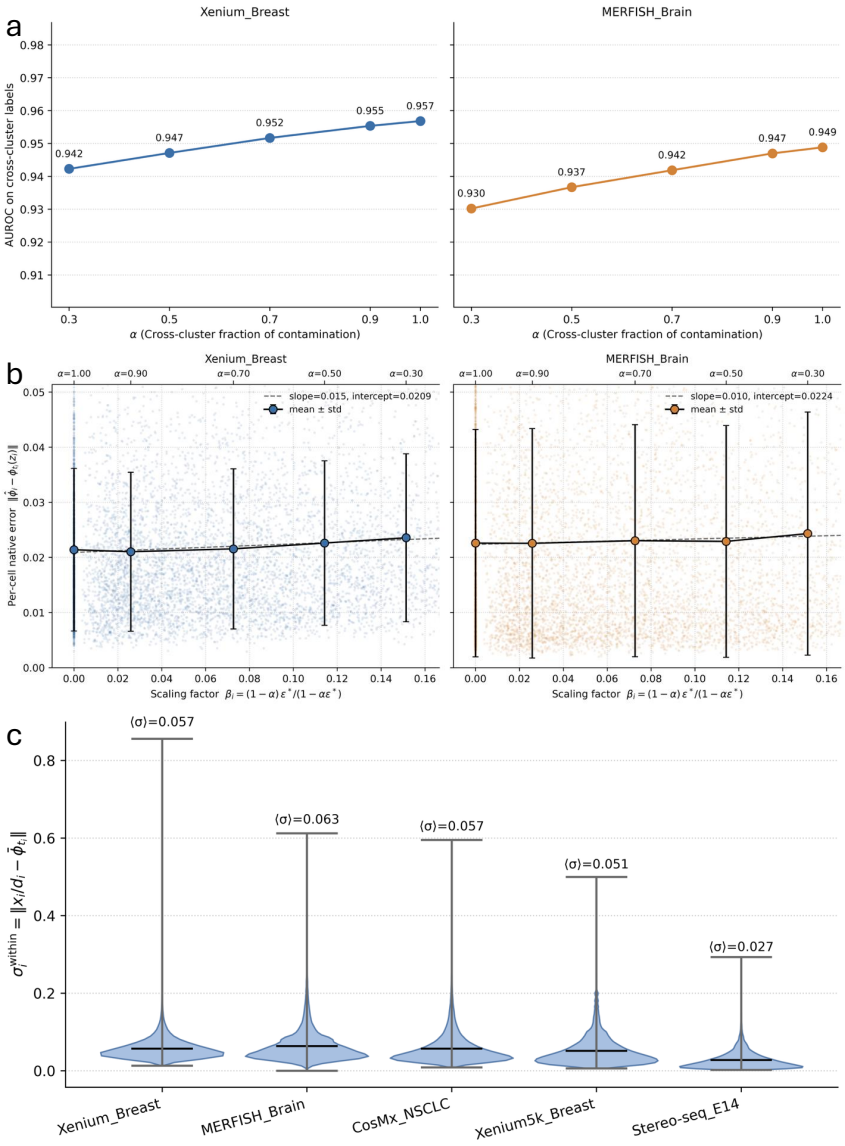

Figure 14: DeSpotX is robust to within-cluster contamination across spike-in mixtures. (a) AUROC for distinguishing cross-cluster contamination from native counts, across $\alpha \in \{0.3, 0.5, 0.7, 0.9, 1.0\}$. (b) Per-cell native-profile error $\|\hat{\phi}_i - \phi_{t_i}(z_i)\|$ versus the scaling factor $\beta_i = (1-\alpha)\varepsilon^*/(1-\alpha\varepsilon^*)$. Markers show mean $\pm$ std per α; dashed line is a linear fit. (c) Distribution of the empirical within-cluster heterogeneity across cells in all five datasets at zero injected contamination, with mean values annotated.

L RUNTIME COMPARISON

We compare end-to-end runtime across decontamination methods on `Xenium5k_Breast` (577,258 cells, 5,101 genes) and `Stereo-seq_E14` (92,928 cells, 18,582 genes), the two largest datasets used in this work (Figure 15). SoupX and DecontX run in 3 to 10 minutes. Among the deep learning-

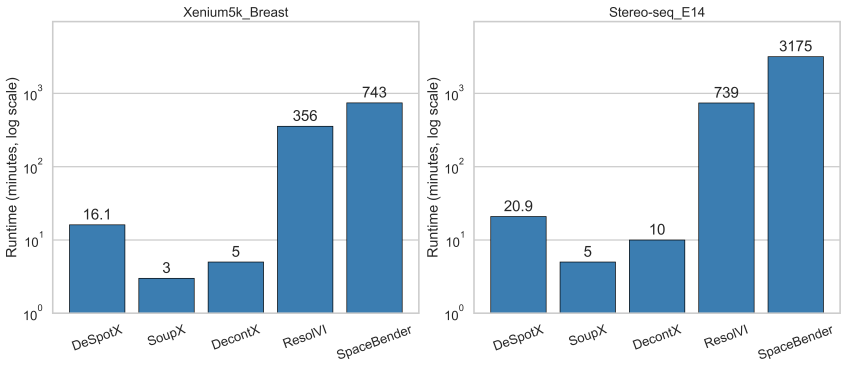

Figure 15: Runtime comparison across decontamination methods on `Xenium5k_Breast` and `Stereo-seq_E14`.

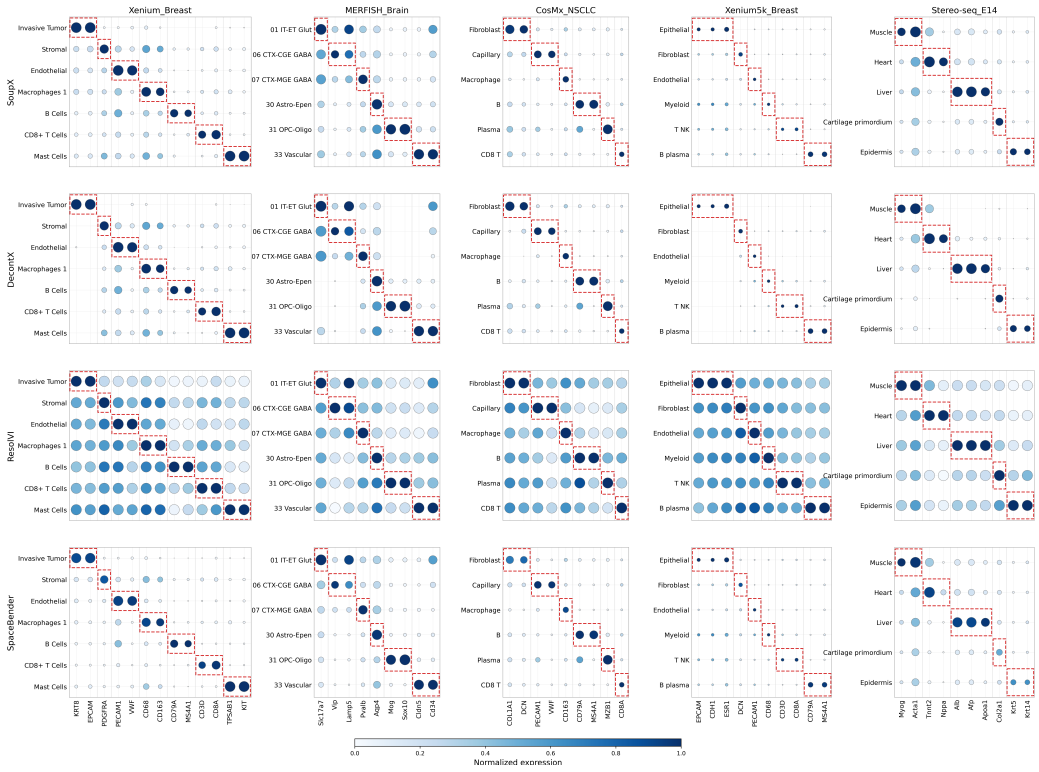

Figure 16: Marker-gene dotplots for the four baseline methods across the five datasets. Red dashed boxes mark canonical marker-cluster pairs.

based methods, DeSpotX is the fastest at 16 to 21 minutes. ResolVI and SpaceBender take 6 to 53 hours. All methods were run on a single NVIDIA Tesla T4 GPU (16 GB).

M Marker-gene specificity of baseline methods

We extend the marker-gene comparison in Figure 3a to the four baseline methods (Figure 16). SoupX produces dotplots that closely resemble the raw counts, with off-target expression largely unchanged. DecontX and SpaceBender reduce off-target expression to varying degrees across datasets, but residual signal in non-canonical clusters remains visible, particularly on the larger panels (Xenium5k_Breast,

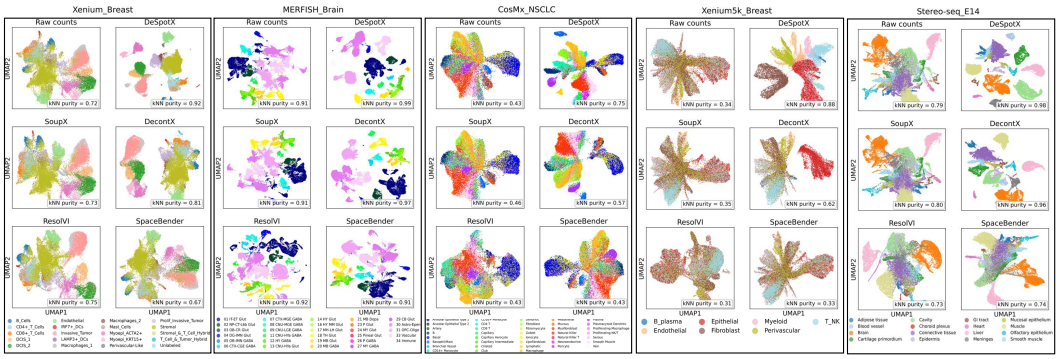

Figure 17: UMAP embeddings of raw counts and decontaminated counts from each method, across the five datasets, colored by cell cluster. kNN purity is shown in each panel.

CosMx_NSCLC). ResolVI shows little difference between canonical and non-canonical clusters. All four baselines yield lower marker-gene specificity than DeSpotX.

N UMAP EMBEDDINGS ACROSS METHODS

Figure 17 shows UMAP embeddings of the raw counts and the decontaminated counts from DeSpotX and the four baselines across the five datasets. Each panel reports the kNN purity at $k = 20$, defined as the fraction of each cell's k nearest neighbors in UMAP space that share its cluster label; higher values indicate cleaner cluster separation. DeSpotX yields the highest kNN purity on every dataset and produces visibly better-separated clusters than the raw counts, while the baselines yield embeddings with less clear cluster separation.

O SPLIT-CLUSTER NEGATIVE CONTROL

To test whether DeSpotX's gains in label-based metrics reflect removed contamination rather than the input labels, we randomly split three Xenium_Breast clusters (Invasive Tumor, CD4+ T cells, macrophages) into two spatially interleaved halves, give the halves different labels, and compare them after decontamination. As a null, we decontaminate with the original labels and apply the same split afterwards. We also run DeSpotX with randomly permuted labels to test whether kNN purity increases for arbitrary labels.

Table 12: Split-cluster negative control on Xenium_Breast. Each cluster is split into two halves with different labels. DE genes is the number of genes that differ between the halves (Wilcoxon test, FDR < 0.05, $|\log_2 \mathrm{FC}| > 0.25$). AUROC measures how well a cross-validated classifier tells the halves apart (0.5 means indistinguishable), using genes detected in more than 30% of the cluster's cells (own genes) or all genes. Default gives each half its own anchor set, and shared uses one anchor set for both. Retention is the total count kept after decontamination, relative to the null.

Cluster	DE genes (split / null)	AUROC (own genes)	AUROC (all genes, default / shared anchors)	Retention
Invasive Tumor	0 / 0	0.51	0.50 / 0.50	0.95
CD4+ T	2 / 0	0.50	0.67 / 0.50	0.96
Macrophages	6 / 0	0.49	0.80 / 0.51	0.96

The halves remain indistinguishable on the cluster's own genes (AUROC 0.49-0.51), with at most 6 DE genes out of 313 and 95-96% retention relative to the null (Table 12). With permuted labels, kNN purity with respect to those labels stays at 0.153 after decontamination. The all-gene separation in CD4+ T cells and macrophages is traced to a few off-type genes near the anchor threshold, as it disappears with a shared anchor set (AUROC 0.50-0.51). This control illustrates that anchor selection can introduce label-dependent differences in off-type genes near the selection threshold.

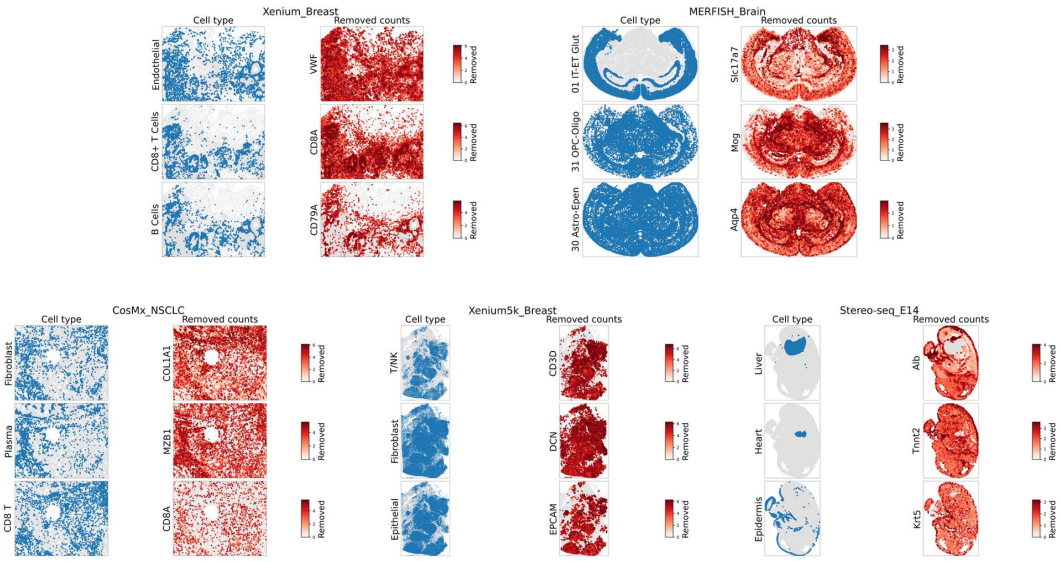

Figure 18: Spatial maps of counts removed by DeSpotX for canonical cell-cluster markers across the five datasets. For each marker, cells of the canonical expressing cluster are highlighted on the left, and the spatial distribution of removed counts is shown on the right.

Off-type genes also enter ligand-receptor analysis, so we checked the ligands behind the results in Section 4.3.3 and Appendix R (PENK and PDYN in MERFISH_Brain, PTN in Xenium_Breast). Across all 25 ligand-cluster pairs, detection rates lie at least 6.6 standard errors from the anchor threshold, and anchor status never changes under within-cluster resampling, including 50% subsamples that match the split in sample size.

P Spatial localization of removed contamination

To examine where DeSpotX removes contamination, we map the per-cell amount removed for canonical markers of representative cell clusters across the five datasets (Figure 18). The removed counts are spatially concentrated near the canonical expressing cluster, consistent with the expected pattern of ambient contamination, in which a marker's signal diffuses from cells that natively express it into nearby cells, and DeSpotX removes the diffused signal from the receiving cells. For example, VWF in `Xenium_Breast` is endothelial-specific and Mog in `MERFISH_Brain` is oligodendrocyte-specific; in both cases, DeSpotX's removal extends beyond the canonical cluster into surrounding tissue, with the highest removal in regions immediately adjacent to the canonical cells. This pattern holds across the five datasets and supports the spatial-locality assumption underlying DeSpotX's contamination model.

Q Spatial coherence of baseline methods

We extend the Moran's I comparison in Figure 4 to the four baseline methods (Figure 19). SoupX and DecontX produce scatters that closely follow the identity line, indicating that their decontamination preserves the raw spatial structure of expression but does not enhance it. SpaceBender's scatters fall on or below the diagonal across all five datasets, suggesting that its decontamination reduces spatial coherence relative to the raw counts. ResolVI shows substantially elevated Moran's I values across all five datasets, often saturating at 0.4-0.8 even for genes with near-zero raw values. This pattern reflects ResolVI's spatial-smoothing decoder, which incorporates neighbor information into each cell's expression estimate and inflates Moran's I regardless of whether the underlying signal is biological or ambient. Such smoothing produces visually coherent maps but does not separate native signal from ambient contamination. By contrast, DeSpotX (Figure 4) raises Moran's I above the

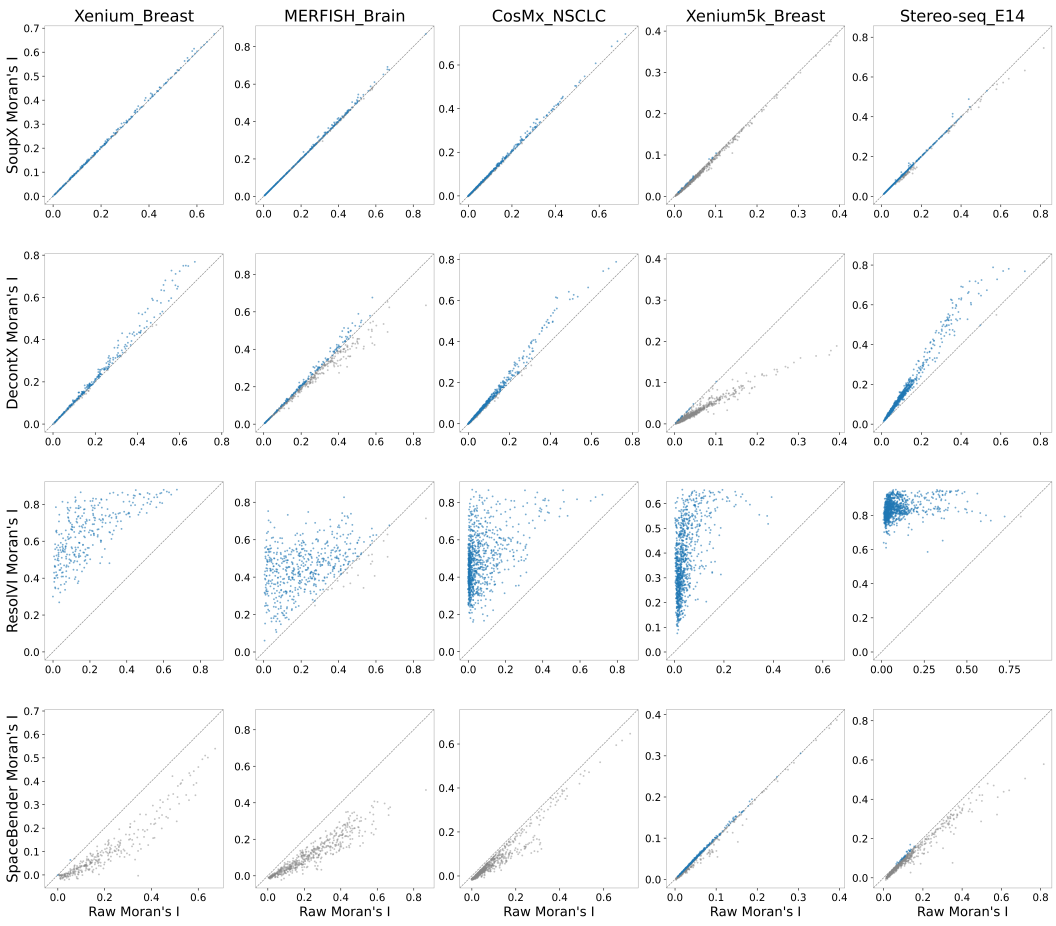

Figure 19: Per-gene Moran's I for raw and decontaminated counts from each baseline method (SoupX, DecontX, ResolVI, SpaceBender) across the five datasets. Each panel plots the raw Moran's I on the x-axis against the decontaminated Moran's I on the y-axis; the dashed line denotes equality.

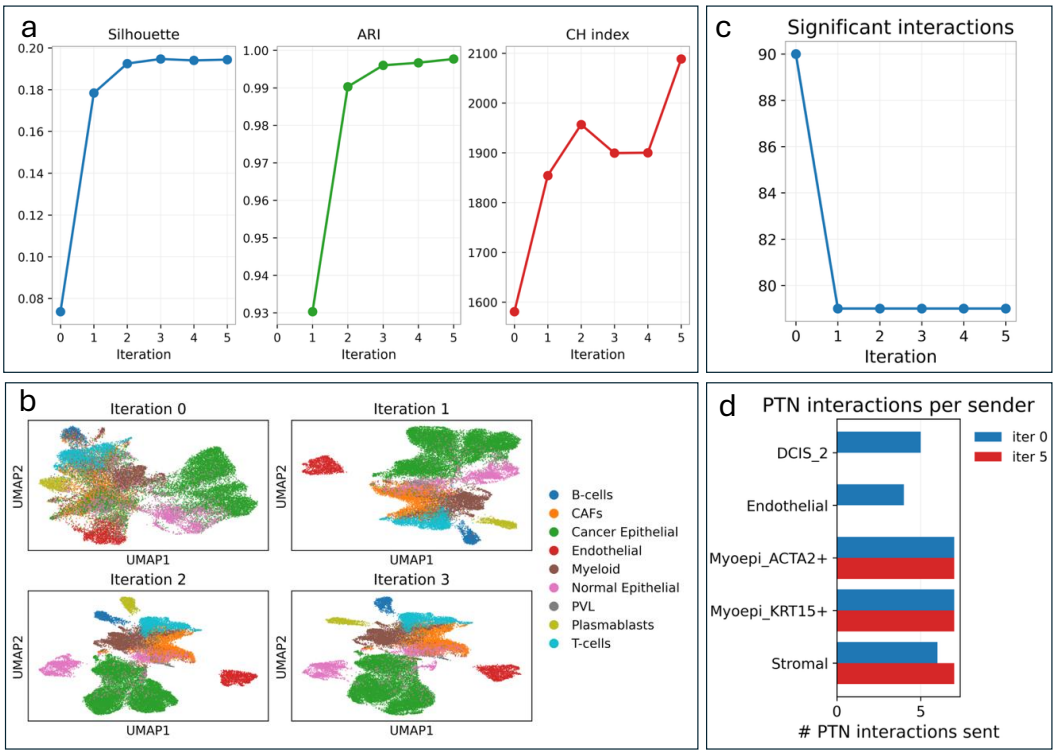

Figure 20: Iterative DeSpotX decontamination and cell-cluster annotation on `Xenium_Breast`. (a) Cluster-validity metrics across iterations. (b) UMAP embeddings of the decontaminated counts at iterations 0-3. (c) Total number of significant cell-cell interactions inferred by CellChat across iterations. (d) Number of significant PTN-pathway interactions per sender cell-cluster at iteration 0 and iteration 5.

diagonal proportionally to the raw value, indicating that improvements arise from removing local contamination rather than from spatial averaging.

R ITERATIVE DECONTAMINATION ON XENIUM_BREAST

We apply the iterative procedure described in Section 4.3.3 to the `Xenium_Breast` dataset. The CellTypist classifier (Domínguez Conde et al., 2022) for this dataset was trained on a published single-cell breast cancer atlas and covers nine broad cell-cluster categories spanning normal and malignant epithelial populations.

The cluster-validity metrics improve significantly from iteration 0 to iteration 1 and stabilize thereafter (Figure 20a), with Silhouette rising from 0.07 to 0.19 and Calinski-Harabasz increasing from 1,590 to 2,089. The ARI between consecutive iterations approaches 1.0 by iteration 2, indicating that the procedure converges to a stable annotation. Additionally, the UMAP embeddings show clearer separation of breast cell types across iterations (Figure 20b).

CellChat (Jin et al., 2021) inference shows a reduction in the number of significant interactions between iteration 0 and iteration 1, after which the network stabilizes (Figure 20c). The reduction is largely driven by removal of interactions involving the secreted ligand PTN (pleiotrophin). At iteration 0, PTN is inferred to be sent from cancer (DCIS_2) and endothelial cells; by iteration 5, these interactions are removed, while PTN signaling from myoepithelial cells, which natively express PTN in mammary tissue, is preserved (Figure 20d).