Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

spAlignDE unifies cross-sample and cross-modal spatial alignment with mismatch-aware differential expression

Dongyuan Song

dosong@uchc.edu

University of Connecticut Health Center https://orcid.org/0000-0003-1114-1215

Su Xu

University of Connecticut Health Center

Yu Wang

University of Connecticut

Lingzhen Meng

University of Connecticut Health Center

Aarya Dalal

University of Connecticut

Yuxin Yin

University of Connecticut Health Center

Article

Keywords: Spatial omics · Spatial alignment · Differential expression analysis · Mismatch-aware inference

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10951651/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

spAlignDE unifies cross-sample and cross-modal spatial alignment with mismatch-aware differential expression

Su Xu[1,†[0009−0003−3000−271X]], Yu Wang[2,†[0009−0000−6595−417X]], Lingzhen Meng[1[0009−0007−6871−4893]], Aarya Dalal[1,3[0009−0008−5448−5783]], Yuxin Yin[1[0009−0005−5377−4928]], and Dongyuan Song[1[0000−0003−1114−1215]*]

[1] Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington, CT 06030, USA
[2] Department of Statistics, University of Connecticut, Storrs, CT 06269, USA
[3] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

Abstract. Comparing spatial omics data across samples and modalities requires aligning them to a common coordinate system. Existing alignment methods are often computationally intensive, and cross-modal alignment remains challenging when datasets lack comparable molecular features. In addition, alignment is imperfect: residual mismatch errors can map locations from different biological regions to the same coordinates, leading to false differential expression signals. Here, we propose spAlignDE, a computational framework that integrates cross-sample, cross-modal alignment with mismatch-aware local differential expression analysis. To enable alignment across diverse modalities including transcriptomics, epigenomics, histology, and anatomical atlases, spAlignDE represents structures from each modality as image-like continuous fields. It aligns these fields using shooting-based diffeomorphic registration, without requiring shared molecular features. In cross-sample benchmarks against 12 methods, spAlignDE achieved the highest agreement in gene expression patterns and anatomical annotations across samples. It also scaled to a large cohort of 20 MERFISH brain sections containing 1.45 million cells. In cross-modal benchmarks, spAlignDE accurately aligned spatial transcriptomics with histological images, the Allen Mouse Brain Common Coordinate Framework, and spatial ATAC-seq data. Furthermore, a central innovation of spAlignDE is its mismatch-aware differential expression test. The test estimates location-resolved mismatch risks and inflates the variance component in statistical comparisons, thus reducing false signals introduced by residual mismatch. Comprehensive simulations verified improved control of false discoveries across alignment methods without systematic power loss. In two real data applications, spAlignDE localized age-associated changes in gene expression and T-cell distribution in the mouse brain and spatially restricted expression differences between normal and injured kidney sections.

Keywords: Spatial omics · Spatial alignment · Differential expression analysis · Mismatch-aware inference.

1 Introduction

Spatial transcriptomics (ST) technologies measure gene expression while preserving the locations of cells, spots or bins within tissue sections [1–3]. This *in situ* measurement links molecular states to tissue architecture and enables studies of gene expression, cellular organization, local tissue niches and anatomical changes during development and in disease [4, 5]. In recent years, the number of profiled tissue sections has increased dramatically, opening new opportunities to decipher how tissues change across conditions and over time [6, 7]. Moreover, new spatial assays also enable spatial profiling of chromatin accessibility, proteins and metabolites [8–10]. These molecular measurements can be analyzed together with histological images and anatomical atlases [2, 11, 12]. Together, these advances support comparisons of tissue sections across conditions and time points, as well as joint analysis of different modalities. However, these joint or comparative analyses are confounded by spatial heterogeneity: tissue sections differ in orientation, scale, sectioning angle, local deformation, and tissue coverage. Spatial alignment is therefore needed to establish anatomical correspondence for downstream analysis [13–16]. Like image registration, spatial alignment matches corresponding structures in a common coordinate system [17]. The resulting correspondence supports label transfer, multi-modal integration and downstream differential expression analysis [18–20].

Existing alignment methods construct correspondence using different approaches, including optimal transport [13, 21, 22], probabilistic latent-coordinate models [15, 23], graph- and representation-learning approaches [14, 16, 24], and feature-based and continuous-field registration [18, 25, 26]. These methods collectively support spatial alignment across adjacent tissue sections, biological replicates, measurement platforms and modalities. However, a recent benchmarking study found that the relative performance of spatial alignment methods varies across datasets and identified several remaining limitations [27]. We summarize four major limitations here. First, high computational demands can limit the use of existing methods for large, cell-resolved datasets or studies with many samples. Second, even close geometric overlap does not necessarily establish biologically meaningful correspondence. Overly flexible transformations may distort local cell neighborhoods, spatial domains or anatomical boundaries. They may also force regions present in only one sample into unsupported correspondence. Third, cross-modal alignment remains difficult when datasets lack directly comparable molecular features. Existing approaches often focus on one of the specific modality pairs, such as ST with histology, anatomical atlases or spatial assay for transposase-accessible chromatin using sequencing (spatial ATAC-seq), rather than providing a unified method for these modalities [19,

*Corresponding author: Dongyuan Song, dosong@uchc.edu
†These authors contributed equally to this work.

22, 26, 28]. Finally, recent methods can provide spatially resolved measures of correspondence confidence during alignment [26], but these measures do not necessarily indicate whether aligned neighborhoods remain comparable after registration. The reliability of local correspondence after alignment can therefore remain difficult to quantify and interpret. A broadly applicable framework should scale to large datasets, preserve biologically meaningful tissue organization without forcing unsupported matches, operate across diverse modalities and quantify the reliability of local correspondence.

After spatial alignment, a central analytical question is which genes increase or decrease significantly at each anatomically matched location, a task we refer to as location-resolved differential expression ("local DE"). However, to the best of our knowledge, no existing methods provide a reliable local DE test. VIMA identifies microniches whose abundance is associated with case–control status and localizes the tissue patches contributing to these associations [29]. Its tests assess microniche abundance across samples and do not directly test gene-specific expression differences at anatomically matched locations. DESpace2 tests domain-by-condition interactions to identify changes in spatial expression patterns and affected domains [30]; consequently, its domain-level aggregation does not resolve which individual locations within a domain show significant expression changes. For aligned tissue sections, CODA uses spatial cross-correlation to rank genes with conserved or divergent spatial expression patterns, but these rankings do not provide directional differential-expression tests at each matched location [20]. STcompare provides gene-level empirical P values for spatial correlation using permutations that preserve spatial autocorrelation, together with pixel-level fold-change maps. However, its pixel-level classifications use predefined fold-change thresholds and do not provide location-specific P values [31]. These advances leave a need for directional gene-expression tests at individual anatomically matched locations that account for residual alignment mismatch. Even after registration, matched coordinates may contain neighborhoods that differ in anatomy, sampling density or cell-type composition. Such residual mismatch can generate apparent local expression differences, and treating estimated correspondences as exact can underestimate uncertainty and inflate false discoveries. Reliable local DE therefore requires quantifying local comparability and incorporating this information into the uncertainty of each expression contrast.

To address these challenges in generalizable alignment and rigorous DE analysis on aligned data, we introduce spAlignDE, an integrated framework that combines spatial alignment across samples and modalities with mismatch-aware local differential expression analysis after alignment. spAlignDE converts modality-specific tissue structures into comparable, image-like continuous fields that guide shooting-based large-deformation diffeomorphic metric mapping [32]. This representation supports multi-sample ST alignment and cross-modal alignment involving histology, anatomical atlases and spatial ATAC-seq, without requiring directly comparable molecular measurements across modalities or feature fusion. To account for the effects of residual mismatch after alignment, spAlignDE estimates expression contrasts on a shared grid and inflates their variance based on mismatch risk estimated from discrepancies in putatively stable gene profiles and local observation density. Poorly supported local comparisons therefore receive lower inferential precision, while the estimated expression contrasts remain unchanged. In cross-sample benchmarks against 12 methods, spAlignDE consistently achieved the highest agreement in spatial gene-expression patterns and anatomical annotations. spAlignDE also scaled to 20 MERFISH brain sections containing 1.45 million cells in 2.43 minutes, excluding preprocessing. In cross-modal benchmarks, spAlignDE yielded high accuracy for ST-histology, ST-atlas and ST-spatial ATAC alignment and outperformed modality-specific competitors. Simulations showed improved false-discovery control without systematic loss of power. Applications to the aging mouse brain and injured kidney resolved localized changes in gene expression and cell organization.

2 Results

2.1 Overview of the spAlignDE framework

spAlignDE has two complementary components: structure-guided spatial alignment and mismatch-aware local differential expression analysis (Fig. 1).In the spatial alignment component, spAlignDE represents modality-specific tissue organization as spatial structures and establishes correspondence using shared structure identities across samples or geometric compatibility across modalities. This design supports both cross-sample and cross-modal alignment. In particular, the cross-modal workflow does not require a shared molecular feature space or feature fusion. In the local differential expression component, spAlignDE quantifies putative residual mismatch after alignment and propagates this risk into inferential uncertainty through location-specific variance inflation, rather than treating the estimated correspondence as fixed. The model details are described in Methods 5.1 and 5.2, with implementation details in Supplementary Methods S2 and S3.

For multi-sample spatial transcriptomics, spAlignDE identifies spatial structures jointly across sections using gene expression and spatial neighborhood information in a batch-adjusted representation [33, 34]. Structures assigned the same identity establish direct query–reference correspondence (Fig. 1A; Methods 5.1.1). For cross-modal alignment, spAlignDE identifies structures independently in spatial transcriptomics, spatial ATAC-seq and histological images, or obtains them from predefined anatomical atlas compartments (Fig. 1B; Methods 5.1.1). These structures provide a modality-agnostic representation, and spAlignDE next identifies their correspondences across modalities based on geometric similarity. Because some geometric similarity metrics are unreliable when samples differ in global properties such as position, orientation, and scale, spAlignDE first performs a global pre-alignment to initialize approximate spatial correspondence (Methods 5.1.2). If automatic initialization remains inaccurate, users can manually adjust the global alignment through an interactive interface (Supplementary

Fig. S1; Methods 5.1.2). spAlignDE then represents candidate structures as masks and assigns each query–reference structure pair a composite score based on five complementary metrics of geometric correspondence (Methods 5.1.3). spAlignDE retains only high-confidence candidate pairs that satisfy modality-specific criteria and does not require every structure to have a corresponding match. For atlas alignment, the anatomical hierarchy provides a multiscale library of candidate regions. spAlignDE progressively refines the ST structure resolution while retaining all eligible atlas-hierarchy candidates at every stage (Supplementary Methods S2.8). Users can also manually specify correspondences through an interactive interface when automatic pairing is ambiguous (Supplementary Fig. S2; Supplementary Methods S2.8).

spAlignDE converts the shared cross-sample structures or accepted cross-modal structure pairs into continuous multichannel fields analogous to multichannel images, with each channel encoding one spatial structure. For cross-sample alignment, spAlignDE aligns samples using smoothed fields of local structure composition. For cell-resolved data, it also includes a cell-density field when density provides useful information for alignment. A spot-density field is not used for spot-based ST because the spots are evenly distributed and provide no informative spatial variation in density. For cross-modal alignment, spAlignDE uses paired signed-distance-transform fields derived from accepted structure pairs and, when informative, an additional field representing the overall tissue shape. spAlignDE then rasterizes these image-like fields for image registration. spAlignDE jointly registers the rasterized fields using shooting-based large deformation diffeomorphic metric mapping (S-LDDMM), which estimates a smooth, invertible transformation with affine and diffeomorphic components [32, 35, 36]. Standard variational LDDMM optimizes a sequence of time-dependent velocity fields [35]. Instead, S-LDDMM estimates a single initial momentum field that determines the full geodesic trajectory [32]. By reducing the number of parameters to optimize and the memory required, this parameterization greatly improves scalability. To accommodate missing tissue and partial overlap, the registration model must allow for locations with no counterpart in the other sample. Treating these locations as matched can distort the estimated transformation. spAlignDE therefore adapts the probabilistic missing-data formulation of Tward et al. [18, 37] and fits an expectation–maximization mixture model with one matched component and two unmatched appearance components. The matched-component posterior probability $W_M(\mathbf{x})$ weights each location during registration, so locations more consistent with the unmatched components contribute less to transformation estimation. Finally, spAlignDE applies the estimated transformation to the original coordinates of cells, spots or bins to align the data at their original spatial resolution (Methods 5.1.4 and 5.1.5).

Even after alignment, nominally matched locations may differ in local anatomy or cell-type organization, so their expression differences may reflect residual mismatch rather than true between-sample changes. spAlignDE addresses this problem by allowing a putative local mismatch risk to increase inferential uncertainty rather than treating the estimated correspondence as exact. spAlignDE constructs a shared testing grid over the aligned tissue sections (Fig. 1C). Each retained grid location serves as a testing unit, and nearby cells or spots form a kernel-weighted neighborhood for estimating the local expression mean μ. For post-alignment inference, sample A denotes the condition of interest (case), and sample B denotes the baseline condition (control), regardless of which sample served as the query or reference during alignment. For gene g at location i, the estimated directional contrast is

$$\widehat{\Delta}_{ig} = \widehat{\mu}_{\mathrm{A},ig} - \widehat{\mu}_{\mathrm{B},ig},$$

and the working model is

$$\widehat{\Delta}_{ig} = \gamma_g + \mathbf{z}_i^\top \boldsymbol{\beta}_g + \theta_{ig} + \varepsilon_{ig}.$$

When included, γ_g represents the comparison-wide baseline of the contrast, $\mathbf{z}_i^\top \boldsymbol{\beta}_g$ captures optional adjustment for confounding variation such as effects of library size and detection rate, and θ_{ig} is the location-specific change of interest. With the comparison-wide intercept included, spAlignDE tests whether the local contrast departs from the fitted baseline defined by the comparison-wide contrast and included confounders. The intercept γ_g can absorb approximately uniform between-section shifts as a practical compromise for mitigating potential batch effects when each condition lacks multiple replicates. When the intercept is omitted, the test instead evaluates whether the covariate-adjusted local contrast differs from zero. Throughout the post-alignment analyses, P denotes a raw P value, and q denotes the corresponding Benjamini–Hochberg-adjusted P value within the testing family specified for each analysis.

The true residual mismatch is not directly observed, so spAlignDE uses a data-derived estimate of mismatch risk as a practical compromise. Similarly to the use of stable control genes for batch-effect correction in scRNA-seq [38], spAlignDE quantifies the severity of potential mismatch from cross-sample discrepancies in the local distributions of putatively stable genes and from local sampling-density discordance (Fig. 1D). These quantities define a normalized local mismatch-risk score R_i^{loc}. When cell-type annotations are available, an optional adjustment compares the kernel-smoothed local cell-type-composition vectors between the two samples to reflect discrepancies in local anatomical structures directly. The final adjusted local variance is

$$\widehat{V}_{ig} = \widehat{V}_{ig}^{\mathrm{base}} \left\{ 1 + \widehat{\lambda}_g \left(R_i^{\mathrm{loc}} \right)^2 \right\} \phi_i^{\mathrm{cell}}.$$

Here, $\widehat{V}_{ig}^{\mathrm{base}}$ represents the baseline variance of the local comparison. The location-specific score R_i^{loc} quantifies the risk that residual alignment mismatch makes the comparison at location i less reliable, and the gene-specific coefficient $\widehat{\lambda}_g$ controls how strongly this risk inflates the variance for gene g. To estimate $\widehat{\lambda}_g$, spAlignDE first computes local statistics without mismatch adjustment, groups locations by R_i^{loc}, and estimates the excess dispersion of the statistics in each risk group relative to the corresponding Student-t null distribution. Then a non-negative weighted regression of excess dispersion on

the squared mismatch risk is fitted; after a bounded reference-bin adjustment, its fitted coefficient gives a contrast-specific estimate of λ_g. When multiple contrasts are available, these estimates are robustly combined to obtain $\widehat{\lambda}_g$. The optional factor ϕ_i^{cell} provides a further adjustment for local cell-type-composition inconsistency between samples. These adjustments increase inferential uncertainty where the aligned samples are less biologically comparable but leave the estimated effect and its direction unchanged. spAlignDE reports local P values, q values and connected significant regions for downstream gene-level and spatial analyses (Fig. 1E). Cells or spots contribute to estimating local spatial patterns but are not treated as independent biological replicates. Full details are provided in Methods 5.2.1–5.2.6.

2.2 spAlignDE enables accurate and scalable cross-sample alignment

Accurate cross-sample alignment of spatial transcriptomics data requires both reliable anatomical correspondence across sections and preservation of local tissue organization during deformation. The spAlignDE cross-sample workflow comprises two stages (Fig. 2A). First, spAlignDE identifies shared spatial clusters from batch-corrected gene expression and spatial neighborhood information. After rasterization, these clusters provide a common image-like representation of anatomical structures that guides coarse alignment and subsequent diffeomorphic registration (Methods 5.1.1, 5.1.2, and 5.1.4). Second, spAlignDE uses shooting-based LDDMM to estimate smooth transformations that correct global and local deformations while preserving tissue continuity (Methods 5.1.5).

We benchmarked spAlignDE against 12 existing spatial alignment methods [13–16, 18, 20, 21, 23–26, 39] using three datasets that differed in spatial resolution and alignment scale. These datasets included a paired MERFISH mouse brain dataset [40], a paired 10x Genomics Visium mouse kidney dataset comparing normal and ischemia–reperfusion conditions [31, 41, 42], and a multi-sample MERFISH mouse brain aging cohort [6, 43]. Details of method implementations and settings are documented in Supplementary Methods S2.9. The benchmark covered cell- and spot-resolution platforms, healthy and disease conditions, and both single-pair and large-scale multi-sample alignment.

The paired Vizgen MERFISH mouse brain benchmark included coronal sections from two biological replicates (S2R2 and S2R3) profiled at single-cell resolution, and we aligned S2R3 to S2R2. Following the benchmarking pipeline described in [27], we evaluated the preservation of spatial gene expression patterns across 483 shared genes using four metrics: the Pearson correlation coefficient (PCC), cosine similarity, structural similarity index measure (SSIM) and mutual information (MI) (Fig. 2B; Methods 5.3.1). spAlignDE achieved the highest mean performance across all four metrics. The rankings of competing methods, however, varied by metric. Because these metrics are calculated after gene expression is aggregated on a spatial grid [27], the results may depend on the degree of spatial averaging. We therefore repeated the evaluation using finer grids. spAlignDE remained top-ranked at both resolutions (Supplementary Figs. S3 and S4). Alignment performance also remained stable across clustering resolutions that produced 15–27 joint spatial clusters (Supplementary Figs. S5A, B and S6; Methods 5.3.1; Supplementary Methods S2.12). These results indicate that spAlignDE is robust to changes in evaluation-grid resolution and the granularity of the spatial structures used for alignment.

We next assessed computational efficiency by measuring core alignment runtime (excluding preprocessing) and peak GPU memory usage (Fig. 2C; Methods 5.5.2; Supplementary Methods S2.10.1). spAlignDE completed the core alignment in 0.1 min, tying for the shortest runtime among the evaluated methods. Wall-clock runtimes are reported separately in Supplementary Methods S2.10; most preprocessing steps are part of standard spatial data analysis workflows. Its peak GPU memory usage was 0.20 GB, the second-lowest value. In summary, spAlignDE achieved scalable alignment without loss of accuracy; its speed and memory efficiency stem from using image-like representations in place of individual cells during registration and from efficient optimization with S-LDDMM.

We further assessed anatomical preservation at global and local scales by visual inspection (Fig. 2D). spAlignDE matched the overall tissue geometry and boundaries and retained the laminar organization of the hippocampus with minimal local distortion. In contrast, competing methods showed excessive global deformation, residual local mismatch or forced alignment in regions with incomplete tissue support. These observations indicate that spAlignDE preserved anatomical structure at both spatial scales in this benchmark.

Beyond alignment accuracy, spAlignDE can quantify alignment uncertainty by repeating its fast alignment using independently subsampled cells from both sections (Fig. 2E; Methods 5.3.1; Supplementary Methods S2.11). Most locations showed small mean Euclidean displacement from their repeat-mean aligned positions across repeated alignments, consistent with locally stable transformation estimates. Larger mean displacement was concentrated in areas with incomplete tissue overlap at the bottom-left corner. We marked locations above the 95th percentile to identify regions with reduced transformation stability. These maps identify regions where downstream results should be interpreted with caution.

We next benchmarked spAlignDE on spot-resolution Visium kidney sections by aligning the ischemia–reperfusion injury section IL3 to the normal section NL3 (Fig. 2F,G). Because injury-associated expression changes make gene-expression concordance unreliable as a measure of alignment accuracy, we did not evaluate this metric. We instead evaluated anatomical correspondence based on overall kidney morphology and annotations for three compartments and other tissue [42]. spAlignDE preserved overall kidney morphology and improved correspondence across the annotated compartments and other tissue. spAlignDE ranked first among all methods for each of the three compartment-label concordance metrics: k-nearest-neighbor (kNN) label matching, weighted Dice and weighted intersection over union (Methods 5.3.1). We subsequently used the aligned sections for local DE analysis (Results 2.5).

We then assessed the scalability and alignment accuracy of spAlignDE using the mouse brain aging MERFISH dataset from Sun et al. [6, 43], which comprised 20 coronal sections and 1,453,144 cells (Fig. 2H–K). We used the 4.3-month section as the reference because it contained the most complete set of anatomical structures and aligned the remaining 19 sections to this reference. We evaluated alignment accuracy for these 19 query-to-reference comparisons using anatomical annotations, spatial overlap and gene-expression concordance (Methods 5.3.1). spAlignDE mapped all sections to a common coordinate system while preserving anatomical organization and representative marker-gene patterns (Fig. 2H). spAlignDE achieved the highest region-level annotation agreement among the 13 methods (Fig. 2I) and remained top-ranked when evaluated using cell-type and subregion annotations (Supplementary Figs. S7 and S8). spAlignDE also maintained consistent whole-section correspondence across the cohort (Supplementary Fig. S9). Using the default evaluation grid, spAlignDE achieved the highest gene-expression concordance across 300 genes. When we repeated the analysis using finer grids, spAlignDE remained top-ranked for all four metrics (Supplementary Figs. S10, S11 and S12). spAlignDE completed the 19 core query-to-reference alignments in 2.43 min, with peak allocated GPU memory below 0.1 GB (Methods 5.5.1 and 5.5.2). We report wall-clock runtimes separately in Supplementary Methods S2.10. These results show that spAlignDE scaled well to a large cohort study.

We also checked the performance of spAlignDE under incomplete spatial coverage in the aging brain cohort. For the 26.7-month query, spAlignDE preserved the lower-right boundary created by incomplete tissue coverage while maintaining overall anatomical correspondence, rather than forcing unsupported overlap with the 4.3-month reference (Fig. 2J; Methods 5.1.5). Mean displacement from the repeat-mean aligned position was elevated in regions with discordant spatial structures or incomplete tissue support (Supplementary Fig. S13; Methods 5.3.1; Supplementary Methods S2.11). This larger mean displacement indicates that the estimated transformation was less stable where correspondence was poorly constrained. The resulting common coordinate system enabled comparison of spatial gene-expression trajectories across age (Fig. 2K). We further analyzed these trajectories in Results 2.5 (Methods 5.2.6 and 5.4.1).

The preceding benchmarks primarily evaluated whole-section alignment between tissue sections with broadly corresponding anatomy and well-defined spatial domains, even when tissue coverage was incomplete. To assess partial alignment between sections with greater anatomical heterogeneity, we aligned two consecutive sections from the 10x Genomics Xenium FFPE human breast cancer dataset [44], using manual pre-alignment for global initialization (Methods 5.1.2). Using the default evaluation grid, spAlignDE ranked first across all four gene-pattern metrics. At finer grid resolutions, spAlignDE remained close to the top-performing method (Supplementary Figs. S14–S16; Methods 5.3.1). Spatial overlays showed close correspondence within the shared tissue without forcing non-overlapping regions to match (Supplementary Fig. S17). Thus, spAlignDE can accommodate partial overlap between anatomically heterogeneous tissue sections when an appropriate global initialization is available.

2.3 spAlignDE enables structure-guided cross-modal alignment without requiring shared molecular features

Cross-modal alignment is more challenging than cross-sample alignment because the datasets may not share directly comparable molecular features. spAlignDE therefore establishes cross-modal correspondence through spatial structures rather than a shared molecular feature space. spAlignDE first identifies modality-specific structures from spatial transcriptomics, spatial ATAC-seq, histological images or anatomical atlases (Methods 5.1.1). After global pre-alignment, spAlignDE represents these structures as masks and pairs query and reference structures according to their geometric compatibility (Methods 5.1.2 and 5.1.3). spAlignDE then converts the accepted structure pairs into matched continuous fields and aligns these fields using S-LDDMM (Fig. 3A; Methods 5.1.4 and 5.1.5). We evaluated this workflow by aligning spatial transcriptomics with histology, an anatomical atlas and spatial ATAC-seq.

We first aligned a Xenium mouse brain section to an H&E image, which contains paired Visium gene-expression measurements [45, 46] (Fig. 3B). We withheld the Visium measurements from the entire alignment and used them only for independent evaluation. Using spatial structures identified from H&E, spAlignDE refined the initial global alignment with S-LDDMM and improved anatomical correspondence (Fig. 3B and Supplementary Fig. S18; Methods 5.1.1, 5.1.2, 5.1.4 and 5.1.5). We evaluated the result by comparing spatial gene-expression patterns between the aligned Xenium data and the held-out Visium measurements. Across 246 shared genes, PCC, cosine similarity, SSIM and MI showed high molecular concordance between aligned data (Fig. 3C; Methods 5.3.1). Given the differences between cell-level Xenium and spot-level Visium measurements, we downweighted genes with sparse detection, limited spatial variation or imbalanced detection across platforms. Weighted mean similarity was higher across all four metrics, with mean cosine similarity increasing from 0.60 to 0.80, supporting cross-platform molecular concordance (Supplementary Methods S2.10.2). Representative genes had spatial PCC values of 0.48–0.73 (Fig. 3D). To evaluate the generalizability to other types of histological images, we aligned a MERFISH mouse brain section to a Nissl-stained section from the Allen Mouse Brain Atlas [11, 40]. The alignment by spAlignDE improved correspondence of the tissue boundary and internal cytoarchitecture (Supplementary Fig. S19; Methods 5.1.1, 5.1.2, 5.1.4 and 5.1.5). These analyses show that spAlignDE can align Xenium and MERFISH data to H&E and Nissl images, respectively, without requiring shared molecular features.

Next, we evaluated alignment to the Allen Mouse Brain Common Coordinate Framework version 3 (CCFv3), a three-dimensional reference atlas of the adult mouse brain with hierarchically organized anatomical regions [12]. Because the MERFISH sample is a single tissue section, we used a two-dimensional coronal section of CCFv3 at the corresponding

anatomical depth. We aligned the MERFISH section to this atlas section and transferred region- and subregion-level atlas labels to the ST cells (Fig. 3E). Across three stages, spAlignDE progressively refined the ST structure partition while retaining all eligible atlas hierarchy candidates from depths 2–10 at each stage, and then applied iterative S-LDDMM refinement (Supplementary Methods S2.8 and Methods 5.1.5). The automatic workflow retained 18 matched ST–atlas structure pairs. We compared spAlignDE with xIV-LDDMM, a task-specific method for cross-modal atlas-to-spatial-omics registration [19, 28] (Supplementary Methods S2.9). spAlignDE more closely reproduced the whole-section geometry and local anatomical organization of the CCFv3 reference. In contrast, xIV-LDDMM showed larger discrepancies in tissue boundaries, section size and local deformation (Fig. 3E, top). After atlas labels were projected back to the native ST coordinates, spAlignDE left only 559 cells unlabeled (0.67%), compared with 4,852 cells (5.81%) for xIV-LDDMM (Fig. 3E, bottom).

We then evaluated whether the transferred atlas labels were consistent with the MERFISH gene-expression profiles. Using RCTD with the Allen mouse cortex and hippocampus single-cell dataset as the reference, we independently assigned transcriptomic subclasses to the MERFISH cells [47–49]. We compared these expression-based subclasses with the corresponding broad CCFv3 regions assigned after alignment. spAlignDE showed higher agreement between the two label sets in all nine evaluated cortical, hippocampal and dentate gyrus categories (Fig. 3F; Methods 5.3.1). Across seven brain region marker genes, spAlignDE produced higher expected-region localization than xIV-LDDMM for all seven genes (Fig. 3G) [19, 50]. The marker-gene patterns were more concentrated within the expected atlas regions and showed less leakage across anatomical boundaries. These results show that spAlignDE improved both geometric atlas alignment and the biological consistency of atlas-label transfer in this comparison.

To assess whether atlas-alignment performance remained stable across brain sections at different coronal depths, we aligned two additional MERFISH sections, S1R1 and S3R1, to the corresponding coronal sections of Allen CCFv3 (Methods 5.1.5 and Supplementary Methods S2.8). For both sections, spAlignDE closely matched the overall tissue shape and local anatomical regions and transferred atlas labels to the MERFISH cells (Supplementary Figs. S20 and S21). These results indicate that the atlas-alignment workflow generalizes across independently profiled sections with different coronal depths and tissue geometries.

Matched structure pairs constrain the estimated transformation, so we tested whether alignment was sensitive to the number retained. The automatic workflow retained 18 pairs for S2R1. Holding all other components fixed, we reran the final-resolution S-LDDMM step once from the same global pre-alignment using all 18 pairs as the baseline. We then independently repeated this step with the top-ranked 16, 14, 12, 10 or 8 pairs (Methods 5.3.1; Supplementary Methods S2.12). Label agreement remained stable across cortical, hippocampal and dentate gyrus categories as fewer pairs were retained. Regionally enriched marker genes also remained localized to their expected atlas regions (Supplementary Figs. S22 and S23). To complement this analysis of automatic pairing, we evaluated an interactive interface through which users can specify correspondences when automatic pairing is uncertain but anatomical knowledge is available (Supplementary Methods S2.5). Alignment using user-specified correspondences closely matched the Allen CCFv3 reference. Overall label agreement was comparable to that of the automatic workflow (37.8% versus 36.3%), with only modest category-specific differences (Supplementary Figs. S24 and S25). Marker-gene localization was also similar between the two workflows (Supplementary Fig. S26). Finally, we evaluated alignment across modalities and spatial resolutions by aligning a spatial ATAC-seq section at 20-μm pixel resolution to a cell-resolved MERFISH ST reference [8, 40]. Because the ATAC-seq section covered only part of the mouse brain, we used the corresponding half-brain region of the MERFISH data as the reference (Methods 5.1.1–5.1.5). We compared spAlignDE with 3d-OT, a method for heterogeneous spatial multi-omics alignment [22] (Supplementary Methods S2.9). Although 3d-OT placed the query in the correct broad brain region, it spread the ATAC-seq pixels across most of the reference, distorting the query geometry and internal spatial organization. spAlignDE preserved the compact query geometry and local structure (Fig. 3H). We quantified neighborhood preservation as the fraction of each pixel's original K-nearest ATAC-seq neighbors retained after alignment. Across the evaluated values of K, mean preservation ranged from 0.89 to 0.98 for spAlignDE and from 0.18 to 0.46 for 3d-OT (Fig. 3I; Methods 5.3.1). We next assessed biological correspondence using published anatomical annotations and gene-activity scores for the ATAC-seq pixels, with the ST reference assigned to the same coarse anatomical categories based on marker patterns [8, 51]. Each aligned ATAC-seq pixel received an ST-derived label through inverse-distance-weighted voting among its ten nearest ST cells. spAlignDE achieved a higher exact annotation match rate than 3d-OT (62.5% versus 49.6%), with higher rates in seven of the eight anatomical categories (Fig. 3J). For representative genes, ATAC-derived gene activity and ST expression showed concordant spatial patterns after spAlignDE alignment, while the ATAC-seq section retained its native geometry (Fig. 3K).

2.4 Mismatch-aware post-alignment testing improves false-discovery control in location-resolved comparisons

After benchmarking spAlignDE's alignment accuracy and efficiency, we next evaluated whether its mismatch-aware local differential expression test improves false-discovery control. The test compares expression at each location on a shared grid and inflates the variance of each local contrast according to the estimated mismatch risk. The mismatch-aware DE test has broader applicability: it can be used downstream of any alignment method, not only the alignment procedure in spAlignDE. To determine whether this adjustment improves post-alignment inference, we constructed semi-synthetic datasets with known local differential expression regions using a real mouse brain spatial template (Fig. 4A). Simulation sample A served as

the baseline reference. Simulation sample B contained spatially localized up- and down-regulated regions and underwent global and local deformations before alignment. We defined the post-alignment contrast as sample B minus sample A. This design retained realistic tissue geometry and spatial sampling while providing known local differential expression regions [52] (Methods 5.3.2; Supplementary Methods S3.10). Because the local differential expression model can be applied to the output of any alignment algorithm, we compared its naïve and mismatch-aware modes across alignments produced by spAlignDE and nine other alignment methods. The naïve mode treats the aligned coordinates as exact. The mismatch-aware mode instead incorporates residual mismatch into the local variance (Methods 5.2.1–5.2.5).

Across ten simulation groups, we evaluated naïve and mismatch-aware post-alignment inference for alignments produced by spAlignDE and nine other methods (Fig. 4B; Methods 5.3.3). Grid-level calls were projected onto cells in sample B and compared with the implanted truth, requiring true-positive calls to have the correct direction. At a target FDR of 0.05, alignments produced by spAlignDE yielded a near-zero median FDP and a median power of approximately 0.7 at FDP $\leq$ 0.05. SPOmiAlign and Spateo also yielded relatively low FDP and high power, whereas less accurate alignments showed higher FDP and lower power. This ordering broadly agreed with the independent alignment benchmark (Results 2.2). Mismatch-aware adjustment reduced FDP across all alignment methods, with larger reductions when residual mismatch was greater. For STalign and CODA, the adjustment lowered median FDP below 0.05. At the same FDP threshold, power was generally maintained and increased for some methods. However, FDP remained above the nominal target for several methods with extreme residual mismatch, indicating that mismatch-aware testing cannot fully compensate for poor alignment and that accurate alignment is essential.

We then examined a representative simulation to assess whether mismatch risk reflects false-positive calls. False-positive calls from the naïve analysis were concentrated in regions with high robust-standardized mismatch-risk scores ζ_i^R, which overlapped the implanted distorted regions (Fig. 4C and Supplementary Figs. S27 and S28). Mismatch-aware testing removed many false-positive calls while retaining most implanted up- and down-regulated regions. FDP decreased from 0.108 to 0.007 for spAlignDE, from 0.210 to 0.015 for STalign and from 0.389 to 0.178 for CAST. We next assessed robustness to differential expression signal strength and deformation severity (Methods 5.3.2; Supplementary Methods S3.10). Across 20 matched simulation templates at each signal level, power increased with the strength of the implanted effects. Mismatch-aware testing maintained a low median FDP at the target FDR of 0.05 and power comparable to that of the naïve analysis at FDP $\leq$ 0.05 (Fig. 4D). Increasing the severity of local deformation produced the same qualitative pattern (Supplementary Fig. S32). Across 15 matched simulation groups, all six alignment methods showed higher median regional mismatch scores inside implanted distorted regions than outside them (Fig. 4E). Regional scores were summarized before median–MAD standardization and normalized using a common min–max transformation (Supplementary Methods S3.3). Mismatch-aware testing with the estimated risk map reduced median FDP relative to naïve testing for all six methods. Permuting the risk values across grid locations increased median FDP relative to testing with the estimated risk map (Fig. 4F). For STalign, median FDP decreased from 0.164 with naïve testing to 0.019 with the estimated risk map, compared with 0.124 after permutation. The spatial pattern of the estimated risk map therefore contributed to the reduction in false discoveries. We also examined controlled coordinate perturbations. Under the perturbation scheme shown in Supplementary Fig. S29, shared-grid inference was more stable than cell-centered local testing (Supplementary Fig. S30). Mismatch-aware adjustment further reduced FDP within the shared-grid workflow, with some loss of power at larger perturbations (Supplementary Fig. S31; Supplementary Methods S3.10.1).

2.5 Mismatch-aware inference identifies localized changes in aging brain and injured kidney

After validating the DE component in simulations, we applied mismatch-aware post-alignment inference to two real datasets: a cell-resolved MERFISH dataset of aging brain samples and a spot-based Visium dataset of injured and normal kidney samples. The aging mouse brain dataset comprised 20 coronal sections spanning distinct ages and comparable anatomical levels [43]. We selected the 4.3-month section as the reference because it retained the most complete anatomy, and then aligned the remaining 19 sections to it (Methods 5.4.1; Supplementary Methods S3.11). The original study reported a broad age-associated decline in *Gamt* expression [43]. To determine the spatial pattern of this decline, spAlignDE estimated local t-statistics and q-values at each shared-grid location and grouped contiguous significant locations into regions (Fig. 5A, B; Methods 5.2.5; Supplementary Methods S3.2). *Gamt* down-regulation became more pronounced at later ages and was concentrated in an arch-shaped white-matter region along the corpus callosum (CC).

To summarize age-related changes at the gene level across locations, we fitted a linear age trend to the unsmoothed adjusted expression at each grid location using mismatch-aware precision weights. We then combined the resulting local trend P values across space using the Aggregated Cauchy Association Test to obtain a gene-level P value (ACAT; Methods 5.2.6; Supplementary Methods S3.9). *Gamt* showed evidence of an age-associated spatial trend ($P = 1.1 \times 10^{-4}$), but *Vip* did not ($P = 0.4182$; Fig. 5C). The *Vip* result was consistent with a previous mouse cortex study that reported little age-related change in *Vip* transcript or VIP protein abundance [53]. To characterize where and how expression changed with age across the tissue, we grouped locations with similar adjusted expression trajectories and tested the age trend within each spatial cluster using spatial-variability-weighted regression (Methods 5.2.6). The automatic procedure selected $K = 5$ for *Gamt* and $K = 2$ for *Vip*. For *Gamt*, clusters C1–C3 yielded two-sided Wald P values below 0.05 ($P = 1.9\times10^{-10}$, 1.8×10^{-4} and 0.0446, respectively). Clusters C4 and C5 showed no evidence of an age trend ($P = 0.4063$ and 0.2743). Neither *Vip* cluster showed

evidence of an age trend ($P = 0.8945$ and 0.9073). Because the cluster number and spatial partition were selected from the observed trajectories, these raw Wald P values should be interpreted as conditional summaries. As a sensitivity analysis, we fixed K at 2, 4, 6 and 8; the resulting partitions mainly subdivided the same broad spatial domains (Supplementary Fig. S33).

spAlignDE supports spatial comparisons beyond differential expression, including analyses of cell-type organization across samples. We used this capability to examine age-associated changes in spatial cell-type organization. As illustrated schematically in Fig. 5D, top, proximity describes how close T cells are to a target cell population, whereas spatial overlap describes how closely their spatial distributions coincide. We quantified T-cell proximity by the median nearest-neighbor distance and spatial overlap by the mismatch-risk-weighted correlation between smoothed cell-density fields (Methods 5.4.1; Supplementary Methods S3.11). Positive proximity and overlap effects indicate decreasing distance and increasing spatial overlap with age, respectively. T-cell proximity to oligodendrocytes, microglia and ependymal cells increased with age ($\beta_{\mathrm{prox}} = 0.025$, 0.010 and 0.028; $q = 0.001$, 0.048 and 0.001, respectively). Spatial overlap with the same cell classes also increased ($\beta_{\mathrm{overlap}} = 0.011$, 0.010 and 0.006; $q < 0.001$, $q < 0.001$ and $q = 0.021$, respectively; Fig. 5D, bottom), consistent with the source study [43]. To determine where T cells were locally concentrated, we mapped local T-cell enrichment relative to its tissue-area-based expectation in the shared coordinate system (Fig. 5E, left, schematic). Positive enrichment indicates local T-cell overrepresentation compared to its expectation. Young (≤ 13.6 months; $n = 7$) and old (≥ 24.3 months; $n = 7$) summaries showed greater enrichment around an oligodendrocyte-associated callosal region in older brains (Fig. 5E, right). Using the callosal-proximal and callosal-distal ROIs illustrated in Fig. 5E, middle, we examined T-cell distributions relative to oligodendrocytes and enrichment trajectories across age. The callosal-proximal enrichment trajectory showed an age association ($P = 2.0 \times 10^{-4}$), whereas the callosal-distal trajectory yielded $P = 0.6353$ (Fig. 5F). These spatial patterns localized age-associated T-cell accumulation near callosal oligodendrocyte neighborhoods.

In the injured kidney application, we treated the normal NL3 Visium section as the reference and the injured IL3 section as the query, following the comparison in STcompare [31] (Methods 5.4.2; Supplementary Methods S3.12). Gene-level ACAT tests identified spatial expression differences for *Cbr1*, *Cd44* and *Myo5a* ($P = 1.7 \times 10^{-14}$, 4.1×10^{-6} and 7.4×10^{-14}, respectively). The local maps showed where and in which direction these differences occurred: *Cbr1* and *Cd44* increased in spatially restricted regions of the injured section, whereas *Myo5a* decreased locally (Fig. 5G). Local DE therefore resolved the direction and spatial extent of expression differences within the aligned NL3–IL3 section pair.

We next asked whether the localized expression differences were associated with compartment-specific functional patterns. We first applied Benjamini–Hochberg adjustment to the gene-level ACAT P values across the predeclared family of 16,446 genes and retained genes with $q_g^{\mathrm{ACAT}} \leq 0.05$. For each retained gene, we tested whether its significant local-grid locations were enriched within the annotated compartments and assigned each spatially concentrated gene to one dominant compartment–direction class. We then performed Gene Ontology (GO) Biological Process over-representation analysis for the genes assigned to each class (Methods 5.4.2; Supplementary Methods S3.12). The five highest-ranked cortex and medulla terms are shown in Fig. 5H, and the 15 highest-ranked terms for cortex, interface and medulla are shown in Supplementary Figs. S34–S35. The enrichments differed by both compartment and direction of change. Decreased-expression classes were enriched for metabolic processes in the cortex and medulla and for mitochondrial energy production at the interface. Increased-expression classes included glutathione metabolism in the cortex, extracellular-matrix organization and interleukin-4 responses at the interface, and translation and ribosome assembly in the medulla. For comparison, the corresponding ORA rankings from compartment-level DE (i.e., comparison of all spots within a compartment between the two samples) are shown in Supplementary Figs. S36–S37. Among the significant GO terms identified by local DE, 92% occurred in one compartment–direction class, 7% in two classes and 1% in at least three classes. The corresponding proportions for compartment-level DE ORA were 53%, 26% and 21%, respectively (Supplementary Fig. S38). Local DE therefore produced more compartment- and direction-specific functional enrichment patterns in this section pair.

Finally, we examined local differences in cell-type composition between the injured and normal kidney sections. Because the data were measured at spot resolution, we first estimated cell-type proportions within each spot using RCTD and the mouse kidney single-cell reference from Balzer et al. [47, 54] (Methods 5.4.2; Supplementary Methods S3.12). Mean estimated monocyte and granulocyte proportions were higher in injured tissue than in normal tissue (0.450% versus 0.075% and 0.244% versus 0.051%, respectively). Aggregating evidence from compartment-restricted local comparisons yielded cell-type-level spatial ACAT P values of 3.1×10^{-12} for monocytes and 2.1×10^{-12} for granulocytes. At the exploratory local threshold of $q < 0.20$, monocyte increases were detected at 20.64% of valid interface grid locations and 16.77% of valid medulla grid locations. Granulocyte increases were detected at 50.79% of valid medulla grid locations, whereas decreases occurred at 15.51% of valid interface grid locations (Fig. 5I). Thus, the higher overall granulocyte proportion was accompanied by opposing local changes across compartments. These results identify spatially heterogeneous changes in estimated myeloid-cell composition in this section pair, consistent with the inflammatory response to renal ischemia–reperfusion injury [55].

3 Discussion

We developed spAlignDE, an integrated framework for generalizable spatial alignment and mismatch-aware local differential expression analysis. spAlignDE unifies alignment across transcriptomics, epigenomics, histology and anatomical atlases through a common representation of tissue structure. It encodes this structure as continuous fields and uses geometric matching to guide S-LDDMM registration without requiring shared molecular features. For post-alignment inference, spAlignDE

incorporates residual mismatch into the variance of local expression contrasts, rather than treating aligned coordinates as exact. Across evaluated datasets, spAlignDE preserved global tissue geometry and local anatomy, scaled to a large multi-sample cohort, and aligned spatial transcriptomics with histology, anatomical atlases, and spatial ATAC-seq. Simulations showed that mismatch-aware variance inflation reduced false discoveries. The aging-brain analysis localized gene-expression and cell-neighborhood changes that section-wide summaries could obscure, and the injured-kidney analysis identified local expression patterns within anatomical compartments that compartment-level aggregation could obscure.

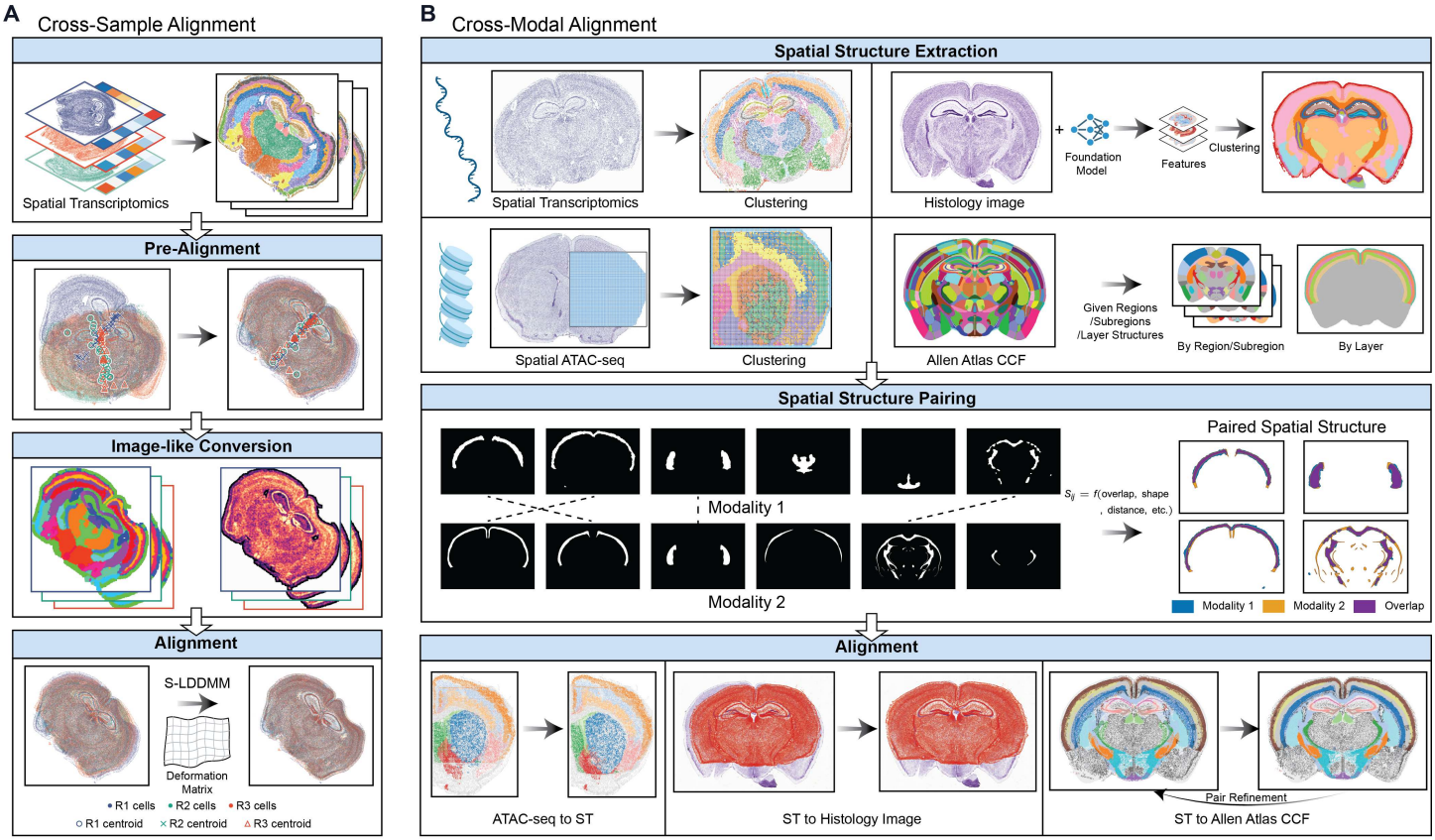
While spAlignDE proposes the first rigorous location-resolved statistical comparison, the interpretation of local expression contrasts depends on the testing target and experimental design. With the comparison-wide intercept included, spAlignDE tests whether the local contrast departs from the gene-specific section-wide baseline after optional adjustment for covariates like local library size and detection rate. Approximately uniform technical or biological shifts may be absorbed by the baseline. With one section per condition, the results describe sample-specific spatial contrasts: additional cells or spots improve local estimation but do not provide biological replication [56, 57]. For multiple aligned sections, the current workflow fits pairwise or sequential contrasts separately according to the scientific comparison. We anticipate that replicated spatial datasets will become more common and provide the data needed to develop and evaluate this extension for population-level inference. One straightforward solution is to incorporate random effect terms to model between-slice variation.

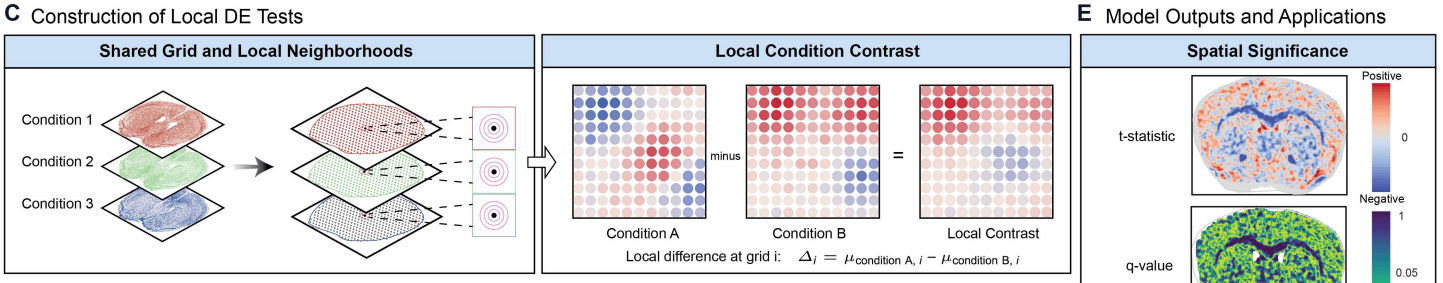
Structure-guided alignment also has several limitations. First, alignment accuracy depends on accurately identifying and matching spatial domains across samples or modalities. The same anatomical structure may be divided differently across datasets because of clustering errors or differences in segmentation granularity. For example, a single domain in one dataset may be divided into several subregions in another, and incorrect pairing of these regions can distort the estimated transformation. spAlignDE merges oversegmented regions automatically and lets users specify grouped one-to-many or many-to-many correspondences through an interactive interface. However, spAlignDE does not yet automatically infer general many-to-many mappings. Automatic resolution of one-to-many and many-to-many region correspondences would reduce sensitivity to clustering resolution and the need for manual specification.

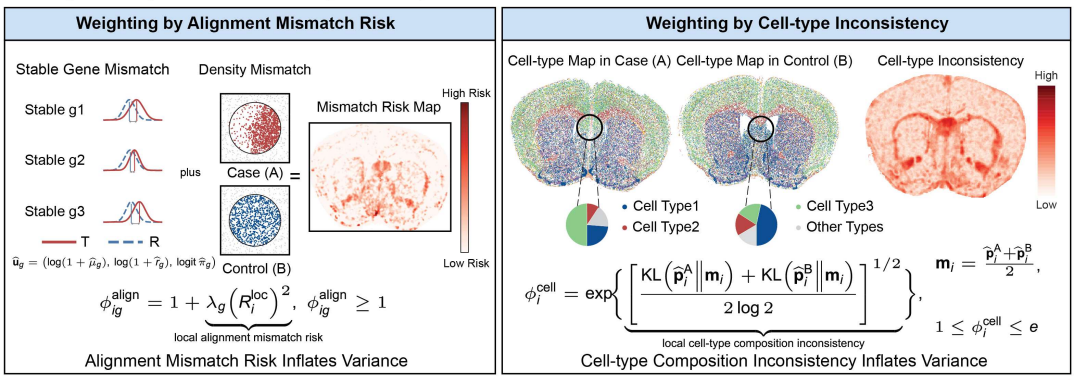
Second, the alignment result depends on the choice of reference sample. Because the reference defines the common coordinate system, its selection can affect both alignment and local inference. This sensitivity should be assessed when several suitable references are available. The reference must also share sufficient anatomy with the query, as S-LDDMM models correspondence through a smooth, one-to-one, topology-preserving transformation. Tissue damage can leave regions without valid counterparts, while differences in sectioning depth can introduce structures present in only one section. Downweighting these unmatched regions limits their influence on registration but cannot make them suitable for location-level comparisons. spAlignDE therefore focuses primarily on two-dimensional sections with sufficiently shared anatomy and locally comparable neighborhoods. Extending the current workflow to three-dimensional reconstruction will require modeling how structures appear or disappear across serial sections.

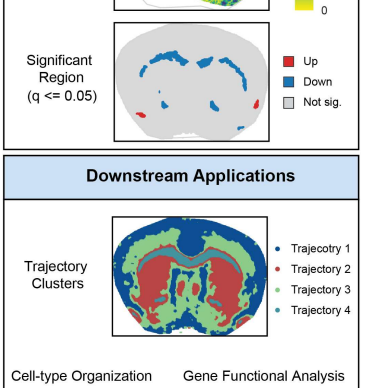
Lastly, we emphasize that the mismatch-risk score is a data-derived proxy for local comparability. It assumes accurately aligned neighborhoods have similar stable-gene expression and local observation density. Genuine biological differences may increase these discrepancies and reduce sensitivity through conservative variance inflation. Variance inflation also cannot fully compensate for severe misalignment that changes the biological neighborhoods being compared. The subsampling-based measure assesses the empirical stability of the estimated transformation, while calibration of alignment uncertainty remains a separate task. One future direction will be deriving uncertainty directly from the alignment model and propagating it through a joint framework for alignment and statistical inference.

4 Figures


A Cross-Sample Alignment
Spatial Transcriptomics
Pre-Alignment
Image-like Conversion
Alignment
S-LDDMM
Deformation Matrix
R1 cells
R2 cells
R3 cells
R1 centroid
R2 centroid
R3 centroid
B Cross-Modal Alignment
Spatial Structure Extraction
Spatial Transcriptomics
Clustering
Histology image
Foundation Model
Features
Clustering
Spatial ATAC-seq
Clustering
Allen Atlas CCF
Given Regions /Subregions /Layer Structures
By Region/Subregion
By Layer
Spatial Structure Pairing
Modality 1
Modality 2
Paired Spatial Structure
Modality 1
Modality 2
Overlap
Alignment
ATAC-seq to ST
ST to Histology Image
Pair Refinement
ST to Allen Atlas CCF


C Construction of Local DE Tests
Shared Grid and Local Neighborhoods
Condition 1
Condition 2
Condition 3
Local Condition Contrast
minus
Condition A
Condition B
Local Contrast
Local difference at grid i:
E Model Outputs and Applications
Spatial Significance
t-statistic
Positive
0
Negative
q-value
1
0.05
0
Significant Region (q <= 0.05)
Up
Down
Not sig.


D Weighted Regression for Post-Alignment Adjustment
Weighting by Alignment Mismatch Risk
Stable Gene Mismatch
Density Mismatch
Stable g1
Stable g2
Stable g3
T
R
plus
Case (A)
Control (B)
Mismatch Risk Map
High Risk
Low Risk
local alignment mismatch risk
Alignment Mismatch Risk Inflates Variance
Weighting by Cell-type Inconsistency
Cell-type Map in Case (A)
Cell-type Map in Control (B)
Cell-type Inconsistency
High
Low
Cell Type1
Cell Type2
Cell Type3
Other Types
local cell-type composition inconsistency
Cell-type Composition Inconsistency Inflates Variance


Downstream Applications
Trajectory Clusters
Trajectory 1
Trajectory 2
Trajectory 3
Trajectory 4
Cell-type Organization
Gene Functional Analysis
Region 1
Region 2
Pathway 1
Pathway 2
Pathway 3
Pathway 4
Pathway 5
High
Low

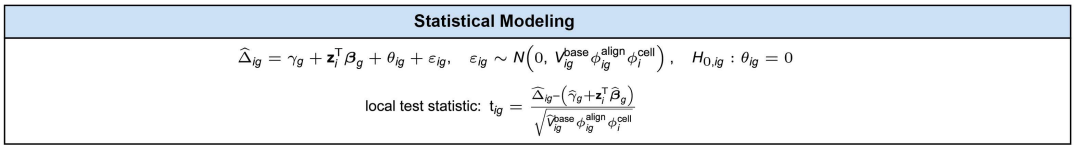

Statistical Modeling
local test statistic:

Fig. 1: **Overview of the spAlignDE framework. A,** Cross-sample alignment of spatial transcriptomics data. Joint clustering identifies shared spatial domains. These domains guide global pre-alignment and the construction of image-like representations for registration by shooting-based large deformation diffeomorphic metric mapping (S-LDDMM). **B,** Cross-modal alignment. Spatial structures are extracted from spatial transcriptomics, spatial assay for transposase-accessible chromatin using sequencing (spatial ATAC-seq), histology and anatomical atlases. Structures from different modalities are paired according to spatial and geometric similarity and aligned using S-LDDMM. **C,** Construction of local differential expression tests. Each retained location on the shared grid defines a testing unit. Kernel-weighted neighborhoods of aligned observations provide the condition-specific local means and baseline variance for the local contrast. **D,** Mismatch-aware local regression. Profiles of putatively stable genes characterize the local transformed mean, negative-binomial size and excess-zero probability. Cross-sample discrepancies in these quantities, together with local sampling-density discordance, define a normalized local alignment-mismatch risk. Gene-specific calibration converts this risk into the spatially varying variance factor ϕ_{ig}^{align}. Kernel-smoothed cell-type composition vectors define the optional inconsistency factor ϕ_i^{cell}. Weighted regression separates each estimated local contrast into a comparison-wide offset, an optional term for confounding variation such as effects of local library size and detection rate, and a local residual. The local statistic standardizes the residual $\widehat{\theta}_{ig}$ using the baseline variance scaled by $\phi_{ig}^{\text{align}}\phi_i^{\text{cell}}$. **E,** Model outputs and downstream applications. spAlignDE reports local test statistics, local q-values and connected significant regions. These outputs support analyses of spatial trajectories, cell-type organization and pathways. Parts of this figure were created with BioRender.

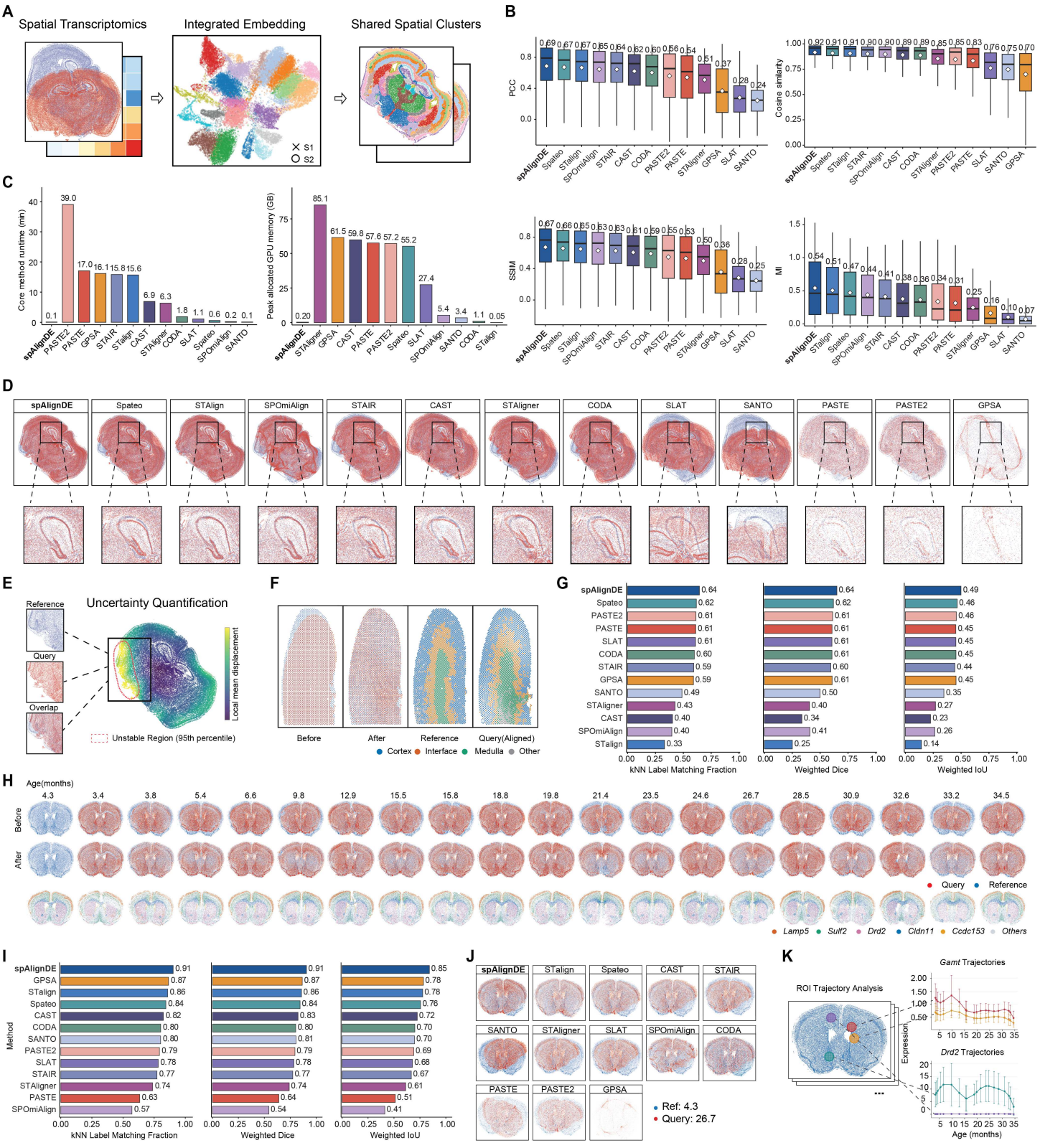

A
Spatial Transcriptomics
Integrated Embedding
Shared Spatial Clusters
S1
S2
B
PCC
Cosine similarity
SSIM
MI
C
Core method runtime (min)
Peak allocated GPU memory (GB)
D
spAlignDE
Spateo
STAlign
SPOmiAlign
STAIR
CAST
STAligner
CODA
SLAT
SANTO
PASTE
PASTE2
GPSA
E
Uncertainty Quantification
Reference
Query
Overlap
Unstable Region (95th percentile)
Local mean displacement
F
Before
After
Reference
Query(Aligned)
Cortex
Interface
Medulla
Other
G
kNN Label Matching Fraction
Weighted Dice
Weighted IoU
H
Age(months)
Before
After
Query
Reference
Lamp5
Sulf2
Drd2
Cldn11
Ccdc153
Others
I
Method
kNN Label Matching Fraction
Weighted Dice
Weighted IoU
J
Ref: 4.3
Query: 26.7
K
ROI Trajectory Analysis
Gamt Trajectories
Drd2 Trajectories
Expression
Age (months)

Fig. 2: **Cross-sample spatial alignment benchmarks and applications. A,** The spatial transcriptomics sections were integrated with Harmony and jointly clustered to define shared structures for alignment. Crosses and circles distinguish the two sections in the uniform manifold approximation and projection (UMAP). **B**–**E,** Vizgen multiplexed error-robust fluorescence in situ hybridization (MERFISH) Mouse Brain Receptor Map benchmark. S2R3 was aligned as the query to the S2R2 reference. **B,** Agreement of spatial gene-expression patterns across 13 methods for 483 shared genes, assessed by the Pearson correlation coefficient (PCC), cosine similarity, structural similarity index measure (SSIM) and mutual information (MI). Boxes indicate interquartile ranges and medians, whiskers extend to 1.5× the interquartile range and diamonds indicate means. Outliers are not shown, and methods are ordered by the mean for each metric. **C,** Core alignment runtime and peak allocated graphics processing unit (GPU) memory. Values are shown above the bars, and methods without recorded memory measurements are omitted. **D,** Whole-section and hippocampal alignment results for all 13 methods (reference, blue; aligned query, red). For **B**–**D**, PASTE and PASTE2 used 30,000 cells per section, and GPSA used 10,000 cells per section; all other methods used the complete sections containing 85,958 query cells and 84,172 reference cells. **E,** Pointwise transformation stability, measured by mean Euclidean displacement from the repeat-mean aligned position across ten repeated alignments using independent 80% cell subsamples. Contours mark locations above the 95th percentile. **F**–**G,** 10x Genomics Visium kidney benchmark. The injured IL3 section was aligned as the query to the normal NL3 reference. **F,** Sections before and after alignment with spAlignDE, together with tissue-compartment annotations. **G,** Region agreement across 13 methods, assessed by the k-nearest-neighbor label-matching fraction, weighted Dice coefficient and weighted intersection over union (IoU). **H**–**K,** Aging-brain MERFISH series comprising 20 sections from 3.4 to 34.5 months, all aligned to the 4.3-month reference. **H,** Sections before and after alignment, together with representative aligned gene-expression patterns. **I,** Region-annotation agreement across 19 source-to-reference alignments, assessed using the metrics in **G**. Bars indicate means across alignments. **J,** Alignment results for the 26.7-month section across all 13 methods (reference, blue; query, red). The query lacks the lower-right tissue boundary. **K,** Region-of-interest expression trajectories for *Gamt* and *Drd2* across age.

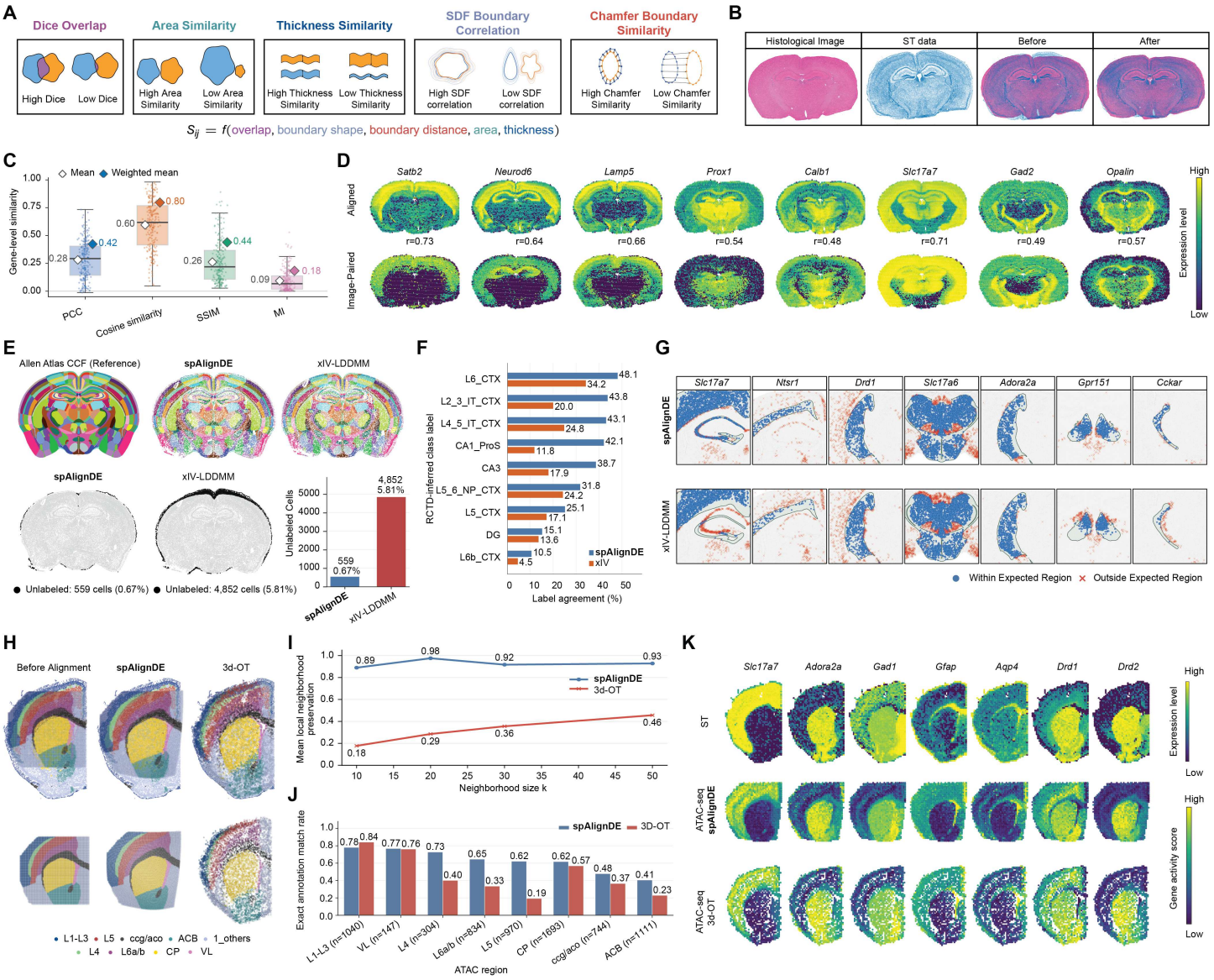


Fig. 3: **Cross-modal spatial alignment benchmarks. A,** Candidate spatial structures are paired according to a modality-specific score, s_{ij}, that combines overlap, boundary geometry, area and thickness. **B–D,** Alignment of Xenium spatial transcriptomics (ST) data to a hematoxylin and eosin (H&E)-stained image. **B,** H&E image, Xenium section and overlays before and after alignment with spAlignDE. **C,** Spatial concordance between aligned Xenium expression and Visium expression paired with the reference H&E image. Each dot represents one of $n = 246$ shared genes evaluated across 5,091 matched bins using the Pearson correlation coefficient (PCC), cosine similarity, structural similarity index measure (SSIM) and mutual information (MI). Boxes indicate interquartile ranges and medians; open and filled diamonds indicate unweighted and reliability-weighted means, respectively. **D,** Representative expression patterns in the aligned Xenium data and the H&E-paired Visium data. **E–G,** Alignment of ST data to the Allen Mouse Brain Common Coordinate Framework version 3 (CCFv3), compared with xIV-LDDMM. **E,** Atlas alignment and label transfer. The fixed-seed automatic workflow retained 18 matched spatial-structure pairs. Cells mapped outside the annotated atlas are shown in the bottom row. xIV-LDDMM left 4,852 cells unlabeled (5.81%), compared with 559 cells (0.67%) for spAlignDE. **F,** Agreement between transferred atlas labels and RCTD-inferred class labels. For each anatomically informative RCTD-inferred class, bars show the percentage of eligible ST cells assigned to one of its predefined compatible cortical, hippocampal or dentate-gyrus regions. **G,** Localization of regionally enriched marker genes. Blue circles and red crosses denote high-expression cells mapped inside and outside their expected atlas regions, respectively; outlines mark the expected regions. **H–K,** Alignment of a partial spatial assay for transposase-accessible chromatin using sequencing (spatial ATAC-seq) section to ST, compared with 3d-OT. **H,** Tissue overlays before and after alignment. **I,** Mean neighborhood preservation among 20-μm ATAC-seq pixels. For each pixel, neighborhood preservation is the fraction of its original K-nearest neighbors retained after alignment. **J,** Agreement between the original ATAC-seq anatomical annotations and labels inferred from the ten nearest ST observations after alignment. **K,** Representative ST expression and ATAC-derived gene-activity patterns after alignment with spAlignDE or 3d-OT.

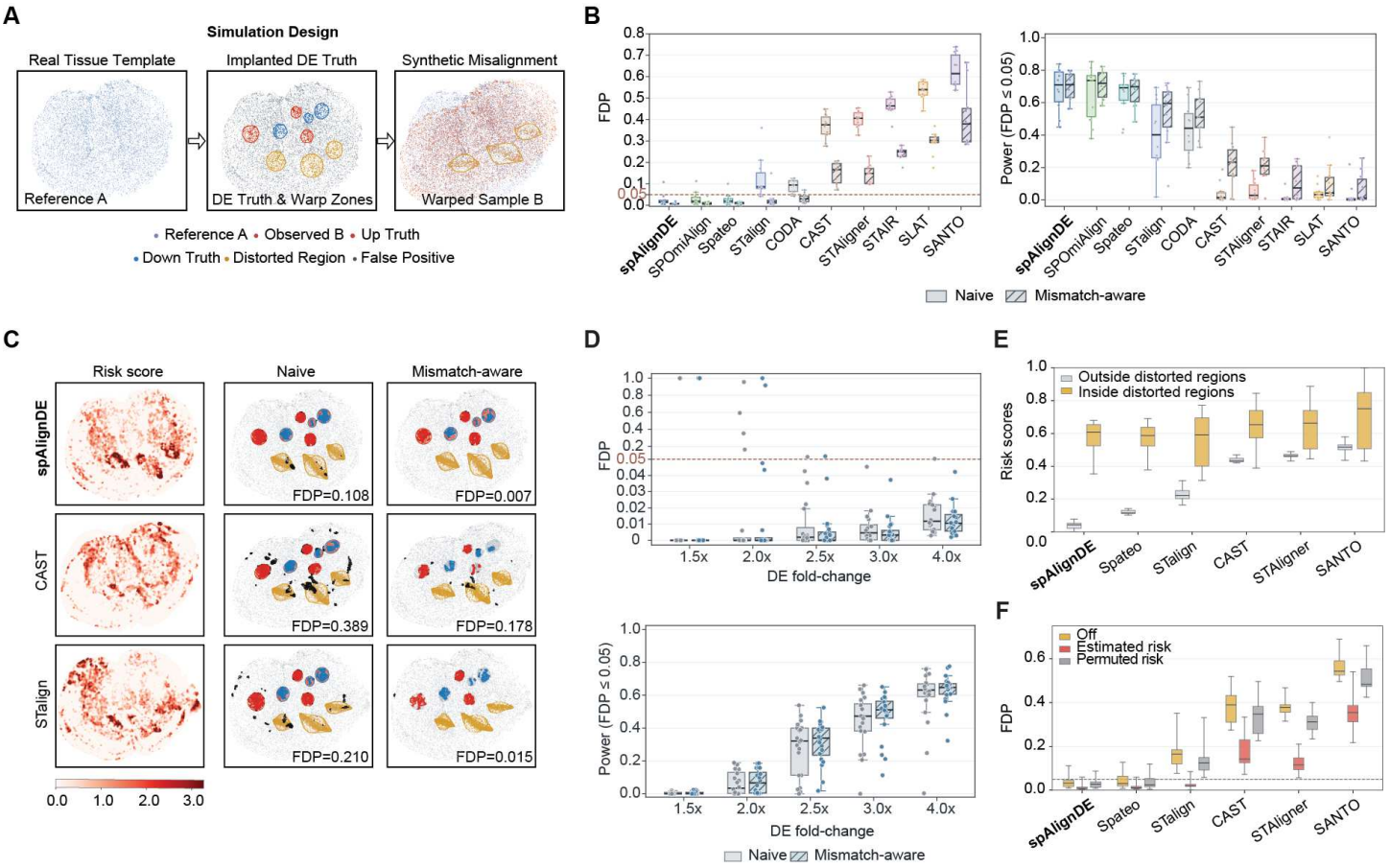


Fig. 4: **Simulation benchmark of mismatch-aware post-alignment spatial testing. A,** Semi-synthetic design using a real mouse brain tissue template. Local upregulated and downregulated regions and separate distorted regions were implanted in sample B before synthetic deformation. **B,** naïve and mismatch-aware testing across ten alignment methods and ten matched simulation groups. Left, location-level FDP at a target FDR of 0.05; the dashed line indicates the target. Grid calls and statistics were projected onto sample-B cells before evaluation. Right, maximum correctly signed true DE cell recovery among q-value-ranked call sets with FDP ≤ 0.05. Solid and hatched boxes indicate naïve and mismatch-aware testing, respectively; colors identify alignment methods. **C,** Representative results for spAlignDE, CAST and STalign, from top to bottom. Columns show the robust-standardized risk score ζ_i^R before capping and rescaling, naïve testing and mismatch-aware testing. Grid calls are projected onto sample-B cells. Red and blue denote correctly signed true positives in upregulated and downregulated regions; orange denotes distorted regions, black denotes false positives and gray denotes all cells. Values indicate location-level FDP at a target FDR of 0.05. **D,** Signal-strength sensitivity for spAlignDE across 20 matched templates per level. Upregulated fold-changes were 1.5×, 2.0×, 2.5×, 3.0× or 4.0×, with reciprocal downregulated effects. Top, FDP at a target FDR of 0.05. Bottom, power at FDP ≤ 0.05. The FDP axis allocates equal vertical space to 0–0.05 and 0.05–1 using a piecewise linear scale. **E,** Regional medians of pre-standardization mismatch scores outside and inside implanted distorted regions for six alignment methods across 15 matched simulation groups, normalized to $[0, 1]$ using a common min–max transformation. **F,** FDP at a target FDR of 0.05 for the same methods and simulation groups. "Off" indicates no mismatch adjustment; "Estimated risk" and "Permuted risk" indicate mismatch-aware testing with the estimated risk map and its spatially permuted version, respectively. Boxes show medians and interquartile ranges; whiskers follow the 1.5× interquartile-range rule in **B** and **D** and span minima to maxima in **E** and **F**.

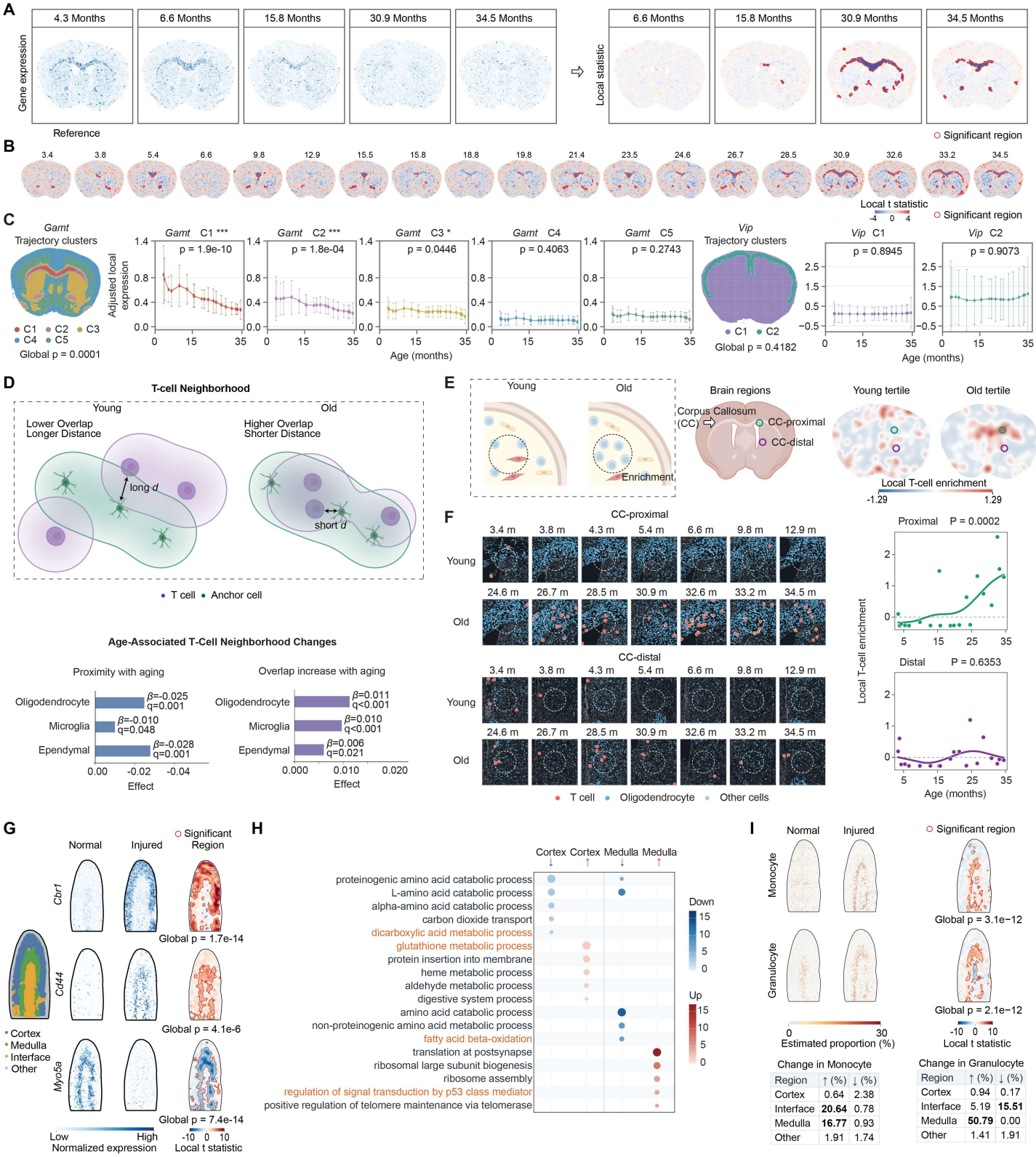

A
4.3 Months
6.6 Months
15.8 Months
30.9 Months
34.5 Months
Gene expression
Reference
Local statistic
Significant region
B
3.4
3.8
5.4
6.6
9.8
12.9
15.5
15.8
18.8
19.8
21.4
23.5
24.6
26.7
28.5
30.9
32.6
33.2
34.5
Local t statistic
Significant region
C
Gamt
Trajectory clusters
C1 C2 C3
C4 C5
Global p = 0.0001
Adjusted local expression
Gamt C1 ***
p = 1.9e-10
Gamt C2 ***
p = 1.8e-04
Gamt C3 *
p = 0.0446
Gamt C4
p = 0.4063
Gamt C5
p = 0.2743
Age (months)
Vip
Trajectory clusters
C1 C2
Global p = 0.4182
Vip C1
p = 0.8945
Vip C2
p = 0.9073
Age (months)
D
T-cell Neighborhood
Young
Old
Lower Overlap
Longer Distance
Higher Overlap
Shorter Distance
long d
short d
T cell
Anchor cell
Age-Associated T-Cell Neighborhood Changes
Proximity with aging
Oligodendrocyte
Microglia
Ependymal
β=-0.025 q=0.001
β=-0.010 q=0.048
β=-0.028 q=0.001
Effect
Overlap increase with aging
β=0.011 q<0.001
β=0.010 q<0.001
β=0.006 q=0.021
Effect
E
Young
Old
Enrichment
Brain regions
Corpus Callosum (CC)
CC-proximal
CC-distal
Young tertile
Old tertile
Local T-cell enrichment
-1.29
1.29
F
CC-proximal
Young
Old
3.4 m
3.8 m
4.3 m
5.4 m
6.6 m
9.8 m
12.9 m
24.6 m
26.7 m
28.5 m
30.9 m
32.6 m
33.2 m
34.5 m
CC-distal
T cell
Oligodendrocyte
Other cells
Proximal P = 0.0002
Distal P = 0.6353
Local T-cell enrichment
Age (months)
G
Normal
Injured
Significant Region
Cbr1
Cd44
Myo5a
Global p = 1.7e-14
Global p = 4.1e-6
Global p = 7.4e-14
Cortex
Medulla
Interface
Other
Low
High
Normalized expression
Local t statistic
H
Cortex ↓
Cortex ↑
Medulla ↓
Medulla ↑
proteinogenic amino acid catabolic process
L-amino acid catabolic process
alpha-amino acid catabolic process
carbon dioxide transport
dicarboxylic acid metabolic process
glutathione metabolic process
protein insertion into membrane
heme metabolic process
aldehyde metabolic process
digestive system process
amino acid catabolic process
non-proteinogenic amino acid metabolic process
fatty acid beta-oxidation
translation at postsynapse
ribosomal large subunit biogenesis
ribosome assembly
regulation of signal transduction by p53 class mediator
positive regulation of telomere maintenance via telomerase
Down
Up
I
Normal
Injured
Monocyte
Granulocyte
Estimated proportion (%)
Change in Monocyte
Region | ↑ (%) | ↓ (%)
Cortex | 0.64 | 2.38
Interface | 20.64 | 0.78
Medulla | 16.77 | 0.93
Other | 1.91 | 1.74
Significant region
Global p = 3.1e−12
Global p = 2.1e−12
Local t statistic
Change in Granulocyte
Region | ↑ (%) | ↓ (%)
Cortex | 0.94 | 0.17
Interface | 5.19 | 15.51
Medulla | 50.79 | 0.00
Other | 1.41 | 1.91

Fig. 5: **Mismatch-aware post-alignment inference in aging brain and injured kidney. A,** Representative *Gamt* expression and local t-statistic maps relative to the 4.3-month reference. **B,** Local *Gamt* statistics for all 19 non-reference sections. Red solid contours in A and B mark regions with local $q < 0.05$. **C,** Spatial trajectory clusters and adjusted expression trajectories for *Gamt* and *Vip*. Points show cluster means, and bars indicate the mean ± 1.96 spatial standard deviations. Labels give raw two-sided Wald P values for cluster-level age trends and gene-level spatial ACAT P values. **D,** Top, schematic illustrating T-cell proximity and spatial overlap in young and old tissue. Purple cells represent T cells, and green anchor cells represent the target population. Bottom, age-associated proximity to and spatial overlap with oligodendrocytes, microglia and ependymal cells. Proximity effects are the negatives of the labeled log-distance age slopes; positive overlap effects indicate increasing spatial overlap. Labels give BH-adjusted q values. **E,** Left, schematic illustrating local T-cell enrichment. Middle, schematic showing the callosal-proximal and callosal-distal ROI positions. Right, local T-cell enrichment maps calculated from young (≤ 13.6 months; $n = 7$) and old (≥ 24.3 months; $n = 7$) sections. Enrichment is shown in percentage points. Green and purple circles mark the callosal-proximal and callosal-distal ROIs, respectively. **F,** Spatial views of the two ROIs in young and old sections and their enrichment trajectories across age. Red, blue and gray points represent T cells, oligodendrocytes and other cells, respectively. Circles mark the ROIs. Trajectory points show section-level local enrichment in percentage points, curves show age-smoothed trajectories, and labels give age-association P values. **G,** Expression and mismatch-aware local differential expression for *Cbr1*, *Cd44* and *Myo5a* in normal NL3 and injured IL3 kidney sections. All six expression maps share a common color scale, and local t-statistic maps in G and I share a common color scale. Red contours mark local $q < 0.05$, and labels give raw gene-level ACAT P values. **H,** The five highest-ranked GO Biological Process terms for each displayed cortex or medulla compartment–direction class. Circle size indicates within-class rank, with the largest circle representing rank 1. Blue and red indicate decrease and increase, respectively; color intensity represents $-\log_{10}(q)$, where q is the BH-adjusted term P value. Blank entries indicate terms outside a class's top five. Orange labels denote kidney-injury-related terms annotated after ranking. **I,** RCTD-estimated monocyte and granulocyte proportions in normal and injured sections and their local composition statistics. Red contours mark joint BH-adjusted local $q < 0.20$, and labels give raw cell-type-level spatial ACAT P values. Tables report the percentages of valid grid locations within each compartment showing significant increases or decreases. Parts of this figure were created with BioRender.

5 Methods

5.1 Structure-guided spatial alignment

spAlignDE aligns a query spatial dataset to a fixed reference coordinate system through five stages: spatial-structure construction, global pre-alignment, structure-pair identification, conversion to continuous multichannel fields and shooting-based large deformation diffeomorphic metric mapping (S-LDDMM) [35–37]. The query is the dataset whose coordinates are transformed, whereas the reference remains fixed. The same workflow accommodates cross-sample spatial transcriptomics and cross-modal alignment.

5.1.1 Construction of modality-specific spatial structures Alignment begins by defining spatial structures in each dataset. Depending on the modality, these structures may correspond to spatial domains, anatomical annotations or image-derived segmentation. They are typically identified using spatial clustering or segmentation procedures that produce spatially coherent domains, although spAlignDE can also accept user-defined structures. The resulting labels are subsequently converted into masks or continuous structural channels for searching cross-modality structure correspondence and S-LDDMM registration.

Spatial transcriptomics. For cross-sample ST alignment, BANKSY combines gene-expression profiles with local spatial-neighborhood information within each section, after which the sample-specific representations are integrated with Harmony [33, 34]. Leiden clustering of the integrated representation assigns shared spatial-structure identities across sections; therefore, identically labeled structures have natural correspondence across samples. For cross-modality alignment, BANKSY clustering is performed independently within each ST sample, as correspondence is established later based on geometric compatibility. Detailed preprocessing and clustering settings are provided in Supplementary Methods S2.1.1.

Histological images. Histology-derived spatial structures are constructed from feature embeddings extracted using pretrained computational pathology models, including HIPT [58] and UNI [59]. Candidate representations based on HIPT embeddings alone or combined HIPT–UNI embeddings are clustered and spatially refined to obtain coherent tissue regions for alignment. Stain-specific preprocessing, feature-model configurations and final structure-map selection are described in Supplementary Methods S2.1.2.

Anatomical reference atlases. Atlas structures are obtained from predefined annotation labels and an anatomical hierarchy rather than inferred by clustering. In the Allen CCFv3 analysis, descendant labels are combined along hierarchical structure paths to generate candidate masks at multiple anatomical depths, with cortical layers additionally represented as layer-specific structures [11, 12]. Details are provided in Supplementary Methods S2.1.4.

Spatial ATAC-seq. Spatial ATAC-seq structures are identified by applying BANKSY to an ATAC-derived gene-activity score matrix and its spatial coordinates. The ST reference is clustered independently from its expression matrix, so ATAC and ST structures are connected only during geometric structure pairing. Gene-activity processing and clustering settings are provided in Supplementary Methods S2.1.3.

5.1.2 Global pre-alignment Global pre-alignment corrects large differences in tissue position, orientation and scale, allowing subsequent S-LDDMM registration to focus on residual affine differences and local nonlinear tissue deformation (Section 5.1.5). For cross-modal alignment, this correction also brings independently identified structures into approximate spatial correspondence, enabling meaningful geometric comparison and selection of candidate structure pairs (Section 5.1.3).

spAlignDE provides three strategies for global pre-alignment: shared-structure-centroid alignment, whole-tissue-mask-overlap alignment and interactive similarity alignment.

When shared structure identities are available, spAlignDE estimates a weighted similarity transformation from their centroids. Let $\boldsymbol{\mu}_r^{\mathrm{qry}}$ and $\boldsymbol{\mu}_r^{\mathrm{ref}}$ denote the query and reference centroids of shared structure r. The transformation is estimated by solving

$$(\widehat{s}, \widehat{\mathbf{R}}, \widehat{\mathbf{t}}) = \arg \min_{\substack{s>0,\ \mathbf{R}\in \mathrm{SO}(2),\\ \mathbf{t}\in\mathbb{R}^2}} \sum_{r\in\mathcal{R}_{\mathrm{shared}}} \omega_r \left\| s\mathbf{R}\boldsymbol{\mu}_r^{\mathrm{qry}} + \mathbf{t} - \boldsymbol{\mu}_r^{\mathrm{ref}} \right\|_2^2,$$

where s is an isotropic scaling factor, $\mathbf{R} \in \mathrm{SO}(2)$ is a two-dimensional rotation matrix that excludes reflection, and $\mathbf{t} \in \mathbb{R}^2$ is a translation vector. The set $\mathcal{R}_{\mathrm{shared}}$ contains the structure identities represented in both datasets. For structure r, let n_r^z denote the number of observations assigned to that structure in dataset $z \in \{\mathrm{qry}, \mathrm{ref}\}$. The weights satisfy $\omega_r \propto \min(n_r^{\mathrm{qry}}, n_r^{\mathrm{ref}})$ and are normalized to sum to one.

When shared structure identities are unavailable but the datasets have comparable overall tissue shapes, spAlignDE estimates global alignment by maximizing whole-tissue-mask overlap. Tissue centers determine translation, while rotation

and scale are selected through an overlap-based search. Let $\mathbf{c}_{\mathrm{qry}}$ and $\mathbf{c}_{\mathrm{ref}}$ denote the estimated query and reference tissue centers. For candidate scale $s > 0$ and rotation angle θ, a query coordinate $\mathbf{x}$ is transformed as

$$T_{s,\theta}(\mathbf{x}) = s\mathbf{R}(\theta)\,(\mathbf{x} - \mathbf{c}_{\mathrm{qry}}) + \mathbf{c}_{\mathrm{ref}},$$

where $\mathbf{R}(\theta)$ is the two-dimensional rotation matrix associated with θ. This formulation maps the query center to the reference center, determining translation for each candidate s and θ without an independent translation search. The transformed query coordinates are rasterized to obtain the binary mask $\mathcal{M}^{\mathrm{qry}}_{\mathrm{tissue}}(s,\theta)$ on the same grid as the fixed reference mask $\mathcal{M}^{\mathrm{ref}}_{\mathrm{tissue}}$. Their intersection over union is

$$\mathrm{IoU}(s,\theta) = \frac{\left|\mathcal{M}^{\mathrm{qry}}_{\mathrm{tissue}}(s,\theta) \cap \mathcal{M}^{\mathrm{ref}}_{\mathrm{tissue}}\right|}{\left|\mathcal{M}^{\mathrm{qry}}_{\mathrm{tissue}}(s,\theta) \cup \mathcal{M}^{\mathrm{ref}}_{\mathrm{tissue}}\right|},$$

where $|\cdot|$ denotes the number of foreground pixels. The candidate rotation and scale yielding the largest IoU are retained.

When neither automatic strategy provides reliable global correspondence, spAlignDE supports interactive similarity pre-alignment, allowing users to manually adjust translation, rotation and isotropic scale.

In the analyses reported here, centroid-based pre-alignment was used for the primary cross-sample ST analyses. Whole-tissue-mask-overlap pre-alignment was used for ST-to-atlas and Xenium-to-H&E alignment. Interactive pre-alignment was used for MERFISH-to-Nissl and spatial ATAC-to-ST alignment, as well as the Xenium breast-cancer ST-to-ST partial-alignment example. Exact search ranges, tissue-center estimation, mask operations and interface behavior are described in Supplementary Methods S2.3.1.

5.1.3 Structure-pair identification After global pre-alignment, spAlignDE determines which query and reference structures provide corresponding anatomical support. For cross-sample alignment, joint clustering in a batch-adjusted representation yields shared structure identities that directly define query–reference correspondences. For cross-modal alignment, structures are identified independently and represented as masks on a common raster domain after global pre-alignment. Correspondences are then established using a composite score that combines multiple measures of geometric similarity between candidate structures. Only candidate pairs satisfying the pairing criteria are retained, allowing structures without suitable counterparts to remain unmatched.

Cross-sample correspondence. For cross-sample ST alignment, shared structure labels obtained through joint clustering define query–reference correspondences. Only structures represented in both samples are used for correspondence-based pre-alignment and transformation estimation. The resulting transformation is applied to all original observations, including those outside the shared structures.

Cross-modal correspondence. Candidate cross-modal structures are represented as binary masks on a shared raster domain. For point-based structures derived from cells, spots or bins, observations are filtered, rasterized, adaptively smoothed and refined to connect small gaps, fill holes and remove isolated components. Filtering is used only for mask construction; all original observations receive the final transformation. Histology and atlas masks are obtained directly from their labeled image regions. Detailed mask-construction procedures are provided in Supplementary Methods S2.3.2.

For query structure i and reference structure j, geometric correspondence is quantified using the composite score

$$S^{\mathrm{comp}}_{ij} = \sum_{m \in \mathcal{C}} w_m C^{(m)}_{ij}, \qquad \mathcal{C} = \{\mathrm{Dice}, \mathrm{SDF}, \mathrm{Ch}, \mathrm{area}, \mathrm{thick}\}, \quad w_m \geq 0, \quad \sum_{m \in \mathcal{C}} w_m = 1.$$

The five components capture complementary aspects of geometric similarity. *Dice similarity* measures foreground-mask overlap. *Signed-distance-field (SDF) similarity* compares signed distances within a boundary band to assess boundary geometry under modest displacement. *Chamfer similarity* converts the mean bidirectional nearest-boundary distance into a decreasing similarity score. *Area similarity* assesses agreement in spatial extent through a symmetric relative-size comparison. *Thickness similarity* compares distance-transform summaries of characteristic width and is particularly informative for narrow or layer-like structures.

Candidate pairs are screened using modality-specific criteria, including minimum overlap or maximum boundary displacement where appropriate. Candidates that pass screening are ranked by their composite scores and selected greedily. Each primitive query and reference structure is used at most once, unless structures have first been combined into an explicit correspondence group. Exact component definitions, weights, thresholds and soft penalties are provided in Supplementary Methods S2.4. When automatic pairing is ambiguous, users can interactively define correspondence groups, which enter the same SDT construction and S-LDDMM registration workflow (Supplementary Methods S2.5).

5.1.4 Conversion to continuous multichannel fields S-LDDMM registration in spAlignDE operates on continuous multichannel fields, requiring discrete spatial observations and their structure labels to be converted into continuous representations. For cross-sample ST, observations assigned to each shared structure are rasterized and spatially smoothed to construct a local structure-composition channel. When spatial sampling is sufficiently dense, an additional channel represents

the smoothed local density of observations assigned to shared structures. This density channel is omitted when sampling is too sparse to support reliable density estimation.

For cross-modal alignment, the discrete cells, spots or pixels assigned to each accepted structure are first rasterized as binary masks on a common spatial domain. Directly matching binary masks provides piecewise-constant fields whose spatial variation is concentrated at structure boundaries and can therefore provide limited guidance when paired structures remain separated after global pre-alignment. Each paired mask is instead converted into a signed-distance-transform (SDT) field, whose magnitude records distance to the nearest boundary and whose sign distinguishes the structure interior from the exterior. The SDT converts boundary displacement into a spatially graded matching signal, providing useful correspondence information even before the paired boundaries overlap and increasing the effective capture range of the deformable registration [60]. Smoothing, distance truncation and normalization reduce sensitivity to irregular mask boundaries and prevent distant interior or background locations from dominating the matching objective.

The query and reference fields are sampled on the same regular raster and arranged in the same channel order. For cross-sample alignment, the raster covers the joint spatial extent of the pre-aligned query and reference, and its dimensions are determined by the selected grid spacing rather than by the number of cells or spots. For cross-modal alignment, the raster is inherited from the fixed reference image, atlas section or common pre-alignment canvas, with optional resolution adjustment while preserving the same coordinate extent. Let $\Omega \subset \mathbb{R}^2$ denote this common raster domain and let $F^{\mathrm{qry}}, F^{\mathrm{ref}} : \Omega \to \mathbb{R}^C$ denote the query and reference fields, where C is the total number of input channels. Detailed rasterization, normalization and grid-construction procedures are provided in Supplementary Methods S2.6.

5.1.5 Structure-guided shooting LDDMM spAlignDE registers the corresponding continuous multichannel fields using shooting-based large deformation diffeomorphic metric mapping (S-LDDMM) to refine global alignment and resolve local tissue deformation. Whereas conventional time-discretized LDDMM directly optimizes a sequence of velocity fields, S-LDDMM generates the full deformation trajectory from a single initial momentum field. This formulation reduces the number of independently optimized deformation variables while retaining the smooth, invertible and topology-preserving transformations of LDDMM [35, 36].

To accommodate missing tissue, incomplete overlap and structures without valid counterparts, spAlignDE adapts the probabilistic missing-data formulation of Tward et al. [37] to multichannel structural fields. A three-component mixture, fitted by expectation–maximization, distinguishes matched locations from two unmatched appearance components. The posterior probability of belonging to the matched component, $W_M(\mathbf{x})$, weights each reference-grid location in the field-matching objective. Locations with low match probabilities therefore contribute less to transformation estimation, reducing their influence on the inferred deformation.

Registration jointly estimates an initial momentum field and a residual affine transformation. The shooting equations generate the nonlinear deformation from the initial momentum, while optimization balances the probability-weighted multichannel matching loss against an LDDMM kinetic-energy penalty. The resulting query-to-reference mapping is

$$T_{\mathrm{final}} = A \circ \phi_1 \circ T_{\mathrm{pre}},$$

where T_{pre} is the global pre-alignment, ϕ_1 is the diffeomorphic transformation generated by shooting, and A corrects residual affine differences. This mapping is applied to all original query observations. The shooting equations, matched/unmatched mixture, channel mapping, regularization and optimization procedure are detailed in Supplementary Methods S2.7. For ST-to-atlas alignment, structure pairing and S-LDDMM registration are repeated across progressively finer ST partitions, while all eligible atlas-hierarchy candidates remain available at each stage (Supplementary Methods S2.8).

5.2 Mismatch-aware post-alignment inference

The alignment component maps spatial transcriptomic samples into a shared coordinate system. The post-alignment model then defines contrasts according to the scientific comparison rather than the direction used to estimate the alignment. Throughout the post-alignment inference model, sample A denotes the case, treatment or time point of interest, and sample B denotes the control or baseline condition. These inferential labels are independent of which sample was treated as the moving query or fixed reference during alignment. Cells or spots provide the local observations, shared-grid locations are the local testing units, tissue sections are the biological sample units, and case–control or sequential condition comparisons define the fitted contrasts. To keep the notation readable, the location-level model below is written for a generic single contrast and suppresses the contrast index c; c is restored only where multiple contrasts are calibrated or analyzed jointly.

5.2.1 Definition and interpretation of the local DE test For gene g at grid location i, let $\mu_{\mathrm{A},ig}$ and $\mu_{\mathrm{B},ig}$ denote the underlying local means in the case and control samples, respectively. The directional contrast is

$$\Delta_{ig} = \mu_{\mathrm{A},ig} - \mu_{\mathrm{B},ig}.$$

The primary local estimand is the departure of this contrast from a comparison-wide baseline and, when enabled, a grid-varying confounding-covariate term,

$$\theta_{ig} = \Delta_{ig} - \left(\gamma_g + \mathbf{z}_i^\top \boldsymbol{\beta}_g\right), \qquad H_{0,ig} : \theta_{ig} = 0.$$

Here, γ_g represents the component of the between-section contrast that is approximately constant over the tested grid, while $\mathbf{z}_i^\top \boldsymbol{\beta}_g$ optionally captures confounding variation such as effects of local library size and detection rate. Including γ_g defines local DE as a spatial departure from the comparison-wide shift; the intercept can be omitted when an approximately uniform difference is itself part of the inferential target.

Cells or spots supply the local measurements used to estimate a section-level spatial contrast but do not constitute independent biological replicates. In an ordered series such as the aging-brain analysis, each non-baseline age section is treated as sample A and the common baseline-age section as sample B before evidence is aggregated across sections. The current formulation does not estimate within-age replication because only one section was available at each age.

5.2.2 Shared grid, kernel estimators and base local-contrast variance Because aligned cells or spots are not themselves one-to-one matched observations, local contrasts are evaluated on a shared grid of common spatial locations. Let $\boldsymbol{\xi}_i$, $i = 1, \ldots, N_{\text{grid}}$, denote candidate grid locations in the aligned coordinate system. The retained testing set consists of locations passing the common-support mask and sample-specific minimum-support filters; the exact support-mask construction and analysis-specific grid settings are reported in Supplementary Methods S3.1 and S3.2. Let $\widetilde{\mathbf{q}}_j^{(s)}$ and $X_{jg}^{(s)}$ denote the aligned coordinate and expression of observation j in sample s. Around retained grid location i, spAlignDE defines the Gaussian-weighted neighborhood

$$w_{ij}^{(s)} = \exp\left[-\frac{\|\widetilde{\mathbf{q}}_j^{(s)} - \boldsymbol{\xi}_i\|^2}{2h_{\text{loc}}^2}\right] \mathbf{1}\left\{\|\widetilde{\mathbf{q}}_j^{(s)} - \boldsymbol{\xi}_i\| \le h_{\text{loc}}\right\}.$$

Here, h_{loc} is the local spatial bandwidth; it sets the Gaussian distance scale and excludes aligned observations farther than h_{loc} from grid location i. With normalized weights $a_{ij}^{(s)} = w_{ij}^{(s)} / \sum_j w_{ij}^{(s)}$, the local mean estimator is $\sum_j a_{ij}^{(s)} X_{jg}^{(s)}$. Under a working local approximation in which contributing observations are independent and have common variance σ^2, its variance is

$$\operatorname{Var}\left(\sum_j a_{ij}^{(s)} X_{jg}^{(s)}\right) = \sigma^2 \sum_j \left(a_{ij}^{(s)}\right)^2 = \frac{\sigma^2}{n_{s,i}^{\text{eff}}},$$

which gives the Kish effective local sample size [61],

$$n_{s,i}^{\text{eff}} = \frac{1}{\sum_j \left(a_{ij}^{(s)}\right)^2} = \frac{\left(\sum_j w_{ij}^{(s)}\right)^2}{\sum_j \left(w_{ij}^{(s)}\right)^2}.$$

Thus, concentrated kernel weights provide less effective information than equal weights over the same number of observations. This calculation is a working local-sampling approximation: residual spatial correlation among nearby cells or spots is not modeled explicitly, and overlapping neighborhoods induce dependence among neighboring grid tests.

The kernel-weighted local mean, variance and contrast are

$$\widehat{\mu}_{s,ig} = \frac{\sum_j w_{ij}^{(s)} X_{jg}^{(s)}}{\sum_j w_{ij}^{(s)}}, \qquad \widehat{\sigma}_{s,ig}^2 = \frac{\sum_j w_{ij}^{(s)} \left(X_{jg}^{(s)} - \widehat{\mu}_{s,ig}\right)^2}{\sum_j w_{ij}^{(s)}},$$

$$\widehat{\Delta}_{ig} = \widehat{\mu}_{\text{A},ig} - \widehat{\mu}_{\text{B},ig}.$$

The effective-sample-size-adjusted Welch-type variance estimate of the local contrast is

$$\widehat{V}_{ig}^{\text{base}} = \frac{\widehat{\sigma}_{\text{A},ig}^2}{n_{\text{A},i}^{\text{eff}}} + \frac{\widehat{\sigma}_{\text{B},ig}^2}{n_{\text{B},i}^{\text{eff}}}.$$

For gene-expression inference, grid locations with insufficient effective support in either sample are excluded. The effective spatial resolution is therefore determined by observation density relative to h_{loc}, not by shared-grid density alone. In ordered multi-sample analyses, each case–baseline or sequential contrast is fitted independently; local means are not pooled across contrasts.

5.2.3 Putatively stable genes and local mismatch-risk estimation The mismatch-risk map is estimated separately for each contrast from a panel of *putatively stable* genes. These genes are not assumed to be known negative controls; rather, they are selected as internal proxies for local comparability because they show informative spatial variation, positive cross-sample spatial correspondence and a small comparison-wide difference. For the generic single contrast considered here, define

$$\widehat{\ell}_{s,ig} = \log(1 + \widehat{\mu}_{s,ig}), \qquad \bar{\widehat{\ell}}_{s,g} = |\mathcal{I}|^{-1} \sum_{i \in \mathcal{I}} \widehat{\ell}_{s,ig}.$$

The putatively stable panel is

$$\mathcal{G}_0 = \left\{ g : \begin{array}{l} \max\left[\mathrm{Var}_i(\widehat{\ell}_{A,ig}), \mathrm{Var}_i(\widehat{\ell}_{B,ig})\right] \geq \tau_{\mathrm{var}}, \\ \left|\bar{\widehat{\ell}}_{A,g} - \bar{\widehat{\ell}}_{B,g}\right| \leq \tau_{\Delta}, \\ \mathrm{Corr}_i(\widehat{\ell}_{A,ig}, \widehat{\ell}_{B,ig}) \geq \tau_{\rho} \end{array} \right\}.$$

Exact thresholds, ranking rules and the maximum panel size are given in Supplementary Methods S3.3.

The risk map is constructed once for each pairwise or sequential contrast and is then shared across all target-gene fits for that contrast; it is not recomputed after removing the tested gene. Thus, if a target gene also belongs to $\mathcal{G}_0$, it remains in the panel used to construct the common risk map. Its three distributional channels form only one part of a pooled profile containing the corresponding channels from the full stable-gene panel together with the density channel, so its inclusion does not make the risk score gene specific. The use of empirically selected stably expressed genes as negative-control features for estimating shared unwanted variation follows related precedent in single-cell integration [38].

For a putatively stable gene $g \in \mathcal{G}_0$, let $\widehat{v}_{s,ig}$ denote its kernel-weighted local variance. A moment estimator of the local negative-binomial size is

$$\widehat{r}_{s,ig} = \frac{\widehat{\mu}_{s,ig}^2}{\max\{\widehat{v}_{s,ig} - \widehat{\mu}_{s,ig}, \epsilon_r\}},$$

with corresponding fitted zero probability

$$\widehat{q}_{s,ig}^{\mathrm{NB}} = \left(1 + \frac{\widehat{\mu}_{s,ig}}{\widehat{r}_{s,ig}}\right)^{-\widehat{r}_{s,ig}}.$$

If $p_{s,ig}^0$ is the kernel-weighted observed zero fraction, the excess-zero probability is

$$\widehat{\pi}_{s,ig} = \min\left[1 - \epsilon_\pi, \max\left\{0, \frac{p_{s,ig}^0 - \widehat{q}_{s,ig}^{\mathrm{NB}}}{\max(1 - \widehat{q}_{s,ig}^{\mathrm{NB}}, \epsilon_\pi)}\right\}\right].$$

The three-channel local profile

$$\widehat{\mathbf{u}}_{s,ig} = [\log(1 + \widehat{\mu}_{s,ig}), \log(1 + \widehat{r}_{s,ig}), \mathrm{logit}(\widehat{\pi}_{s,ig})]^\top$$

summarizes local expression magnitude, overdispersion and excess sparsity. Each channel is standardized across valid grid locations within its own sample; let $\mathcal{Z}_s$ denote this sample-specific standardization. Local sampling density is summarized by

$$d_{s,i} = \log\left\{1 + \sum_j \exp\left[-\frac{\|\widetilde{\mathbf{q}}_j^{(s)} - \boldsymbol{\xi}_i\|^2}{2h_{\mathrm{loc}}^2}\right] \mathbf{1}\left\{\|\widetilde{\mathbf{q}}_j^{(s)} - \boldsymbol{\xi}_i\| \leq h_{\mathrm{loc}}\right\}\right\}.$$

Let $d_{\mathrm{gene}} = 3|\mathcal{G}_0|$ denote the number of standardized stable-gene channels. For a target density-energy share $s \in (0, 1)$, the density multiplier is

$$\eta = \left(\frac{s}{1 - s} d_{\mathrm{gene}}\right)^{1/2}.$$

Thus, s controls the target contribution of the standardized density channel to the combined profile energy; it is not a direct weight on the final risk map or variance factor. The augmented profile is

$$\mathbf{U}_{s,i} = [\{\mathcal{Z}_s(\widehat{\mathbf{u}}_{s,ig}) : g \in \mathcal{G}_0\}, \eta \mathcal{Z}_s(d_{s,i})].$$

The raw mismatch score is the symmetric cosine dissimilarity

$$M_i = 1 - \frac{\mathbf{U}_{\mathrm{A},i}^\top \mathbf{U}_{\mathrm{B},i}}{\|\mathbf{U}_{\mathrm{A},i}\| \, \|\mathbf{U}_{\mathrm{B},i}\|}.$$

It is smoothed over the grid,

$$\widetilde{M}_i = \frac{\sum_{i'} G_\tau(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_{i'}\|) M_{i'}}{\sum_{i'} G_\tau(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_{i'}\|)},$$

and robustly normalized within the contrast,

$$\zeta_i^R = \left[\frac{\widetilde{M}_i - \mathrm{median}_{j\in\mathcal{I}}(\widetilde{M}_j)}{\mathrm{MAD}_{j\in\mathcal{I}}(\widetilde{M}_j) + \epsilon_R} \right]_+, \qquad R_i^{\mathrm{loc}} = \min\left\{ \frac{\zeta_i^R}{Q_{0.95}(\zeta_j^R : j \in \mathcal{I}) + \epsilon_R}, 1 \right\}.$$

Thus, $R_i^{\mathrm{loc}} \in [0,1]$. This score ranks relative local non-comparability within a contrast; it is neither a calibrated probability of misalignment nor a gene-specific biological effect.

5.2.4 Empirical mismatch calibration and cell-type-composition adjustment For target gene g, the alignment-mismatch variance factor for the generic single contrast is

$$\phi_{ig}^{\mathrm{align}} = 1 + \lambda_g \left(R_i^{\mathrm{loc}}\right)^2, \qquad \lambda_g \geq 0.$$

The location-specific risk map R_i^{loc} identifies where residual alignment mismatch may reduce precision within the current contrast, and the gene-specific coefficient λ_g determines how strongly this local risk inflates the variance for gene g. Calibration is based on the expectation that, if residual mismatch adds uncertainty, higher-risk strata will show greater excess dispersion in the initial local statistics relative to their corresponding Student-t null distributions.

The quadratic form is motivated by a local Taylor approximation of a smooth spatial expression field. A small alignment displacement $\boldsymbol{\delta}_i$ gives

$$\text{expression-error magnitude} \propto \|\boldsymbol{\delta}_i\|, \qquad \text{additional variance} \propto \|\boldsymbol{\delta}_i\|^2.$$

We therefore model alignment-induced variance inflation as quadratic in the local mismatch risk. This form also limits inflation at low-risk locations while concentrating the adjustment in regions with stronger mismatch.

To describe calibration in a form that covers both single- and multi-contrast analyses, we restore the contrast index c. The local model is first fitted with mismatch inflation disabled. Each contrast is calibrated separately using its initial local statistics $t_{igc}^{(0)}$ and risk map R_{ic}^{loc}. Locations are grouped into risk strata, the statistics are median-centered within each stratum and their robust scale is compared with the corresponding Student-t null scale. For bin b, with median risk r_{bc}, size n_{bc} and residual degrees of freedom ν_{gc},

$$y_{bgc} = \left[\left\{ \frac{\mathrm{MAD}_{bgc}}{F^{-1}_{t_{\nu_{gc}}}(0.75)} \right\}^2 - 1 \right]_+.$$

Positive excess dispersion is constrained to be nondecreasing with risk by weighted isotonic regression,

$$\widetilde{\mathbf{y}}_{gc} = \arg\min_{z_1 \leq \cdots \leq z_B} \sum_b n_{bc}(z_b - y_{bgc})^2,$$

and a non-negative quadratic relation through the origin is fitted,

$$\widehat{B}_{gc} = \arg\min_{B\geq 0} \sum_b n_{bc} \left(\widetilde{y}_{bgc} - B r_{bc}^2\right)^2.$$

A bounded reference-bin adjustment converts $\widehat{B}_{gc}$ into a provisional contrast-specific coefficient $\widetilde{\lambda}_{gc}$. In a multi-contrast analysis, valid provisional coefficients are combined by an equal-weight Huber robust center to obtain one shared gene-specific coefficient $\widehat{\lambda}_g$. This avoids dependence on an arbitrarily selected anchor contrast and limits the influence of a contrast with unusually strong biological change or unstable calibration. If only one contrast is available, its valid provisional coefficient is used directly, leaving the single-contrast procedure unchanged. The resulting $\widehat{\lambda}_g$ is held fixed when the final variances and local statistics are constructed, while every contrast retains its own risk map. Exact validity, aggregation and fallback rules are provided in Supplementary Methods S3.4.

This calibration assumes that risk-associated increases in the dispersion of the median-centered initial local statistics reflect residual mismatch-induced variability rather than systematic true local signal. The risk map, $\widehat{\lambda}_g$ and the confounding-covariate basis are treated as plug-in quantities in the final test; their estimation uncertainty is not propagated separately.

Cell-type-composition adjustment. Returning to the generic single-contrast notation, when cell-type annotations or spot-level deconvolution estimates are available, let $\widehat{\mathbf{p}}_i^{\mathrm{A}}$ and $\widehat{\mathbf{p}}_i^{\mathrm{B}}$ denote the kernel-smoothed local cell-type-composition vectors in the case and control samples. Their midpoint, normalized Jensen–Shannon distance and resulting variance factor are

$$\begin{aligned}
\mathbf{m}_i &= \frac{\widehat{\mathbf{p}}_i^{\mathrm{A}} + \widehat{\mathbf{p}}_i^{\mathrm{B}}}{2}, \\
D_i^{\mathrm{cell}} &= \left[\frac{\mathrm{KL}\left(\widehat{\mathbf{p}}_i^{\mathrm{A}} \| \mathbf{m}_i\right) + \mathrm{KL}\left(\widehat{\mathbf{p}}_i^{\mathrm{B}} \| \mathbf{m}_i\right)}{2\log 2} \right]^{1/2}, \\
\phi_i^{\mathrm{cell}} &= \exp\left(D_i^{\mathrm{cell}}\right), \\
\widehat{V}_{ig} &= \widehat{V}_{ig}^{\mathrm{base}} \phi_{ig}^{\mathrm{align}} \phi_i^{\mathrm{cell}}.
\end{aligned}$$

The Kullback–Leibler divergences use natural logarithms. The Jensen–Shannon divergence ranges from zero to $\log 2$, so division by $\log 2$ maps it to $[0,1]$. Its square root gives the bounded composition distance $D_i^{\mathrm{cell}} \in [0,1]$. The exponential link then maps this distance to $1 \le \phi_i^{\mathrm{cell}} \le e$. Identical local compositions give a variance factor of one, and increasing compositional inconsistency smoothly increases the local variance. The factor governs inferential precision; cell-type-specific expression effects require a separate model. Local composition estimation and fallback settings are described in Supplementary Methods S3.5.

5.2.5 Local regression with optional confounding-covariate adjustment For sample s and grid location i, let L_{is} and D_{is} denote local library size and detection rate, and define

$$\mathbf{q}_{is} = [\log(1 + L_{is}), D_{is}]^\top.$$

Sample B, or the designated set of control samples when several controls are supplied, defines the centering and scaling of these quality-control profiles. Singular value decomposition of the standardized control profiles supplies the retained directions, and the pooled sample-A batch profile is projected onto those directions to give $\mathbf{z}_i$. The resulting basis captures confounding variation such as effects of local library size and detection rate. Its construction is independent of which sample served as the query or reference during alignment and is detailed in Supplementary Methods S3.6.

For one contrast, the working observation model is

$$\widehat{\Delta}_{ig} = \gamma_g + \mathbf{z}_i^\top \boldsymbol{\beta}_g + \theta_{ig} + \varepsilon_{ig}, \qquad \varepsilon_{ig} \sim N(0, \widehat{V}_{ig}).$$

Let

$$\boldsymbol{\alpha}_g = \left(\gamma_g, \boldsymbol{\beta}_g^\top\right)^\top$$

and let $\boldsymbol{\alpha}_g^{(0)}$ denote the gene-specific anchor estimate described in Supplementary Methods. The baseline coefficients are estimated by

$$\widehat{\boldsymbol{\alpha}}_g = \arg\min_{\boldsymbol{\alpha}} \sum_{i \in \mathcal{I}_g} \frac{\left[\widehat{\Delta}_{ig} - (1, \mathbf{z}_i^\top)\boldsymbol{\alpha}\right]^2}{\widehat{V}_{ig}} + \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}_g^{(0)}\right)^\top \boldsymbol{\Lambda} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}_g^{(0)}\right).$$

The confounding-covariate coefficients receive the default ridge penalty. When an intercept anchor is available, the intercept is also shrunk toward its anchor; otherwise its corresponding penalty entry is zero and the intercept is unpenalized. The adjusted residual and local statistic are

$$r_{ig} = \widehat{\Delta}_{ig} - \widehat{\gamma}_g - \mathbf{z}_i^\top \widehat{\boldsymbol{\beta}}_g, \qquad t_{ig} = \frac{r_{ig}}{\sqrt{\widehat{V}_{ig}}}.$$

Let n_g be the number of valid grid locations for gene g, and let d_g be the number of fitted columns in the baseline design. The implementation uses

$$\nu_g = \max(n_g - d_g, 5),$$

where $d_g = 0$ when neither the intercept nor a confounding-covariate basis is included and $d_g = 1$ for an intercept-only baseline. Two-sided plug-in P values are

$$P_{ig} = 2F_{t_{\nu_g}}(-|t_{ig}|).$$

The denominator $\widehat{V}_{ig}$ contains the estimated base local-contrast variance and the specified variance-inflation factors, but it does not include a separate weighted-regression leverage correction or an additional variance term for estimating $\widehat{\gamma}_g$ and $\widehat{\boldsymbol{\beta}}_g$. Accordingly, t_{ig} is interpreted as a working plug-in statistic.

For each gene, local q-values are computed across valid grid locations,

$$q_{ig} = \mathrm{BH}\{P_{ig} : i \in \mathcal{I}_g\}.$$

At target level α, connected components of $\{i : q_{ig} \le \alpha\}$ define local DE regions. Each region is summarized by the median adjusted residual, its direction and its minimum local q-value. Connected regions inherit evidence from the locally adjusted grid calls and do not receive a separate region-level FDR guarantee. Numerical safeguards and null-reference details are described in Supplementary Methods S3.8 and S3.7.

5.2.6 Gene-level, multi-contrast and spatial-trajectory summaries For a single contrast, spatially dependent local P values are combined by the Cauchy combination test (ACAT) [62]. For non-negative weights a_i normalized over valid locations,

$$T_g = \sum_{i \in \mathcal{I}_g} a_i \tan\{(0.5 - P_{ig})\pi\}, \qquad P_g^{\mathrm{ACAT}} = 0.5 - \frac{1}{\pi}\arctan(T_g).$$

For an ordered sample series, such as aging, the gene-level test targets a spatially distributed trend rather than a difference in any single contrast. Let $\widehat{\mu}_{A,ig}^{(c)}$ denote the kernel-smoothed local expression estimate for sample A in contrast c. The unsmoothed adjusted expression used in the age-trend analysis is

$$\widetilde{\mu}_{ig}^{(c)} = \widehat{\mu}_{A,ig}^{(c)} - \mathbf{z}_{ic}^{\top}\widehat{\boldsymbol{\beta}}_{gc},$$

where $\mathbf{z}_{ic}^{\top}\widehat{\boldsymbol{\beta}}_{gc}$ is the fitted grid-varying confounding component derived from local library size and detection rate. The comparison-wide intercept $\widehat{\gamma}_{gc}$ is not removed, so broad expression differences among the ordered sections remain available to the trend analysis. Here, "unsmoothed" means that no trajectory-level temporal smoothing has yet been applied; $\widehat{\mu}_{A,ig}^{(c)}$ is still a spatially kernel-smoothed local expression estimate.

In the aging analysis, c indexes the 19 non-reference sections. The 4.3-month reference is used to construct each age-versus-reference local fit but is not included as an additional observation in the age regression. At every valid grid location i, the adjusted expression estimates are fitted across the ordered sample variable x_c,

$$\widetilde{\mu}_{ig}^{(c)} = \alpha_{ig} + \beta_{ig}x_c + e_{igc},$$

using the mismatch-aware inverse-variance weights from the corresponding local fits. The two-sided local trend P values P_{ig}^{trend} are then combined across valid spatial locations by equally weighted ACAT,

$$T_g^{\mathrm{trend}} = \frac{1}{|\mathcal{I}_g^{\mathrm{trend}}|}\sum_{i \in \mathcal{I}_g^{\mathrm{trend}}} \tan\{(0.5 - P_{ig}^{\mathrm{trend}})\pi\}, \qquad P_g^{\mathrm{trend}} = 0.5 - \frac{1}{\pi}\arctan(T_g^{\mathrm{trend}}).$$

The global null is $\beta_{ig} = 0$ at every tested location. This gene-level trend test uses neither trajectory smoothing nor trajectory clusters and is therefore independent of the selected K. Both the single-contrast and ordered-series ACAT values are raw gene-level P values unless a separate across-gene adjustment is stated.

For visualization of ordered samples, adjusted local-expression trajectories are clustered into spatial domains. When K is not supplied, spAlignDE first evaluates whether cluster-specific time trends improve held-out prediction relative to a shared time trend. If no reliable improvement is detected, it returns the smallest candidate K; otherwise, it retains candidates within one standard error of the best held-out dynamic gain. Spatial resolution is then examined from the largest retained K toward smaller values. The procedure stops when the next coarser candidate would increase the fraction of grid locations in components smaller than one R_{map} footprint and retains the current, finer-side local minimum. If this fragmentation measure decreases throughout the scan and therefore supplies no elbow, the procedure makes one conservative coarsening step from the finest retained candidate rather than continuing to the coarsest candidate. This selection uses neither the gene-level trend test nor the subsequent cluster-level Wald P values.

For cluster $\mathcal{C}_{g\kappa}$ and contrast c, spAlignDE reports a reliability-weighted section-level mean $\bar{\mu}_{g\kappa}^{(c)}$ and weighted within-cluster spatial standard deviation $s_{g\kappa}^{(c)}$. The plotted interval $\bar{\mu}_{g\kappa}^{(c)} \pm 1.96 s_{g\kappa}^{(c)}$ describes spatial dispersion among grid locations and is not a confidence interval for the section-level mean. A spatial-variability-weighted linear trend is fitted across section-level summaries using $\{s_{g\kappa}^{(c)}\}^{-2}$ as a reliability weight, yielding a raw two-sided Wald P value for each selected cluster.

Because the trajectory partition and K are estimated from the same observed trajectories that are subsequently summarized, the cluster-level Wald values are conditional downstream localization summaries rather than tests independent of the clustering step. Selection-aware null calibration would require rerunning trajectory smoothing, clustering, K selection and trend fitting under the null and is not part of the current implementation. Complete clustering and trend-summary details are provided in Supplementary Methods S3.9.

The multiplicity scopes are distinct. For analyses with multiple contrasts, restoring the contrast index gives local q_{igc}-values computed across grid locations within each gene and contrast; for a single contrast these reduce to q_{ig}-values. Cluster-level trajectory trends are reported as raw Wald P values without an additional within-gene BH adjustment. Gene-level ACAT values are also raw P values unless an across-gene adjustment is explicitly applied, and connected regions receive no additional region-level adjustment. In the injured-kidney analysis, valid local tests were equally weighted within the single IL3-versus-NL3 contrast. The resulting q_g^{ACAT}-values were computed across the complete predeclared family of 16,446 genes before construction of the local-ORA gene sets.

5.3 Benchmark and simulation design

5.3.1 Alignment evaluation and robustness

Alignment benchmarking. Alignment performance was evaluated using information not supplied to transformation estimation whenever possible. Following the cross-sample spatial-alignment evaluation framework of Yan et al. [27], shared-gene expression was aggregated into corresponding spatial bins and evaluated using four metrics: Pearson correlation coefficient, cosine similarity, structural similarity index measure and mutual information. Compatible annotations were assessed by nearest-neighbor label transfer, weighted Dice coefficient and weighted intersection over union. For cross-modal tasks, evaluation used task-specific evidence: molecular measurements paired with the reference histology but withheld from alignment; atlas-label coverage, label agreement between transferred atlas labels and RCTD-inferred class labels, and regional marker localization for ST-to-atlas alignment; and annotation agreement, gene-activity/expression concordance and native-neighborhood preservation for spatial ATAC-to-ST alignment. Wall-clock runtime and peak allocated GPU memory were recorded separately. Complete metric definitions are provided in Supplementary Methods S2.10.

Subsampling-based transformation stability. Ten replicate inputs were generated by independently retaining 80% of the cells in both sections and rerunning the complete cross-sample workflow. Each replicate-specific transformation was applied to the same fixed set of pre-aligned query points, and the mean Euclidean displacement from the replicate-mean aligned position was calculated. This quantity measures empirical sensitivity to input subsampling rather than a calibrated confidence interval for the unknown deformation. Full formulas are provided in Supplementary Methods S2.11.

Sensitivity to spatial-structure choices. Cross-sample sensitivity was assessed by varying the Leiden resolution used for joint structure identification, whereas cross-modal sensitivity was assessed by reducing the number of retained ST–atlas structure pairs while holding unrelated settings fixed. Complete designs are provided in Supplementary Methods S2.12.

5.3.2 Simulation data generation The simulation benchmark used a real cell-resolved 3.8-month MERFISH mouse-brain section as the tissue and sampling template. Gene-wise negative-binomial spatial generalized additive models preserved the observed library-size, cell-type and spatial-expression structure when generating conditionally independent count profiles for a baseline reference (simulation sample A) and a signal-bearing sample (simulation sample B). Six non-overlapping circular DE regions, comprising three upregulated and three downregulated regions, were implanted in the signal-bearing sample. Three separate distorted regions were placed outside the DE regions and subjected to smoothly blended local affine deformation; global affine and radial-basis-function deformations and coordinate noise were then added to the signal-bearing sample. The simulation A/B labels refer only to data generation; in post-alignment inference, the signal-bearing sample was the condition of interest and the baseline reference was the control. This data-informed construction retained realistic tissue geometry and spatial sampling while preserving known expression and deformation ground truth [52]. Ten matched simulation groups were generated by varying the DE- and distorted region placements while limiting pairwise overlap among their spatial masks. Further details are provided in Supplementary Methods S3.10.

5.3.3 Simulation evaluation Post-alignment inference was performed at shared-grid locations. To compare the calls with the simulation truth, grid-level q-values, local statistics and significance calls were projected to sample-B cells, where the implanted DE regions and directions were defined. For one simulation group, let D be the number of projected cells called significant at the target FDR of 0.05, F the number of called cells outside the implanted DE regions and T the number of correctly signed called cells within those regions. Let M_1 be the total number of cells in the implanted DE regions. The location-level false discovery proportion and power are

$$\mathrm{FDP} = \frac{F}{\max(D, 1)}, \qquad \mathrm{Power} = \frac{T}{M_1}.$$

These quantities are calculated across sample-B spatial locations rather than across genes. Each simulation group therefore contributes one location-level FDP and one power value to the summaries in Fig. 4 and the Supplementary Figures.

For power at FDP $\leq$ 0.05, valid projected cells were ranked by increasing q-value, with ties ordered by decreasing absolute local statistic. FDP and power were evaluated along this ranking, and the largest power among prefixes satisfying the location-level FDP constraint was reported,

$$\mathrm{Power}_{\mathrm{FDP}\leq 0.05} = \max_{k:\,\mathrm{FDP}(k)\leq 0.05} \frac{T(k)}{M_1}.$$

This truth-dependent thresholding is used only to compare sensitivity at a common realized error level in simulations; it is not an output of the real-data analysis.

5.4 Application-specific analyses

5.4.1 Aging-brain analysis The processed coronal aging-brain MERFISH series comprised 20 sections from 3.4 to 34.5 months. The 4.3-month section was used as the reference because it retained the most complete overall anatomy, and each of the remaining 19 sections was aligned to it. Local expression testing was performed separately for every age-versus-reference contrast. Gene-level age evidence was obtained by fitting mismatch-aware local linear trends across the ordered sections and combining the resulting local trend P values across space by ACAT. Adjusted local-expression trajectories were clustered and summarized as described above.

Cell-type neighborhood analyses were performed on aligned coordinates. For each section, T-cell proximity to oligodendrocytes, microglia and ependymal cells was summarized by the median nearest-neighbor distance, transformed so that a positive age coefficient denotes decreasing distance. Spatial overlap was summarized by the mismatch-risk-weighted correlation between kernel-smoothed T-cell and target-cell density fields. BH adjustment was applied across target cell types separately for proximity and overlap. Local T-cell enrichment maps contrasted normalized T-cell mass with its smoothed expectation under local tissue area. Young and old summaries used the age groups shown in Fig. 5E (young, ≤ 13.6 months, $n = 7$; old, ≥ 24.3 months, $n = 7$). Detailed formulas are provided in Supplementary Methods S3.11.

Counts were normalized to a target total of 250. The shared grid contained 77,056 valid locations with spacing 27.05; the local neighborhood bandwidth was 121.1 and $R_{\mathrm{map}} = 40.57$. The analyses in Fig. 5A–C included the comparison-wide intercept, adjusted for local library size and detection rate, used mismatch-aware variance inflation with density-energy share $s = 0.25$, and left cell-type adjustment off. For each gene, provisional mismatch-inflation coefficients were estimated separately in the 19 age-versus-reference contrasts and robustly combined into one shared gene-specific coefficient. Putatively stable genes were selected using the default screen in Supplementary Methods S3.3. Gene-level aging summaries combined equally weighted local age-trend P values across valid grid locations.

5.4.2 Injured-kidney analysis The injured IL3 Visium section was aligned as the query to the normal NL3 section as the reference. For post-alignment inference, injured IL3 was sample A (case) and normal NL3 was sample B (control), so positive local contrasts indicate higher expression in the injured section. The analysis estimated local expression contrasts on their shared aligned support and summarized gene-level evidence by equally weighted ACAT across valid grid locations. The resulting q_g^{ACAT}-values were computed across the complete predeclared 16,446-gene testing family, and $q_g^{\mathrm{ACAT}} \leq 0.05$ was required for entry into the local-ORA gene sets. For each gene and direction, significant local-grid locations were tested for enrichment in the annotated compartments after accounting for compartment area. A gene satisfying the spatial concentration requirements was assigned to one dominant compartment–direction class, and each class-specific gene set was tested for Gene Ontology Biological Process over-representation against the complete testing family, yielding term q-values within that class. For display, terms within each compartment–direction class were ranked by increasing term q-value. Ties were resolved successively by increasing raw over-representation P value, decreasing overlap gene count and alphabetical GO term name.

For comparison, spots were assigned to the cortex, interface or medulla and injured-versus-normal expression was tested within each compartment. Genes passing compartment-specific BH adjustment were divided by direction and subjected to the same GO over-representation framework. Term recurrence was summarized by the number of compartment–direction classes in which each significant term appeared. Detailed formulas are provided in Supplementary Methods S3.12.

Counts were library-size normalized to a target total of 10,000, and genes detected in at least 10 spots with at least 10 total counts were retained. The shared tissue grid contained 6,187 locations with spacing 75.49; the local neighborhood bandwidth was 191.9 and $R_{\mathrm{map}} = 113.2$. The formal full-gene and displayed-gene fits omitted the comparison-wide intercept. This choice treated broad injury-associated expression shifts as part of the injured-versus-normal contrast rather than as a nuisance baseline. The fits adjusted for local library size and detection rate, enabled mismatch-aware variance inflation with density-energy share $s = 0.75$, disabled cell-type adjustment, and used $\alpha = 0.05$ with connected-region cleanup disabled. Gene-level Aggregated Cauchy Association Test values were adjusted across the complete 16,446-gene family before local DE-region over-representation analysis. The final Gene Ontology analysis used `clusterProfiler` 4.18.4 and `org.Mm.eg.db` 3.22.0. Spot-level cell-type proportions were estimated with RCTD [47] using the mouse kidney single-cell reference from Balzer et al. [54]. The mismatch-aware local model was applied to these proportions on the same aligned grid, using only spots assigned to the same compartment as each grid location. No minimum effective-support threshold was applied to the cell-composition tests. Local P values were jointly BH-adjusted across 14 cell types and valid grid locations; Fig. 5I displays calls at $q < 0.20$. Cell-type-level spatial P values were obtained by equally weighted ACAT across valid grid locations. Reference preparation and compartment-level coverage summaries are described in Supplementary Methods S3.12.

5.5 Software and computational resources

5.5.1 Software and reproducibility spAlignDE analyses were performed using spAlignDE 0.1.0 under Python 3.10.14. The software environment and workflow-specific random seeds are described in Supplementary Methods S1. Comparator

versions, configurations and method-specific input limits are reported in Supplementary Methods S2.9. Runtime and GPU-memory measurements are defined in Supplementary Methods S2.10.1.

5.5.2 Computational resources Computational analyses were performed on a workstation equipped with an AMD Ryzen Threadripper PRO 7985WX 64-Core Processor with 64 physical cores / 128 threads, 503 GB of system memory, and an NVIDIA RTX PRO 6000 Blackwell Max-Q Workstation Edition GPU with 96 GB of GPU memory. GPU-accelerated analyses used NVIDIA driver 570.195.03 with CUDA 12.8 support.

5.5.3 Use of generative AI tools We used OpenAI's ChatGPT and Codex for language polishing and limited programming assistance during the preparation of this work. For manuscript preparation, these tools were used selectively to improve readability, clarity, and flow while preserving the authors' intended scientific meaning. Programming assistance was limited to code debugging, refactoring, and implementation suggestions. The study conception, methodological design, scientific analyses, interpretation of results, and conclusions were developed by the authors. All AI-assisted text was reviewed and revised by the authors, and all AI-assisted code was reviewed, tested, and modified as appropriate. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript and associated code.

6 Data Availability

All datasets used in this study are publicly available. The 10x Genomics Visium mouse kidney sections NL3 (normal control) and IL3 (ischemia–reperfusion injury), together with their region annotations, were obtained from the STcompare Zenodo record (https://zenodo.org/records/20647680). The mouse kidney single-cell reference used for RCTD deconvolution was obtained from GEO accession GSE180420 [54] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE180420). Replicate 1 of the 10x Genomics Fresh Frozen Mouse Brain Replicates Xenium dataset was obtained from 10x Genomics (https://www.10xgenomics.com/datasets/fresh-frozen-mouse-brain-replicates-1-standard). The H&E image and paired Visium spatial gene-expression measurements for control Replicate 1 were obtained from the 10x Genomics Visium CytAssist Gene Expression Libraries of Post-Xenium Mouse Brain (FF) dataset (https://www.10xgenomics.com/datasets/visium-cytassist-gene-expression-libraries-of-post-xenium-mouse-brain-ff-using-the-mouse-whole-transcriptome-probe-set-2-standard). The Xenium In Situ Human Breast Cancer data, comprising In Situ Sample 1 Replicates 1 and 2 from serial sections of the same formalin-fixed, paraffin-embedded tissue block, were obtained from 10x Genomics and were originally described by Janesick et al. [44] (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast). The MERFISH Mouse Brain Receptor Map datasets were obtained from Vizgen (https://info.vizgen.com/mouse-brain-map). The processed coronal mouse brain aging MERFISH dataset was obtained from the Zenodo record "Processed MERFISH Datasets for Brain Aging (Coronal, Sagittal) and Rejuvenation (Exercise, Partial Reprogramming)" (https://doi.org/10.5281/zenodo.13883177). The P22 mouse brain spatial ATAC–RNA-seq data originally reported by Zhang et al. [8] were obtained from the UCSC Cell Browser (https://brain-spatial-omics.cells.ucsc.edu) and the Gene Expression Omnibus under accession GSE205055. The processed data and manual anatomical annotations used in the COSMOS study [51] were obtained from Zenodo (https://doi.org/10.5281/zenodo.13932144). The Allen Mouse Brain Common Coordinate Framework version 3 (CCFv3) annotation volume was obtained from the Allen Institute (https://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/annotation/ccf_2022/). The Allen Institute Mouse Whole Cortex and Hippocampus 10x cell metadata and HDF5 gene-expression matrix used to construct the RCTD reference were obtained from the Allen Brain Map data portal [48, 49] (https://brain-map.org/our-research/cell-types-taxonomies/cell-types-database-rna-seq-data/mouse-whole-cortex-and-hippocampus-10x). The Nissl-stained reference section (Mouse, P56, Coronal; atlas image 100960252, section 269) was obtained from the Allen Mouse Brain Atlas (https://mouse.brain-map.org/static/atlas).

7 Code Availability

The spAlignDE Python package, including the source code, executable notebooks, interactive region-pairing interface, testing suite and environment specifications, is publicly available at https://github.com/dsong-lab/spAlignDE/. Comprehensive documentation and step-by-step tutorials are available at https://dsong-lab.github.io/spAlignDE/. The tutorials describe the input and output formats, model parameters and example workflows for the main functionalities of spAlignDE. The source notebooks for reproducing the analyses are available at https://github.com/dsong-lab/spAlignDE/tree/main/source_notebooks, and the archived source code and materials for reproducing the results are available in the Zenodo repository at https://zenodo.org/records/21983769. All datasets analyzed in this study are publicly available from the repositories listed in the Data Availability section. Large datasets, the Allen CCF annotation volume and pretrained HIPT checkpoints are not redistributed with the software; their sources and required input formats are provided in the corresponding tutorials.

8 Competing interests

The authors declare no competing interests.

9 Acknowledgements

The authors appreciate the comments and feedback from members of the DS Lab at UConn Health (https://dsong-lab.github.io/).

10 Author information

Authors and affiliations

Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington, CT 06030, USA
Su Xu, Lingzhen Meng, Aarya Dalal, Yuxin Yin and Dongyuan Song

Department of Statistics, University of Connecticut, Storrs, CT 06269, USA
Yu Wang

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA
Aarya Dalal

Contributions

D.S., S.X. and Y.W. conceived the study. L.M., A.D. and Y.Y. contributed to the data analysis. S.X., Y.W. and D.S. wrote the manuscript. S.X. and Y.W. contributed equally to this work.

Corresponding author

Correspondence to Dongyuan Song (dosong@uchc.edu).

11 Funding

This work was supported by UConn Health faculty start-up funds (to D.S.).

References

[1] Kok Hao Chen et al. "Spatially resolved, highly multiplexed RNA profiling in single cells". In: *Science* 348.6233 (2015), aaa6090.

[2] Patrik L Ståhl et al. "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics". In: *Science* 353.6294 (2016), pp. 78–82.

[3] Samuel G Rodriques et al. "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution". In: *Science* 363.6434 (2019), pp. 1463–1467.

[4] Vivien Marx. "Method of the Year: spatially resolved transcriptomics". In: *Nature methods* 18.1 (2021), pp. 9–14.

[5] Luyi Tian, Fei Chen, and Evan Z Macosko. "The expanding vistas of spatial transcriptomics". In: *Nature Biotechnology* 41.6 (2023), pp. 773–782.

[6] E. Sun. *Processed MERFISH Datasets for Brain Aging (Coronal, Sagittal) and Rejuvenation (Exercise, Partial Reprogramming)*. 2024. DOI: 10.5281/zenodo.13883177. URL: https://doi.org/10.5281/zenodo.13883177.

[7] Yang Liu et al. "Spatial multi-omics landscape of colorectal cancer macro-and micrometastases". In: *Cancer cell* (2026).

[8] Di Zhang et al. "Spatial epigenome–transcriptome co-profiling of mammalian tissues". In: *Nature* 616.7955 (2023), pp. 113–122.

[9] Charlotte Giesen et al. "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry". In: *Nature methods* 11.4 (2014), pp. 417–422.

[10] Zhiyuan Yuan et al. "SEAM is a spatial single nuclear metabolomics method for dissecting tissue microenvironment". In: *Nature Methods* 18.10 (2021), pp. 1223–1232.

[11] Hong Wei Dong. *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse*. John Wiley & Sons Inc, 2008.

[12] Quanxin Wang et al. "The Allen mouse brain common coordinate framework: a 3D reference atlas". In: *Cell* 181.4 (2020), 936–953.e20.

[13] Ron Zeira et al. "Alignment and integration of spatial transcriptomics data". In: *Nature Methods* 19.5 (2022), pp. 567–575.

[14] Xiang Zhou, Kangning Dong, and Shihua Zhang. "Integrating spatial transcriptomics data across different conditions, technologies and developmental stages". In: *Nature Computational Science* 3.10 (2023), pp. 894–906.

[15] Andrew Jones et al. "Alignment of spatial genomics data using deep Gaussian processes". In: *Nature methods* 20.9 (2023), pp. 1379–1387.

[16] Chen-Rui Xia et al. "Spatial-linked alignment tool (SLAT) for aligning heterogenous slices". In: *Nature Communications* 14.1 (2023), p. 7236.

[17] Barbara Zitova and Jan Flusser. "Image registration methods: a survey". In: *Image and vision computing* 21.11 (2003), pp. 977–1000.

[18] Kalen Clifton et al. "STalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping". In: *Nature communications* 14.1 (2023), p. 8123.

[19] Kaitlin M Stouffer et al. "Cross-modality mapping using image varifolds to align tissue-scale atlases to molecular-scale measures with application to 2D brain sections". In: *Nature communications* 15.1 (2024), p. 3530.

[20] Yecheng Tan et al. "Integrative cross-sample alignment and spatially differential gene analysis for spatial transcriptomics". In: *Nature Communications* 17.1 (2026), p. 5577.

[21] Xinhao Liu, Ron Zeira, and Benjamin J Raphael. "Partial alignment of multislice spatially resolved transcriptomics data". In: *Genome Research* 33.7 (2023), pp. 1124–1132.

[22] Bingjie Dai et al. "3d-OT: a deep geometry-aware framework for heterogeneous slices alignment of spatial multi-omics". In: *Nature Methods* (2026), pp. 760–771.

[23] Xiaojie Qiu et al. "Spatiotemporal modeling of molecular holograms". In: *Cell* 187.26 (2024), 7351–7373.e61.

[24] Zefang Tang et al. "Search and match across spatial omics samples at single-cell resolution". In: *Nature methods* 21.10 (2024), pp. 1818–1829.

[25] Haoyang Li et al. "SANTO: a coarse-to-fine alignment and stitching method for spatial omics". In: *Nature Communications* 15.1 (2024), p. 6048.

[26] Yi Wang, Zihang He, and Yunjie Yan. "SPOmiAlign: a modality-agnostic computational framework for multimodal spatial omics alignment enabled by a feature matching foundation model". In: *Briefings in Bioinformatics* 27.3 (2026), bbag331.

[27] Yunzhi Yan et al. "Benchmarking alignment methods for spatial transcriptomics data". In: *Nature Computational Science* (2026), pp. 524–541.

[28] Kaitlin M Stouffer et al. "The xIV-LDDMM toolkit of image-varifold based technologies for mapping 3D images and spatial-omics across scales". In: *Communications biology* 8.1 (2025), p. 1401.

[29] Yakir A Reshef et al. "Accurate and well-powered case–control analysis of spatial molecular data". In: *Nature Methods* (2026), pp. 1–11.

[30] Peiying Cai, Mark D. Robinson, and Simone Tiberi. "DESpace2: detection of differential spatial patterns in spatial omics data". In: *bioRxiv* (2025). DOI: 10.1101/2025.06.30.662268.

[31] Kalen Clifton et al. "STcompare: comparative spatial transcriptomics data analysis of structurally matched tissues to characterize differentially spatially patterned genes". In: *Bioinformatics* 42.9 (Sept. 2026), btag644. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btag644. eprint: https://academic.oup.com/bioinformatics/article-pdf/42/9/btag644/70806525/btag644.pdf. URL: https://doi.org/10.1093/bioinformatics/btag644.

[32] François-Xavier Vialard et al. "Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation". In: *International Journal of Computer Vision* 97.2 (2012), pp. 229–241.

[33] Vipul Singhal et al. "BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis". In: *Nature genetics* 56.3 (2024), pp. 431–441.

[34] Ilya Korsunsky et al. "Fast, sensitive and accurate integration of single-cell data with Harmony". In: *Nature methods* 16.12 (2019), pp. 1289–1296.

[35] M Faisal Beg et al. "Computing large deformation metric mappings via geodesic flows of diffeomorphisms". In: *International journal of computer vision* 61.2 (2005), pp. 139–157.

[36] John Ashburner and Karl J Friston. "Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation". In: *Neuroimage* 55.3 (2011), pp. 954–967.

[37] Daniel Tward et al. "Diffeomorphic registration with intensity transformation and missing data: Application to 3D digital pathology of Alzheimer's disease". In: *Frontiers in neuroscience* 14 (2020), p. 52.

[38] Yingxin Lin et al. "scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets". In: *Proceedings of the national academy of sciences* 116.20 (2019), pp. 9775–9784.

[39] Yuanyuan Yu and Zhi Xie. "Spatial transcriptomic alignment, integration, and 3D reconstruction by STAIR". In: *Genome Biology* 26.1 (2025), p. 427.

[40] Vizgen. *MERFISH Mouse Brain Receptor Map*. Vizgen Data Release V1.0, May 2021. 2021. URL: https://info.vizgen.com/mouse-brain-map.

[41] Sepideh Gharaie et al. "Single cell and spatial transcriptomics analysis of kidney double negative T lymphocytes in normal and ischemic mouse kidneys". In: *Scientific reports* 13.1 (2023), p. 20888.

[42] Kalen Clifton, Jean Fan, and Hamid Rabb. *Single-Cell and Spatial Transcriptomics Analysis of Kidney Double-Negative T Lymphocytes in Normal and Ischemic Mouse Kidneys*. Spatial transcriptomics dataset. Nov. 2025. URL: https://doi.org/10.5281/zenodo.17676992.

[43] Eric D Sun et al. "Spatial transcriptomic clocks reveal cell proximity effects in brain ageing". In: *Nature* 638.8049 (2025), pp. 160–171.

[44] Amanda Janesick et al. "High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis". In: *Nature communications* 14.1 (2023), p. 8353.

[45] 10x Genomics. *Xenium In Situ Dataset: Mouse Brain Neurobiology*. Fresh-frozen mouse brain coronal section profiled using the Xenium Mouse Brain Gene Expression panel. 2023. URL: https://www.10xgenomics.com/products/xenium-in-situ/mouse-brain-dataset-explorer.

[46] 10x Genomics. *Visium CytAssist Gene Expression Libraries of Post-Xenium Mouse Brain (FF)*. Spatial Gene Expression dataset analyzed using Space Ranger 2.1.0. 2023. URL: https://www.10xgenomics.com/datasets/visium-cytassist-gene-expression-libraries-of-post-xenium-mouse-brain-ff-using-the-mouse-whole-transcriptome-probe-set-2-standard.

[47] Dylan M Cable et al. "Robust decomposition of cell type mixtures in spatial transcriptomics". In: *Nature biotechnology* 40.4 (2022), pp. 517–526.

[48] Zizhen Yao et al. "A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation". In: *Cell* 184.12 (2021), 3222–3241.e26.

[49] Allen Institute for Brain Science. *Mouse Whole Cortex and Hippocampus 10x*. Cell metadata and cell-by-gene HDF5 expression matrix; accessed 16 July 2026. 2020. URL: https://brain-map.org/our-research/cell-types-taxonomies/cell-types-database-rna-seq-data/mouse-whole-cortex-and-hippocampus-10x.

[50] Ed S Lein et al. "Genome-wide atlas of gene expression in the adult mouse brain." In: *Nature* 445.7124 (2007), pp. 168–176.

[51] Yuansheng Zhou et al. "Cooperative integration of spatially resolved multi-omics data with COSMOS". In: *Nature communications* 16.1 (2025), p. 27.

[52] Dongyuan Song et al. "scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics". In: *Nature Biotechnology* 42.2 (2024), pp. 247–252.

[53] Patrick M Loerch et al. "Evolution of the aging brain transcriptome and synaptic regulation". In: *PloS one* 3.10 (2008), e3329.

[54] Michael S Balzer et al. "Single-cell analysis highlights differences in druggable pathways underlying adaptive or fibrotic kidney regeneration". In: *Nature Communications* 13.1 (2022), p. 4018.

[55] Arthur Lau et al. "Dipeptidase-1 governs renal inflammation during ischemia reperfusion injury". In: *Science advances* 8.5 (2022), eabm0142.

[56] Jordan W Squair et al. "Confronting false discoveries in single-cell differential expression". In: *Nature communications* 12.1 (2021), p. 5692.

[57] Kip D Zimmerman, Mark A Espeland, and Carl D Langefeld. "A practical solution to pseudoreplication bias in single-cell studies". In: *Nature communications* 12.1 (2021), p. 738.
[58] Richard J Chen et al. "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 16144–16155.
[59] Richard J Chen et al. "Towards a general-purpose foundation model for computational pathology". In: *Nature medicine* 30.3 (2024), pp. 850–862.
[60] Nikos Paragios, Mikael Rousson, and Visvanathan Ramesh. "Non-rigid registration using distance functions". In: *Computer Vision and Image Understanding* 89.2-3 (2003), pp. 142–165.
[61] L. Kish. *Survey Sampling*. Wiley, 1965.
[62] Yaowu Liu and Jun Xie. "Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures". In: *Journal of the American Statistical Association* 115.529 (2020), pp. 393–402.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- spAlignDEsupp.pdf
- spAlignDEsupp.pdf