Supplementary Text S1

Section S1. Supplementary figures

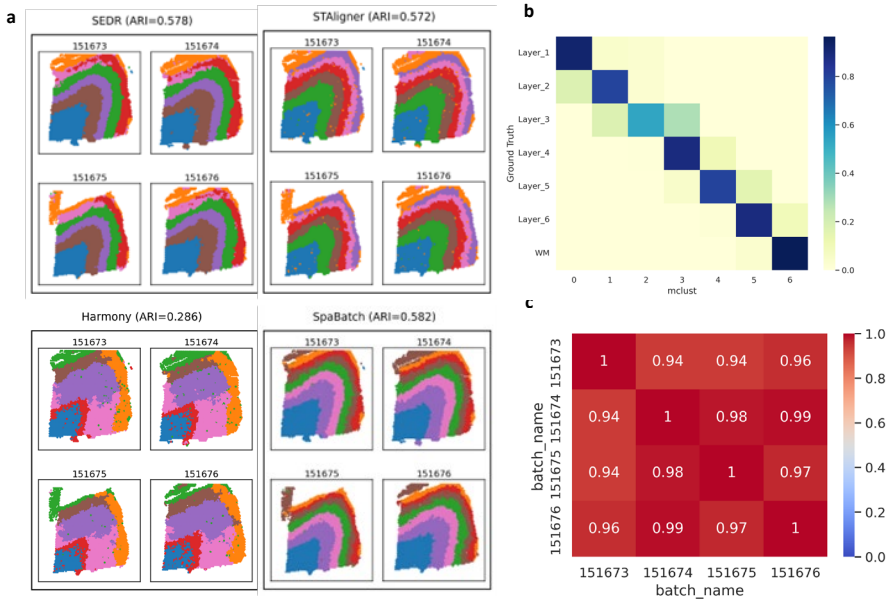

Figure S1. Supplementary DLPFC method comparisons and agreement analyses. Spatial-domain maps show representative executions for SEDR (ARI 0.578), STAligner (0.572), Harmony (0.286) and SpaBatch (0.582) across sections 151673–151676. Additional panels compare reference cortical layers with mclust clusters and summarize pairwise correlations of predicted domain proportions. The displayed scores belong to the corresponding single visualization runs; they are not six-execution averages. Agreement and composition are complementary domain-structure diagnostics and do not independently establish exact anatomical correspondence.

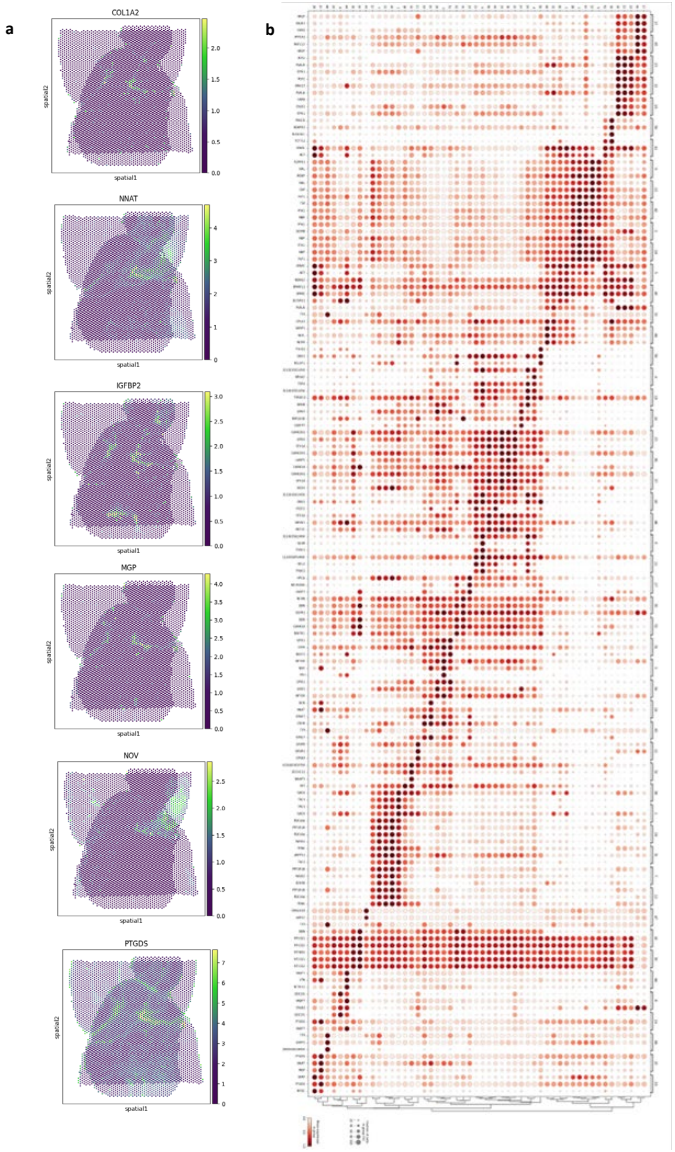

Figure S2. Supplementary mouse-brain marker-expression maps and dot plot. (a) Spatial expression of COL1A2, IGFBP2, MGP, NNAT, NOV and PTGDS in the partially overlapping mouse-brain setting. (b) Domain-associated marker dot plot. Dot size indicates the percentage of spots expressing each marker, and color indicates mean expression in each domain. These expression-derived patterns provide corroboration of spatial-domain interpretation rather than independent molecular validation. Exact correspondence retrieval and zero-overlap unmatched-anchor errors are evaluated separately in Figure S4.

Section S2. Dataset and resource information

Dataset composition, data sources and accession information, and runtime and memory summaries are provided in Tables 1, 2 and 3 of the main manuscript, respectively.

Section S3. Sensitivity and component validation

The additional evaluations distinguish fixed residual mixing from fusion-gate initialization. Residual coefficients remain fixed throughout training, whereas the gate starts from a spatial-view prior of 0.7 and is subsequently learned. The sensitivity analysis compares six graph/projection mixtures and five gate initializations, with five training seeds per setting (Table S1). These targeted comparisons retain

the original defaults; they do not establish a universally optimal setting or replace evaluation on an independent validation dataset.

The full-batch confidence-weight audit covers 30 runs and 150 saved updates (Table S2). Observed weights ranged from 0.874 to 1.007, so neither the full-batch interval of 0.3–1.2 nor the subgraph interval of 0.5–1.5 would alter these recorded values. The absence of clipping at these inspected updates supports interpreting clipping as a safeguard. This audit concerns full-batch training and does not validate the subgraph interval as an optimized choice.

Training diagnostics from one DLPFC run assess embedding effective rank and within-section tissue purity (Table S3). Effective rank decreases during training, while tissue purity remains close to its value immediately before alignment. These observations provide diagnostics of retained structure; they cannot exclude local over-smoothing or over-alignment. Component ablations in Figure S3 compare the complete model with single-graph, fixed-fusion, uniform-weight, no-triplet and joint-training alternatives. The 1000-update joint-training condition matches the total update budget of the two-stage model.

Table S1. Sensitivity to residual mixing and fusion-gate initialization.

Parameter	Setting	ARI (mean ± SD)
Residual mixing: graph / projection	0.0 / 1.0	0.400 ± 0.098
	0.3 / 0.7	0.548 ± 0.057
	0.5 / 0.5	0.593 ± 0.037
	0.7 / 0.3 (default)	0.627 ± 0.027
	0.9 / 0.1	0.619 ± 0.029
	1.0 / 0.0	0.583 ± 0.004
Fusion-gate initialization	0.1	0.506 ± 0.063
	0.3	0.536 ± 0.060
	0.5	0.615 ± 0.033
	0.7 (default)	0.624 ± 0.034
	0.9	0.626 ± 0.034

Values are mean ARI ± SD across five training seeds per setting. Residual settings are graph/projection weights; gate settings are initialization values. The residual weights remain fixed while the gate is subsequently learned.

Table S2. Full-batch confidence-weight clipping diagnostics.

Diagnostic	Result
Number of runs	30
Saved weight updates	150
Full-batch clipping interval	0.3–1.2
Observed weight range	0.874–1.007
Saved updates triggering clipping	0/150

The observed range and clipping interval are bounds, not means ± SD. No saved update triggered clipping. The separate subgraph default, 0.5–1.5, was not optimized by this audit.

Table S3. DLPFC training diagnostics.

Training checkpoint	Effective rank	Tissue purity@10
Initialization, epoch 0	10.43	74.2%
Before alignment, epoch 500	4.06	79.1%
Final, epoch 1000	4.56	78.8%

One joint-training execution with training seed 666. Effective rank has a maximum of 32. Tissue purity@10 is balanced by reference class within each section and averaged across four sections. Reference labels are used only for evaluation; no replicate SD is reported.

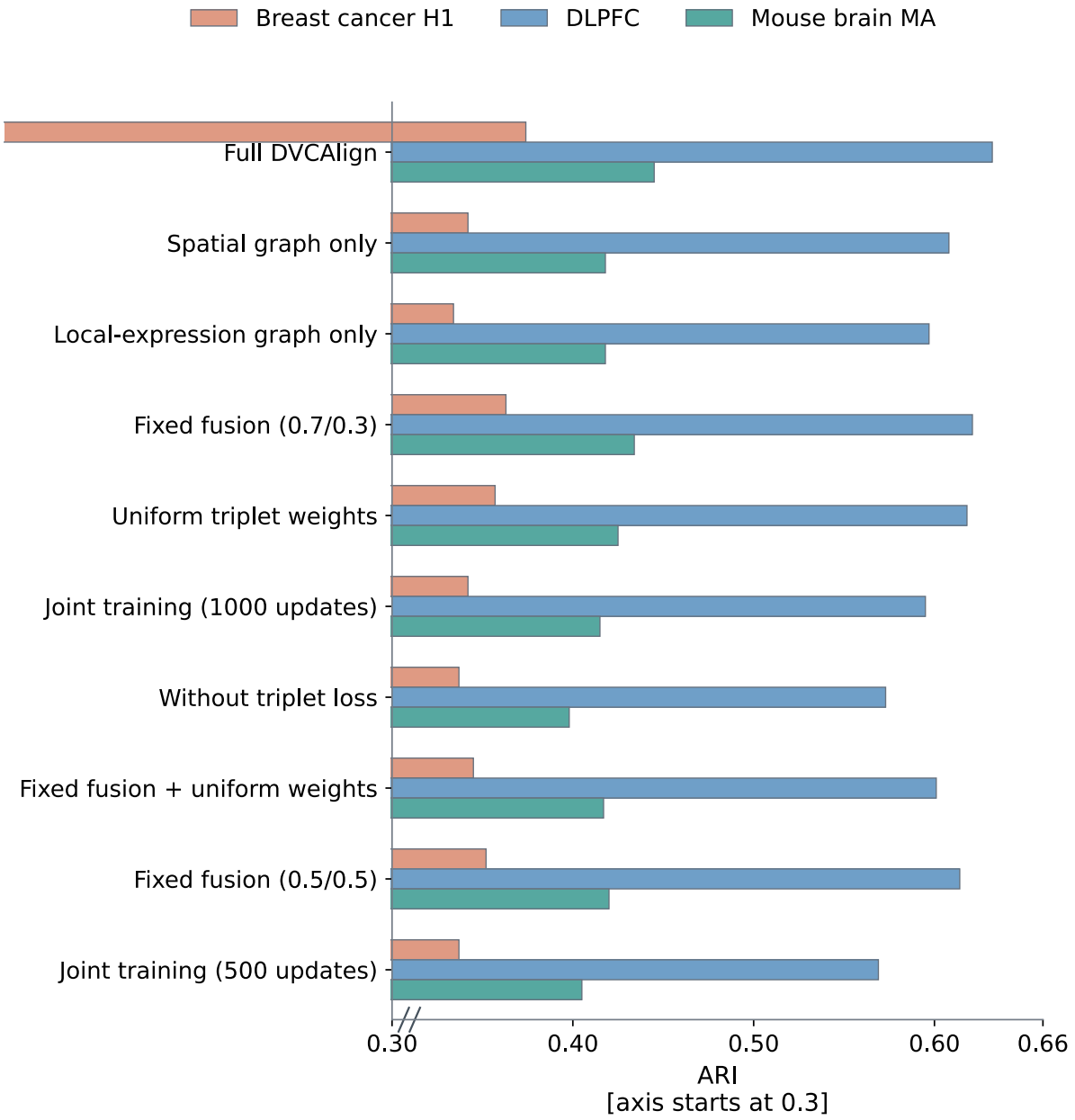

Figure S3. Component ablation results across three datasets. Touching bars compare ARI on Breast cancer H1, DLPFC and Mouse brain MA for each ablation condition. The complete model uses 500 pretraining and 500 alignment updates. Fixed-fusion conditions use the stated spatial/expression weights; uniform weights set all triplet confidence multipliers to one, and the no-triplet condition sets its coefficient to zero. The 1000-update joint condition matches the full training budget. No SD was reported for this comparison. The ARI axis starts at 0.30 and is explicitly marked; visible bar lengths are not ratios from zero.

Section S4. Controlled correspondence and confidence diagnostics

The controlled partial-overlap benchmark uses MA and MP separately to construct pseudo-section pairs along two crop directions. Within each source/crop setting, spot counts and gene panels remain fixed, and MNN settings are matched between methods. Overlap conditions comprise 100%, 75%, 50%, 25% and 0%. Figure S4 reports the available four-setting comparison at 50% overlap and the zero-overlap negative control. The known source-spot identities are evaluation references, not training labels. This design does not independently establish precise anatomical counterparts in the real MA–MP section pair.

Recall@1 is evaluated only on queries with a true counterpart, using all spots in the opposite view as candidates. At 50% overlap, absolute improvements over matched STAligner are 5.19, 4.48, 6.28 and 6.50 percentage points for MA/x4, MA/x5, MP/x4 and MP/x5. Unmatched-anchor false acceptance measures entry into MNN triplet supervision rather than final registration error. At zero overlap, no true counterparts exist and exact Recall@1 is undefined. Mean false-acceptance rates range from 29.39% to 48.53%; consistency weighting does not implement hard rejection of unmatched anchors.

The confidence evaluation uses controlled MA data and compares weighted and uniform alignment from identical pretrained states and initial triplets (Figure S5). A reinforcement-compatible event requires an initially top-decile, exact-match-incorrect pair to remain selected at epoch 900 and to increase both its raw consistency score and fused cosine similarity by more than 10^{-6} from epoch 500. Such events are descriptive evidence compatible with feedback, not proof of individual-pair causality. Confidence weights are therefore treated as relative cross-view-consistency multipliers, not calibrated probabilities of correct correspondence.

A complementary domain-separation diagnostic calculated silhouette coefficients in 2D UMAP using reference tissue labels on four DLPFC sections. The coefficients were 0.0670 before integration, 0.1782 after DVCAlign integration and 0.1775 in a same-settings, same-seed technical repeat. This diagnostic concerns tissue-domain separation in that representation, rather than exact spot-registration accuracy.

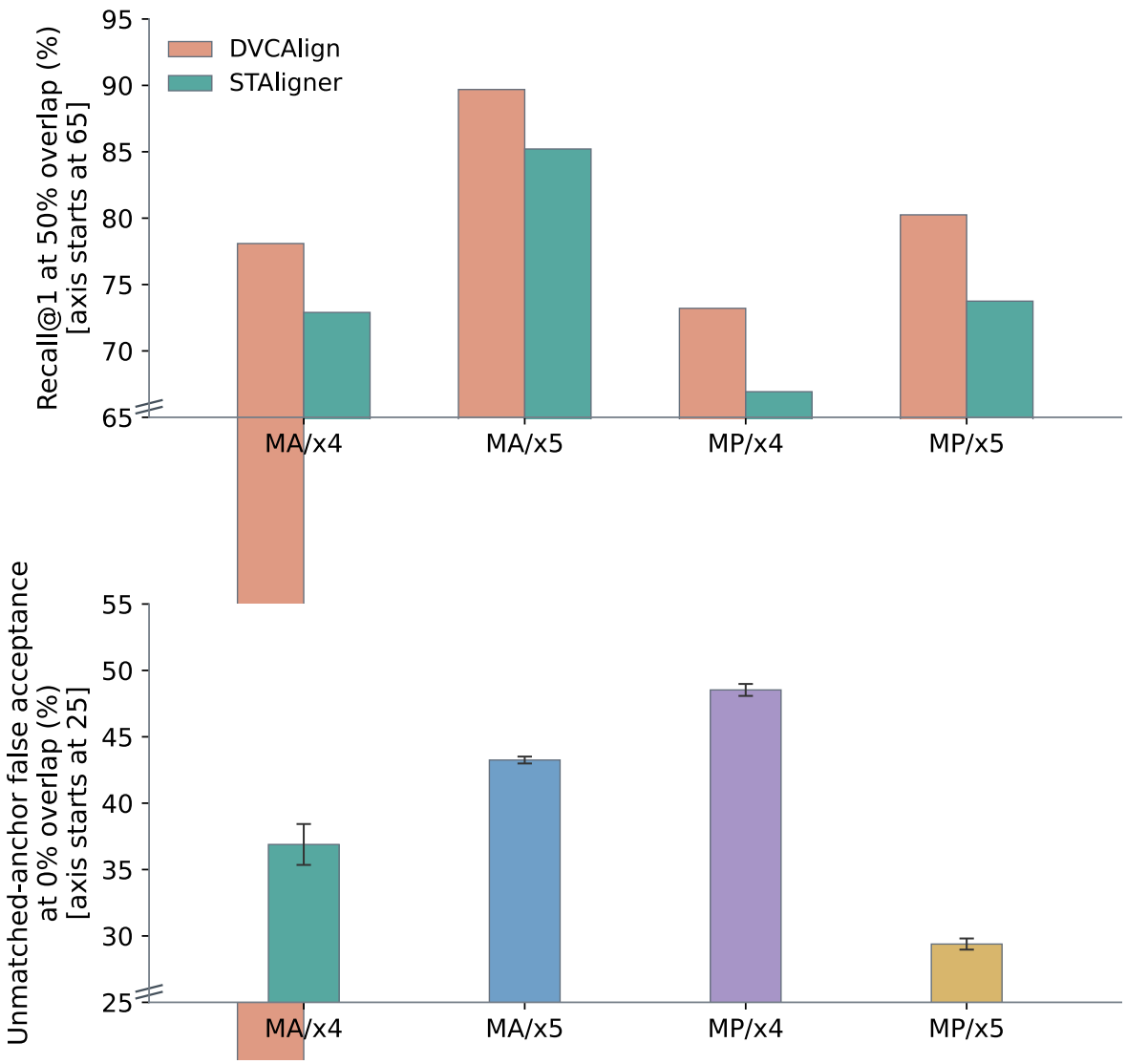

Figure S4. Correspondence retrieval and unmatched-anchor false acceptance. The Recall@1 panel compares DVCAlign and matched STAligner at 50% overlap in the MA/x4, MA/x5, MP/x4 and MP/x5 controlled settings. These Recall@1 values have no reported replicate SD. The false-acceptance panel shows the proportion of unmatched anchors entering MNN triplet supervision at zero overlap, as mean ± SD from three computational executions with training seed 666. Exact Recall@1 is undefined at zero overlap. This supervision-acceptance error is not final registration error. Nonzero numeric-axis origins are explicitly marked.

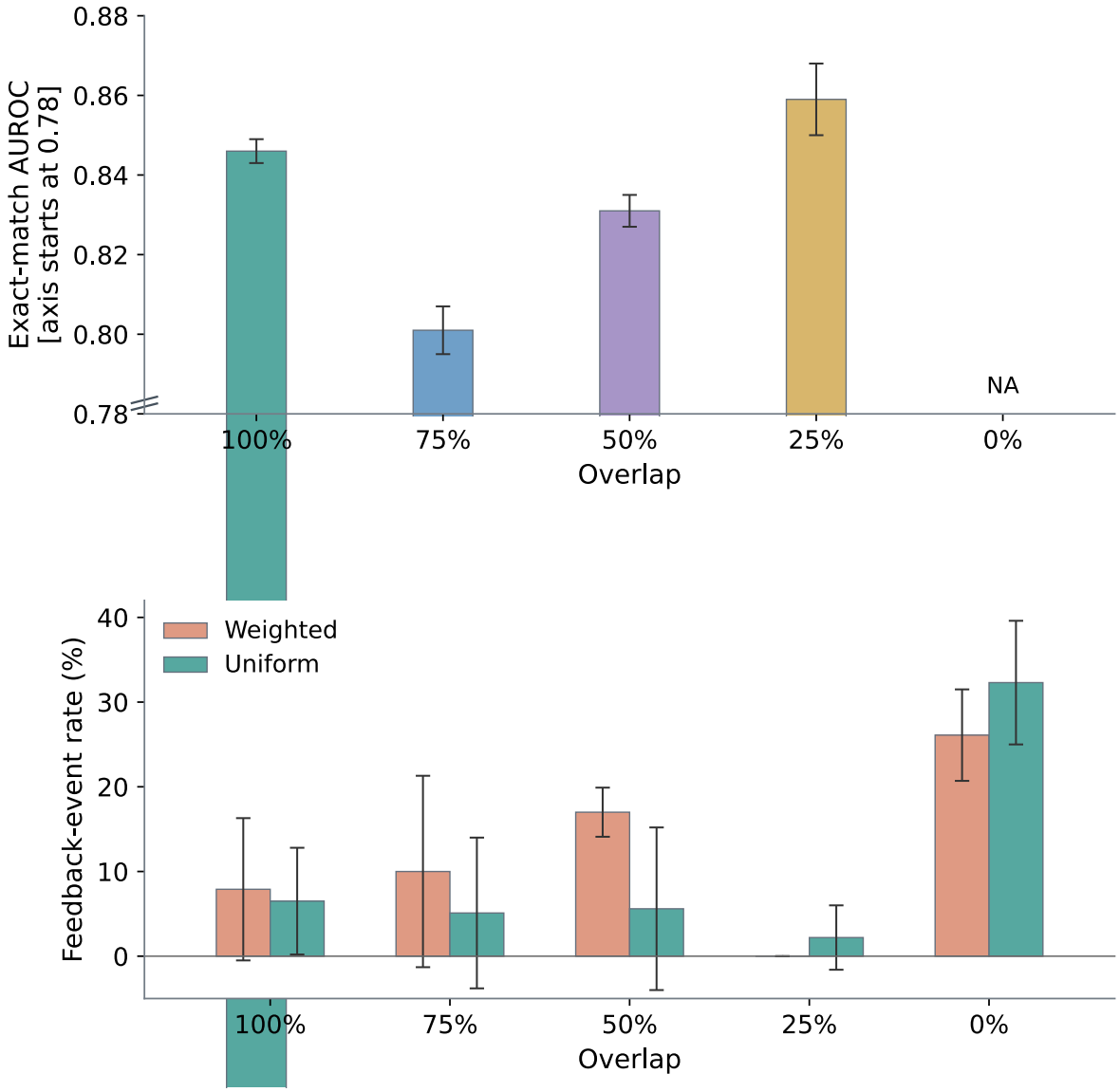

Figure S5. Confidence discrimination and feedback in controlled mouse MA data. Bars show exact-match AUROC and reinforcement-compatible event rates for weighted and uniform training. Error bars are SD across three computational seeds. AUROC is undefined at 0% overlap because no correct matches exist. Events follow the epoch-500 to epoch-900 definition in Section S4 and do not establish individual-pair causality. Symmetric SD intervals may extend below zero and are not probability bounds. The confidence effect is not uniformly beneficial across overlap conditions. The nonzero AUROC-axis origin is explicitly marked.

Section S5. DLPFC execution results and real-slice truncation

Table S4 records all six originally reported pooled ARI values and their DLPFC configuration. Pooled ARI is calculated from the concatenated labels of the four sections. The six executions use training seed 666 and clustering seed 666; their variability describes repeated computation on fixed data, not independent biological replication. Baselines retain their authors' documented or default training parameters, without additional tuning. This is a default-settings comparison, not a comparison of equally optimized methods.

The real-slice truncation comparison holds source section 151673 fixed and progressively crops target section 151674, without extra expression perturbation (Table S5). PASTE2 uses its native GLM-PCA preprocessing and annotation-independent overlap estimation. Both methods are scored on the same target spots receiving PASTE2 correspondences. Coverage is the proportion of retained target spots in

this common set. Each condition is one fixed-settings execution with seed 666. Target retention is a cropping proportion, not a measured anatomical overlap fraction, and LTARI describes label agreement rather than exact registration accuracy.

Table S4. Six reported DLPFC executions and experimental configuration.

Item	Results or settings
Six pooled ARI values	0.635, 0.625, 0.641, 0.632, 0.638, 0.625
Mean ± SD	0.6327 ± 0.0067
Preprocessing and graphs	5000 HVGs/section; spatial radius 200; auxiliary neighbors 40/20
Encoder and training	Dimensions 512/32; Adam, learning rate 0.001; 500 pretraining + 500 joint-training epochs
Triplet settings	MNN neighbors 20; margin 1.0; maximum loss weight 0.7
Seed and clustering	Training seed 666; R/mclust, EEE, 7 clusters, clustering seed 666

All six executions are reported in their original order. The mean and sample SD are 0.6327 and 0.0067. These fixed-data computational executions do not constitute independent biological replicates. The configuration values are the reported DLPFC example settings.

Table S5. Label-transfer performance under real-slice target truncation.

Target spots retained	DVCAlign LTARI	PASTE2 LTARI	Absolute LTARI improvement	Common-target coverage
100%	0.573	0.512	+0.061	60.7%
90%	0.565	0.454	+0.111	80.9%
70%	0.575	0.485	+0.090	73.0%
50%	0.522	0.434	+0.088	65.2%
30%	0.573	0.344	+0.229	58.7%

Each condition is one seed-666 execution, so execution-level SD is not available. Both methods use the common target evaluation set. Target retention is not anatomical overlap; LTARI is not exact spot-registration accuracy. The comparison supports a condition-specific result rather than universal superiority over PASTE2.

Section S6. Histology agreement and marker interpretation

Table S6 evaluates the original H1 clustering against H&E-based pathology annotations under two explicitly defined scoring scopes. The same partition is evaluated on all 613 annotated spots, including undetermined labels, and on 530 spots in six defined categories after excluding undetermined labels; the clustering is not refitted. These scoring counts differ from the 614 H1 input spots in Table 1 of the main manuscript.

All 11 highlighted H1 marker genes were included among the integration features. Their domain-associated patterns provide expression-based corroboration rather than independent molecular validation. Reference pathology agreement is assessed separately. H2 and H3 remain qualitative because corresponding independent pathology annotations are unavailable. Marker-supported labels describe transcriptional programs and are not definitive cell-type assignments or cell-lineage relationships.

Table S6. Agreement with H1 pathology annotations.

H1 evaluation scope	Spots	ARI	NMI
All annotated spots, including undetermined	613	0.372	0.494
Six defined tissue categories, excluding undetermined	530	0.471	0.572

Both scopes evaluate the same saved clustering underlying main Fig. 4. Only the evaluation scope changes. No repeated-execution SD was reported for these values.

Section S7. Supplementary heart enrichment and control analyses

The supplementary stage-wise DVCAlign rerun used the same heart datasets and generated newly fitted domain assignments, with no assumed one-to-one correspondence to the original Fig. 5f domains. The early and later datasets were fitted separately using the retained workflow and seed 666, with two and six clusters, respectively. Each stage uses one computational execution with 500 pretraining epochs and 500 alignment-training epochs, followed by R/mclust EEE clustering with seed 666. Four early-stage sections and nine later-stage sections represent one specimen per stage. Stage is consequently confounded with specimen and technical-batch factors. These analyses characterize stage-associated spatial patterns and functional annotations rather than independently established developmental effects.

Domain-associated genes for the supplementary rerun followed a fixed effect-size and detection rule: at least two eligible sections, median within-section $\log_2$ fold change ≥ 0.25, a positive effect in at least 60% of eligible sections and expression in at least 10% of domain spots. Each section comparison required at least five spots in both groups, and up to 200 genes were retained per domain. These sets are effect- and detection-selected, not gene-level FDR-significant differentially expressed genes.

The following enrichment statistics apply exclusively to the supplementary rerun. One-sided hypergeometric tests used stage-specific backgrounds of GO-annotated genes detected in at least three spots: 9081 early-stage and 11,980 later-stage genes. GO terms with 10–500 background genes were tested. Global Benjamini–Hochberg correction covered all 24,624 tests, including zero-hit terms, across eight prespecified rerun domains. The threshold was $q \leq 0.05$. Table S7 retains all nine selected results, including nonsignificant oxidative/ATP-related terms. These P and q values are not recalculated significance values for the original Fig. 5f domains. The selected cardiac-contraction and ATP-metabolism rows are reported in Table S7.

The original Fig. 5 domains were evaluated separately from the supplementary rerun. D5 contains 279 spots and has atrial-myocardium-associated marker and GO evidence; D6 contains 153 spots and has outflow-tract/large-vessel-associated evidence (Table S8). All 1515 later-stage analyzed spots were matched to the source figure. The GO family for the six original domains contains 19,308 tests with an 11,980-gene background. D6 in Slice13 contains only two spots and is excluded from section-level comparisons requiring five spots per group. MYL7 is exploratory, and consistency across sections is descriptive rather than biological replication.

The shared-encoder and matched-K controls retain contraction- and matrix-associated themes (Table S9). Their domain labels are local to the controls and must not be mapped directly to the original figure labels. Neither cross-sectional transcriptional similarity nor shared-encoder fitting constitutes lineage tracing.

Table S7. Selected GO results from supplementary later-stage rerun domains.

Rerun domain	GO term	Overlap, k/n	Fold enrichment	Raw P	Global BH q
Later R3	Cardiac muscle contraction	11/163	24.50	3.15×10^{-13}	7.05×10^{-10}
Later R3	Sarcomere organization	8/163	18.97	5.98×10^{-9}	4.80×10^{-6}
Later R3	Myofibril assembly	12/163	18.77	9.27×10^{-13}	1.63×10^{-9}
Later R3	Heart contraction	12/163	22.61	7.62×10^{-14}	3.14×10^{-10}
Later R2	ATP metabolic process	11/163	14.44	1.85×10^{-10}	2.53×10^{-7}
Later R5	Extracellular matrix organization	23/163	10.91	8.86×10^{-18}	2.18×10^{-13}
Later R2	Oxidative phosphorylation	2/163	2.49	0.1915	1.0000
Later R2	ATP biosynthetic process	2/163	6.39	0.0386	0.7813
Later R2	Mitochondrial ATP synthesis coupled electron transport	2/163	2.30	0.2164	1.0000

Domain identifiers and enrichment statistics refer exclusively to the supplementary rerun; no one-to-one correspondence with the original Fig. 5f domains is assumed. k/n denotes overlapping genes divided by GO-eligible foreground genes. P is the unadjusted one-sided hypergeometric probability; q is globally BH-adjusted over 24,624 tests. The frozen GO Biological Process 2025 library was used.

Table S8. Functional annotation of original Fig. 5 D5 and D6.

Region (spots)	Interpretation	Marker evidence: mean CP10K, domain / other	GO evidence: overlap k/n; BH q
D5 (279)	Region associated with atrial myocardium	MYH6: 80.23 / 4.39 MYL7: 227.70 / 105.43 Both higher in 9/9 sections	Sarcomere organization: 8/162; 4.40×10^{-6} Heart contraction: 7/162; 2.94×10^{-4}
D6 (153)	Region associated with the outflow tract / large vessels	ELN: 38.31 / 1.28 MYH11: 5.99 / 0.18 ACTA2: 12.73 / 2.39 All higher in 8/8 eligible sections	ECM organization: 16/164; 3.41×10^{-7} Collagen fibril organization: 9/164; 4.73×10^{-6} Vasculature development: 4/164; 0.01345

Marker expression is mean CP10K in the domain versus other spots, not mean ± SD. GO q values use the 19,308-test original-domain family. D5 markers are higher in 9/9 sections and D6 markers in 8/8 eligible sections. These are expression-based annotations, not independent cell-type validation.

Table S9. Functional themes retained in shared-encoder and matched-K controls.

Control	Cardiac muscle contraction: k/n; BH q	ECM organization: k/n; BH q	Setting
Shared encoder	12/163; 5.60×10^{-11}	24/169; 4.50×10^{-14}	All 13 sections; K = 4
Matched K	11/167; 9.10×10^{-10}	21/162; 3.37×10^{-11}	K = 4 in both stages; late-stage results shown

Both controls use repeat 0 and seed 666. The shared encoder fits all 13 sections with K=4; the matched-K control uses K=4 in both stages and reports later-stage results. BH correction is separate across 132,658 control-family tests. These q values are not SD or independent biological replication.

Section S8. Training-seed dependence and conditional uncertainty

The training-seed experiment uses the same four complete DLPFC sections from one donor: 14,364 spots and 1125 shared genes. It comprises three training seeds, 0, 1 and 42, with six executions per seed.

All 18 prespecified executions are retained without score-based run selection (Table S10). Encoder dimensions are 512/32; training lasts 1000 epochs; MNN k is 20; and auxiliary candidate/selected neighborhoods are 40/20. Each execution uses one joint R/mclust EEE fit with K=7 and clustering seed 666. Reference labels are used only for scoring; unknown-label spots remain in training and are excluded only from evaluation.

Table S11 reports mean ± sample SD within each training seed. The average of the three seed-specific mean ARIs is 0.6088, with sample SD 0.0171 between those three means. This between-seed statistic differs from within-seed execution SD and is not the SD of all 18 individual executions. These data characterize computational variability on one donor; they do not establish donor-generalizable performance or a statistically significant advantage over STAligner.

Table S12 refers instead to the original six seed-666 executions in Table S4. Assuming independent computational executions on fixed data, their reported approximate 95% t interval for the mean ARI is [0.6257,0.6396]. This interval is conditional on that assumption and does not quantify biological uncertainty or provide a confidence region for spatial boundaries. Comparable intervals for other methods and spatial confidence maps are not established by these summaries.

Table S10. Individual pooled ARI values for six executions per training seed.

Execution	Training seed 0	Training seed 1	Training seed 42
1	0.645645	0.593716	0.603416
2	0.590741	0.625985	0.589957
3	0.598795	0.632470	0.592797
4	0.648876	0.626829	0.529667
5	0.641376	0.598689	0.626384
6	0.599299	0.620382	0.592706

All 18 prespecified executions are shown. These values describe the same single-donor DLPFC dataset and fixed downstream clustering seed, not 18 independent biological samples.

Table S11. Pooled ARI and NMI summaries within each training seed.

Training seed	Executions	Pooled ARI: mean ± SD	Pooled NMI: mean ± SD
0	6	0.6208 ± 0.0271	0.7419 ± 0.0105
1	6	0.6163 ± 0.0161	0.7338 ± 0.0085
42	6	0.5892 ± 0.0321	0.7248 ± 0.0124

Each SD is the sample standard deviation across six executions within one training seed. These summaries reuse Table S10; they are not additional experiments. The between-seed SD of mean ARIs, 0.0171, is a different statistic.

Table S12. Conditional uncertainty of the six originally reported executions.

Summary	Value
Computational executions	6
Mean pooled ARI	0.6327
Sample SD	0.0067

Summary	Value
Conditional approximate 95% t interval for the mean	[0.6257, 0.6396]

The approximate 95% t interval is for the computational mean and assumes independent executions on fixed data. It is not a donor-level interval, a spatial-boundary region or a significance test against a baseline. These six seed-666 runs differ from the three-seed experiment in Tables S10–S11.

Section S9. Configuration interpretation and nomenclature

The optimizer learns model weights, residual projections and gate parameters; it does not optimize the learning rate, neighborhood settings or loss coefficients. Those hyperparameters follow implementation defaults and empirical graph-selection rules. The domain-count input to neighborhood selection is estimated by unsupervised clustering rather than read from reference annotations. Whole-section training avoids disrupting the graphs by randomly withholding spots, but it does not replace independent validation; no separately held-out hyperparameter-validation dataset is established by these experiments.

Table S13 collects the abbreviations and notation explicitly used in the revised text. The reported architecture uses a 512-dimensional hidden layer and a 32-dimensional latent representation. The gate is initialized with zero weights and bias log(0.7/0.3), then learned; its initialization differs from the fixed 0.7/0.3 graph/residual mixture. Optional cell-type deconvolution could provide composition-based corroboration from the same RNA measurements, but its benefit before integration has not been evaluated and it is not required for the current workflow.

Table S13. Abbreviations and notation used in the revised text.

Term or symbol	Definition and interpretation
DVCAlign	Dual-View Confidence-weighted Alignment; the revised expansion describes dual-view learning and confidence-weighted cross-slice supervision.
ST	Spatial transcriptomics.
DLPFC	Dorsolateral prefrontal cortex.
MA / MP	Mouse sagittal-anterior / sagittal-posterior sections.
PCW	Post-conception weeks.
HVG	Highly variable gene.
PCA	Principal component analysis.
MNN	Mutual nearest neighbor; reciprocal candidates in the learned representation, not certified physical counterparts.
ARI	Adjusted Rand index; clustering/partition agreement.
NMI	Normalized mutual information; clustering/partition agreement.
LTARI	Label-transfer adjusted Rand index; label agreement on the specified target evaluation set.
UMAP	Uniform Manifold Approximation and Projection; the displayed low-dimensional representation.
GO	Gene Ontology.
BH	Benjamini–Hochberg multiple-testing adjustment.
SD	Standard deviation over the stated executions or training seeds; sample SD is used where explicitly reported.
CP10K	Expression counts normalized to a library size of 10,000 per spot.
H&E	Hematoxylin and eosin histology.
X	Normalized expression matrix, with one row per spot and one column per shared feature.

Term or symbol	Definition and interpretation
G_s / G_a	Within-slice spatial graph / auxiliary local-expression graph, respectively.
W	Learnable transformation weight matrix; layer dimensions follow the stated encoder/decoder architecture.
σ / softmax	The hidden graph-layer activation σ in Eq. (1) is the exponential linear unit (ELU); attention-score squashing and the fusion gate use sigmoid, while softmax normalizes neighborhood attention.
K	Specified number of mclust clusters; this differs from neighborhood-size parameters.
k/n in GO summaries	Number of overlapping genes / number of GO-eligible foreground genes.
P / q	Raw enrichment-test probability / BH-adjusted probability within the explicitly defined test family.
EEE	The mclust covariance-model setting used in the reported joint fits.

Symbols are interpreted in the context of the corresponding original equations.