GALA: A Unified Landmark-Free Framework for Coarse-to-Fine Spatial Alignment Across Resolutions and Modalities in Spatial Transcriptomics

Tao Ding[1] and PengCheng Zeng[1,*]

[1]Institute of Mathematical Sciences, ShanghaiTech University, 201210, Shanghai, China
[*]Corresponding author. zengpch@shanghaitech.edu.cn

Abstract

Spatial transcriptomics alignment is challenged by technical variations, including geometric distortions from tissue preparation and platform-driven differences in resolution and modality. These issues create diverse alignment scenarios, from matched and mismatched resolutions to cross-modality integration, while partial tissue coverage further complicates the task. To overcome these limitations, we introduce GALA (Genetic Algorithm–guided Large Deformation Alignment), a unified, landmark-free framework that couples global affine transformation and local diffeomorphic deformation within a single optimisation. Its modality-aware rasterisation harmonises transcriptomic and histological data into a shared grid, enabling landmark-free, multimodal alignment across resolutions and modalities. Evaluated on diverse human and mouse datasets, GALA outperforms existing methods in accuracy, computational efficiency, and biological interpretability for both complete and partial tissue alignment.

Keywords: Spatial Transcriptomics Alignment, Multimodal Integration, Image Registration, Diffeomorphic Deformation, Genetic Algorithm

1 Introduction

Spatial transcriptomics (ST) technologies enable high-throughput mapping of gene expression within intact tissue sections, providing detailed insights into cellular organisation, tissue architecture and disease-associated microenvironments[1,2,3]. Contemporary platforms, including 10x Visium[4], Xenium[5] and MERFISH[6], span a wide range of spatial resolutions, from multi-cellular spots (spot-level) to single-cell or subcellular precision (cell-level), and frequently integrate transcriptomic measurements with histological images. While this diversity expands analytical possibilities, it also introduces substantial variability across tissues, samples and platforms, making direct spatial comparison and integration challenging[7,8,9]. Geometric distortions arising during tissue preparation, such as rotation, shifting, stretching and local warping, further complicate alignment, even between adjacent sections from the same specimen[10,11,12]. Correcting these transformations is essential to preserve tissue morphology and establish accurate spatial correspondence, enabling downstream applications such as cross-sample integration, three-dimensional reconstruction, deconvolution and spatial domain annotation[8,13,14,15].

Beyond geometric distortion, ST datasets often differ in spatial resolution, gene coverage and molecular sensitivity across platforms[3,16]. These discrepancies complicate the integration of measurements ranging from single-cell transcriptomes to multi-cellular spot profiles and spanning modalities such as histology and transcriptomics. In practice, ST alignment encompasses three common scenarios: matched-resolution (spot-to-spot or cell-to-cell), mismatched-resolution (cell-to-spot) and cross-modality (transcriptomics-to-histology). Each scenario imposes unique methodological constraints, influenced by factors such as spatial resolution, tissue integrity, and the availability of complementary histological context.

Spot-to-spot alignment is typically performed using rigid or affine transformations, and in some cases smooth non-rigid deformations informed by spatial features and gene expression[7,8], suitable when each spot aggregates multiple cells and local variation is smoothed. Representative methods include PASTE[13], STAligner[17] and GPSA[18], which perform well under matched resolution and full-tissue overlap but struggle with

partial coverage, cross-platform differences, or large-scale single-cell data. Cell-to-cell alignment requires non-rigid methods to capture fine anatomical variability[3,12]; methods such as STalign[12] estimate diffeomorphic deformation fields but need landmark-based pre-alignment and are sensitive to large displacements or missing tissue. Coarse-to-fine frameworks like ST-GEARS[19] combine global affine transformation and local refinement, improving performance in complex misalignments but mainly for matched-resolution data. Mismatched-resolution alignment, e.g., mapping single-cell profiles to spot-level data, enables cell-type annotation and spatial reconstruction[20,21], though methods like SLAT[22] rely on sparse correspondences and have limited scalability. Tissue morphology, e.g., H&E images, can aid alignment[23]; GPSA[18] and PASTE2[24] incorporate image features but remain spot-level. Cross-modality alignment, such as transcriptomics-to-histology[12,25], is challenging due to unshared features and differing distributions, with existing approaches often requiring manual landmarks (e.g., STalign). Incomplete tissue coverage further complicates alignment[26], as most methods assume full overlap, with only limited probabilistic handling in PASTE2[24]. Collectively, these limitations highlight the need for a unified, scalable alignment approach capable of handling diverse platforms, spatial resolutions, and modalities, while remaining robust to missing or partially overlapping tissue.

To address these challenges, we introduce *GALA* (**G**enetic **A**lgorithm–guided **L**arge Deformation **A**lignment), a general and fully automated framework for multimodal spatial alignment. GALA combines a modality-aware rasterisation module that projects heterogeneous transcriptomic and histological inputs onto a shared spatial grid with a coupled optimisation strategy that jointly estimates global affine transformation and local diffeomorphic deformation under a single objective function. This design enables accurate alignment across matched- and mismatched-resolution settings (spot-to-spot, cell-to-cell and cell-to-spot) as well as transcriptomics-to-histology alignment, while naturally supporting partial tissue overlap without manual landmarks. Across human and mouse datasets from Visium, MERFISH and Xenium, GALA delivers accurate, efficient and biologically coherent alignments, facilitating downstream comparative analyses and multi-sample atlas construction.

2 Results

2.1 Overview of the GALA framework

GALA is a unified, landmark-free framework designed to accurately align ST data across resolutions and modalities by jointly correcting large-scale distortions and fine-scale local variations. As illustrated in Fig. 1a, the workflow comprises four components: Input, Multimodal rasterisation, Coupled alignment process, and Output. The input contains source (to be aligned) and target datasets, each potentially multimodal comprising spatial coordinates, gene expression, and histological images (assumed to be available). During rasterisation, heterogeneous transcriptomic measurements from informative genes and histological images are converted into structured, image-like grids, providing a common representation for efficient joint processing (see Section 4.1). The coupled alignment process consists of three elements: a global affine operator A that captures large-scale linear transformations while identifying the overlapping region between source and target sections; a local diffeomorphic mapping ϕ^v that models smooth nonlinear deformations; and a probabilistic matching term P_M that distinguishes reliably matched areas within the overlapping region. The output stage produces the aligned source data, and for illustration, includes alignment results for a representative gene.

Fig. 1b provides a detailed schematic of the unified objective function and optimisation scheme. The objective includes two fidelity terms for transcriptomics and histology (weighted by α and $1-\alpha$, respectively), together with a regularisation penalty on the deformation field v. Alignment proceeds in two stages. During the coarse stage, the affine matrix A is estimated using a genetic algorithm (GA)[27] to correct global differences such as rotation, translation, scaling, and reflection, and to identify the overlap region for subsequent refinement. During the fine stage, local deformations and correspondence probabilities are refined through a joint update of ϕ^v and P_M using an expectation–maximization (EM)-like large deformation diffeomorphic metric mapping (LDDMM) procedure[12,28,29]. This step captures smooth, invertible deformations associated with tissue distortions while adaptively weighting areas according to their correspondence likelihood. The coupled optimisation framework allows coarse alignment to guide local refinement, while local updates iteratively improve global consistency. Convergence behaviour is illustrated in the rightmost of Fig. 1b over four global–local episodes, where the first subplot shows both the convergence of fidelity and regularisation terms, and the second one only focuses the fidelity to better visualise its convergence trend. A full mathematical formulation and optimisation procedure are provided in Section 4.2 and Supplementary Notes 1.2–1.3.

Compared to existing methods, GALA offers several advantages. Methodologically, it couples global affine

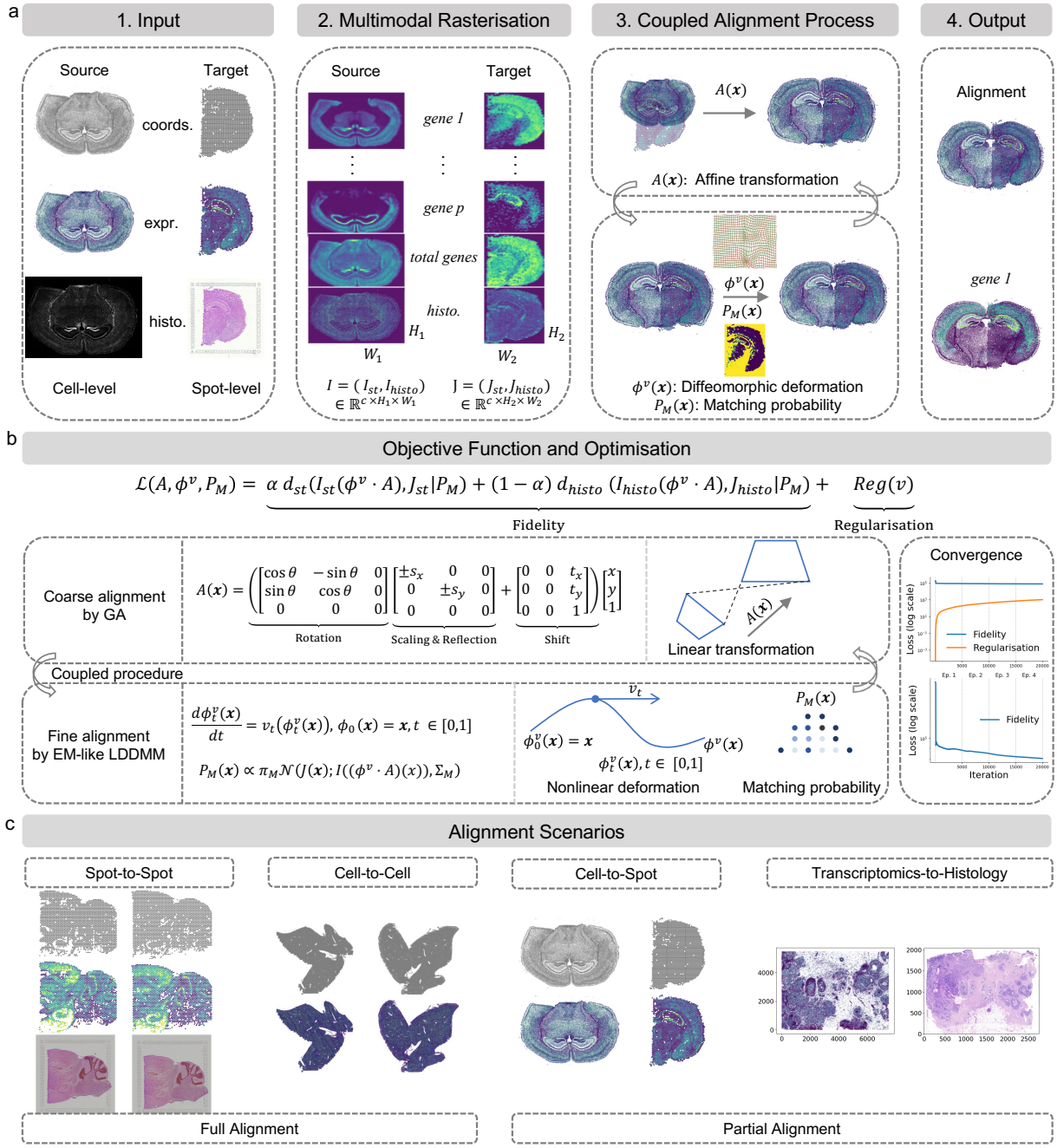

Figure 1: **Overview of the GALA framework and its application scenarios.** **a** Workflow of GALA. The multimodal source (to be aligned) and target datasets, comprising spatial coordinates ('coords.'), gene expression ('expr.'), and histological images ('histo.'), potentially at different resolutions, are rasterised into multi-channel tensors, $I = (I_{\text{st}}, I_{\text{histo}})$ and $J = (J_{\text{st}}, J_{\text{histo}})$, by concatenating the transcriptomic and histology channels. The coupled alignment jointly estimates a global affine transformation A, a local diffeomorphic mapping ϕ^v, and a probabilistic matching field P_M, yielding the final aligned output. An example gene alignment is shown. **b** Schematic of the unified objective function and optimisation. The objective combines transcriptomic and histological fidelity terms (weighted by α and $1-\alpha$) with a regularisation on v. Fidelity is computed as a squared error between the aligned source, $I((\phi^v \cdot A)(\boldsymbol{x})) = I(\phi^v(A(\boldsymbol{x})))$, modulated by the matching probability $P_M(\boldsymbol{x})$, and the target $J(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^2$ denotes a location in the target coordinate system. The global stage estimates A via a GA, while the local stage jointly refines ϕ^v and P_M through an EM-like LDDMM update, guided by a Gaussian posterior with prior π_M and covariance Σ_M. Convergence of fidelity and regularisation across global–local iterations is shown. **c** GALA supports a wide range of alignment tasks, including matched-resolution alignment (spot-to-spot and cell-to-cell), mismatched-resolution alignment (cell-to-spot or spot-to-cell), cross-modality alignment (e.g., transcriptomics-to-histology), and partial alignment with incomplete tissue coverage.

and local diffeomorphic transformations within a single framework using a coordinate descent strategy, enabling mutually informed updates. Practically, it flexibly incorporates histology and performs robustly even with incomplete tissue coverage, while remaining fully automatic and landmark-free. In application, it supports a wide range of scenarios (Fig. 1c): i) matched-resolution (spot-to-spot or cell-to-cell), ii) mismatched-resolution (cell-to-spot), iii) cross-modality (transcriptomics-to-histology), and iv) partial overlap. To our knowledge, no existing method can accommodate all these scenarios effectively.

We next evaluate GALA across diverse ST alignment tasks (Fig. 1c), comparing it to state-of-the-art baselines using established metrics (Section 4.4). Baseline methods vary across scenarios, as most are only applicable to one or two scenarios.

2.2 Spot-to-spot alignment within ST technologies

2.2.1 Human dorsolateral pre-frontal cortex (DLPFC) data by Visium

The first example used to evaluate GALA is the human dorsolateral pre-frontal cortex (DLPFC) ST data generated with the 10x Genomics Visium platform[26]. The dataset comprises three independent samples (A, B, and C), each containing four consecutive tissue slices annotated into seven anatomical regions: six neocortical layers and white matter (WM). Each slice contains 3,000–4,000 spots and approximately 10,000 genes. The spacing between adjacent slices is 10 μm for pairs 1–2 and 3–4, but increases to 300 μm between slices 2 and 3, introducing both gradual and abrupt structural variation (see Sample C in Fig. 2a). We focus on Sample C due to its complete annotation, clear laminar structure, and widespread use in recent comparative studies[13,17,30,31]. GALA was compared against five baseline methods applicable to spot-to-spot alignment, spanning rigid (PASTE, STAligner), non-rigid (GPSA), one-to-one matching (SLAT), and coarse-to-fine (ST-GEARS) strategies. Evaluation metrics include label consistency accuracy, Adjusted Rand Index (ARI), and Normalized Mutual Information (NMI) to quantitatively assess biological coherence after alignment (see Section 4.4).

We first considered three pairwise slice alignments ($2 \rightarrow 1$, $3 \rightarrow 2$, and $4 \rightarrow 3$) for Sample C without histological images. As shown in Fig. 2b, GALA achieves the highest mean label consistency accuracy (0.868) across these comparisons, surpassing top-performing methods such as PASTE (0.839) and GPSA (0.833). Clustering consistency metrics further supported GALA's advantage, with joint Gaussian mixture module (GMM) analysis yielding a mean ARI of 0.396 and NMI of 0.508, outperforming STAligner (0.385 and 0.491, respectively; see Supplementary Tab. S3 and Fig. S6a for detailed metric values). These results demonstrate that GALA preserves label consistency and biological coherence more effectively than all baselines. Visual comparison of aligning slice 3 to slice 2 (Fig. 2c) shows that GALA achieves the most precise alignment of all major cortical layers, whereas STAligner exhibits global structural mismatches and GPSA produces substantial distortions in WM (pink). Visual comparisons for slice pairs $2 \rightarrow 1$ and $4 \rightarrow 3$ are provided in Supplementary Fig. S6b-c.

We further assessed biological consistency by examining canonical layer-specific marker genes. As shown in Fig. 2d, stacked source and target slices display markedly improved spatial correspondence after GALA alignment. Spots with high marker-gene expression are shown in blue on the target slice and are coloured by their annotated layer on the source and aligned source slices; low-expression spots are shown in grey in all panels. Representative markers (*HPCAL1*, *PCP4*, *MOBP*) exhibit increased overlap after alignment, indicating enhanced layer preservation and improved biological coherence. Using pairwise alignment, 3D reconstruction was performed by setting slice 1 as the template and aligning other slices to a common coordinate system. Fig. 2e shows smoother transitions from WM to Layer 1 and better-preserved boundaries in GALA compared to other baselines.

Next, we evaluated the effect of incorporating H&E-stained histology images on alignment between slices 3 and 2. GPSA was included as the only baseline supporting image integration. As shown in Fig. 2f (bar chart), GALA achieved label accuracies of 0.869 and 0.875 with and without histology, respectively, outperforming GPSA (0.833 and 0.830). Visual inspection confirms that GALA produces more accurate and smoother alignments, indicating that inclusion of morphological information benefits GALA while introducing greater variability in GPSA's Gaussian process. As discussed in Section 4.3, the fusion weight α is set to 0.5 by default when histology is available. Additional experiments (line chart in Fig. 2f) show GALA remains stable for α near 0.5, demonstrating robustness to the relative weight.

Third, we explored partial alignment by cropping slice 3 to retain either the left or bottom regions across overlap ratios of 50–90%, incorporating histology in both source and target. GALA was compared with PASTE2, a recent method designed for partial ST alignment. Fig. 2g shows label consistency accuracy for raw

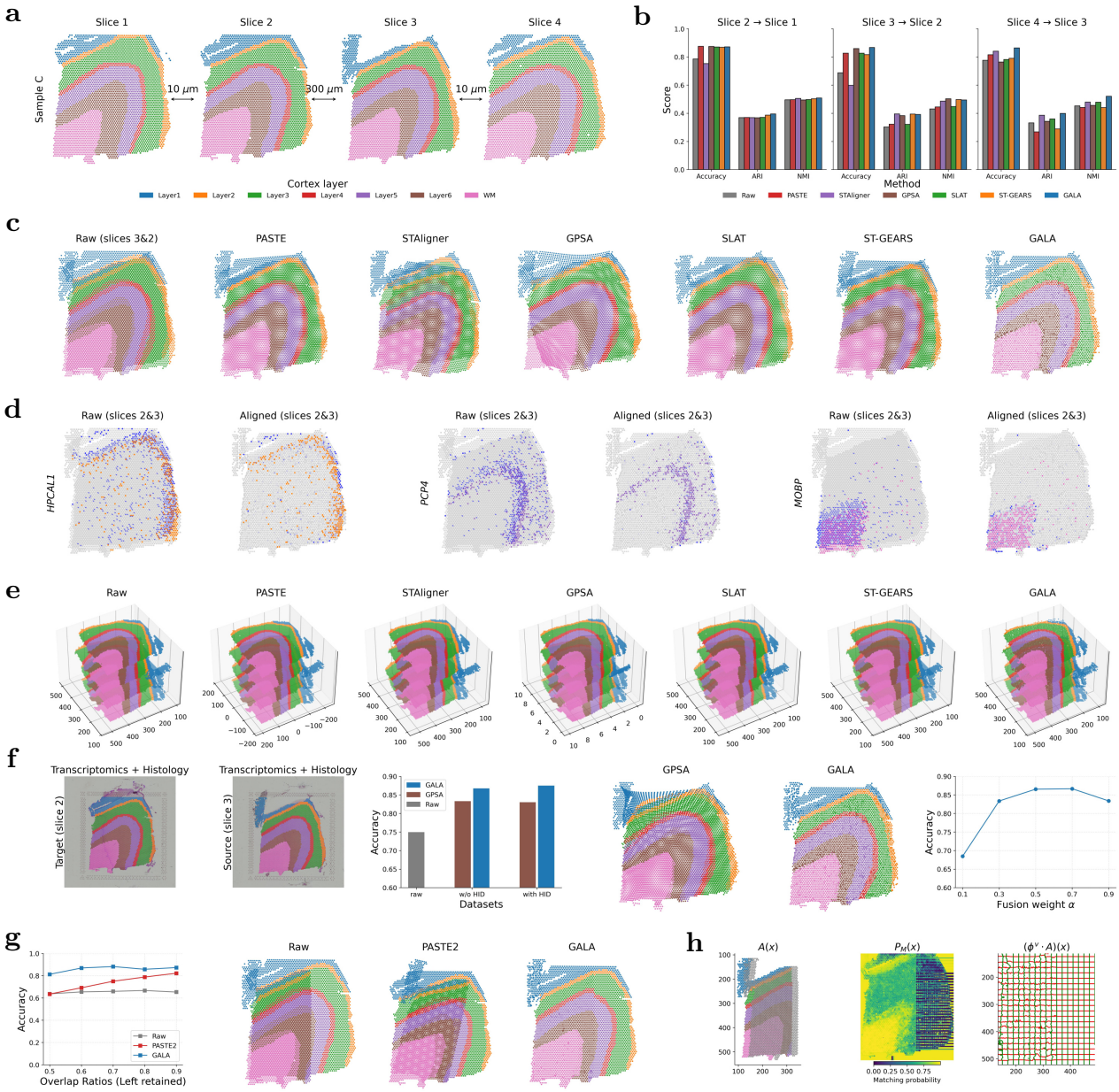

Figure 2: **Evaluation of full/partial alignment on Sample C of DLPFC data across GALA and baselines.** **a** Four raw tissue slices from Sample C, with adjacent distances of 10 μm, 300 μm, and 10 μm. Spots are annotated with six neocortical layers (Layer 1–6) and white matter (WM). **b** Quantitative comparison of alignment performance for three pairwise slice alignments ($2 \to 1, 3 \to 2, 4 \to 3$) based on label consistency accuracy (Accuracy), ARI, and NMI. 'Raw' indicates unaligned slices. **c** Visual comparison of aligning slice 3 to slice 2 across all methods (except 'Raw'). **d** Visual comparison for spatial distribution of the three layer marker genes before and after alignment by GALA. Spots with high marker-gene expression in the target slice are blue; low-expression spots are grey. In the source and aligned source slices, high-expression spots are coloured according to their annotated layer in (a), while low-expression spots are grey. **e** 3D reconstruction of all four slices for GALA and baselines. **f** Comparison of alignment between slices 3 and 2 when histological image data is included, across GPSA and GALA. The rightmost panel shows sensitivity to the relative weight α. 'w/o HID' indicates no histology. **g** Label consistency accuracy for partial alignment (left-retained slice 3) across overlap ratios 50%–90%; right panels show the visual comparison of the corresponding partial alignment in the case of overlap ratio 60%. **h** Illustration of the GALA mechanism for 'left 60% retained': global affine transformation $A(x)$, matching probabilities $P_M(\boldsymbol{x})$, and local diffeomorphic mapping $(\phi^v \cdot A)(\boldsymbol{x})$.

data, PASTE2, and GALA when the left section is retained; results for the bottom section are in Supplementary Fig. S7a. Across all overlap ratios in both partial settings, GALA consistently achieved higher and more stable accuracy than PASTE2. Visual results for 60% overlap further demonstrate that only GALA preserves laminar continuity with minimal distortion. Sensitivity analysis of α (Supplementary Fig. S7c) confirms robustness consistent with Fig. 2f.

Finally, we illustrated why GALA performs well in partial alignment by visualising learned parameters for a 60% overlap (first type of partial overlap; second type shown in Supplementary Fig. S7b). The left panel of Fig. 2h shows the global affine transformation of the cropped source from the original location (grey) to the transformed location (coloured) in the target coordinate system. The middle panel indicates the learned matching probability P_M for each target spot, with higher values (brighter) highlighting the reliably matched areas. The right panel shows the local refinement via the learned diffeomorphic mapping ($\phi^v \cdot A$), producing smooth, topology-preserving deformations. Only the lattice in the overlapped part, corresponding to high P_M values, is distorted. The global affine transformation and local diffeomorphic mapping are iteratively optimised under the objective function, explaining GALA's effectiveness for partial ST alignment.

2.2.2 Mouse brain sagittal-posterior (MBSP) sections by Visium

The second example of spot-to-spot alignment involves mouse brain sagittal-posterior (MBSP) sections produced by 10x Genomics Visium. The two sections contain 2,805 and 2,840 spots, respectively, each profiled for 32,285 genes, and are accompanied by H&E-stained histological images (left panel of Fig. 3a).

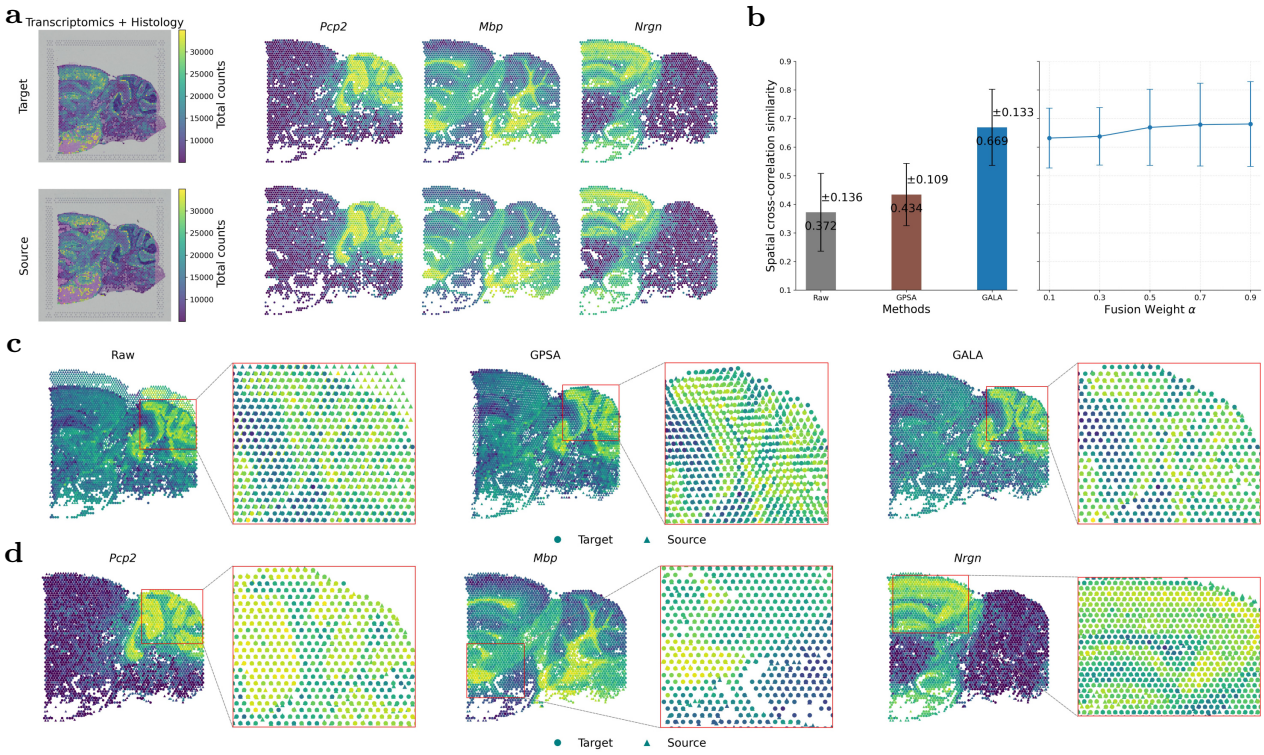

Figure 3: **Alignment of two MBSP sections.** **a** Raw slices overlaid on H&E images and spatial expression of three top informative genes (*Pcp2*, *Mbp*, *Nrgn*). Brighter spots indicate higher gene expression. **b** Left: mean spatial cross-similarity scores ($\pm$ s.d.) across the top 135 informative genes for raw slices (Raw) and slices aligned by GPSA and GALA ($\alpha = 0.5$). Right: cross-similarity $\pm$ s.d. across fusion weights from 0.1 to 0.9. **c** Visual comparison of alignment: left, raw stacked slices; middle and right, slices aligned by GPSA and GALA. Zoomed regions correspond to red boxes. **d** Visual alignment of the three informative genes by GALA, with zoomed-in regions for each panel.

We first examined the role of spatially informative genes in GALA, defined as genes ranked by the R^2-based strategy from GPSA[18], which orders genes according to how well their expression can be predicted

from local spatial context (see Section 4.3). The top three genes automatically selected were *Pcp2*, *Mbp*, and *Nrgn*. These genes exhibit complementary anatomical specificity (Fig. 3a, right panels): *Pcp2* is restricted to cerebellar Purkinje cells and selected cortical layers; *Mbp* marks oligodendrocyte-enriched white matter tracts; *Nrgn* is expressed in excitatory neurons within cortical and hippocampal layers. Their distinct localisation patterns provide strong signals for guiding alignment, effectively serving as 'landmarks' in our landmark-free GALA framework.

Unlike the previous example, ground-truth spot annotations of MBSP sections are unavailable. Therefore, we used spatial cross-correlation similarity (instead of label consistency accuracy, ARI, or NMI) to quantify alignment quality (Section 4.4). Using GPSA's protocol, we selected the top 135 informative genes with $R^2 > 0.6$, each reporting one correlation score after alignment. We compared GALA with GPSA only, as no other baselines support histological image data. The bar charts in Fig. 3b show that: (i) both methods substantially improve alignment relative to raw slices (mean 0.372 ± 0.136); (ii) GALA achieves a higher mean score of 0.669 ± 0.133, an improvement of over 23% relative to GPSA (0.434 ± 0.109). Visual comparison of stacked target and aligned source slices (Fig. 3c) highlights that anatomical features are highly displaced in raw slices. GPSA corrects misalignment globally but shows local distortions, whereas GALA provides finer alignment with better preservation of structural boundaries.

We next examined the alignment of the top three informative genes (Fig.3d). Compared with their raw, unaligned spatial patterns (see Supplementary Fig. S8), all three genes display markedly improved localisation following GALA alignment. In the cerebellar cortex, *Pcp2* becomes more sharply confined to Purkinje cell layers, with clearer lobular boundaries and reduced cross-layer bleeding. *Mbp*, which marks white-matter tracts, shows more continuous and anatomically coherent fibre bundles, while *Nrgn* exhibits a smoother and more spatially consistent distribution across granule and cortical layers. These improvements, further highlighted in the magnified insets (red boxes), demonstrate GALA's ability to recover biologically coherent expression domains by correcting structural variability across slices. Sensitivity analysis of the fusion weight α (default 0.5) further confirmed GALA's robustness, as shown in Fig.3b.

2.3 Cell-to-cell alignment within ST technologies

In the third example, we applied GALA to subcellular-resolved ST data from the MERFISH mouse liver map datasets (Vizgen) without histological image data. The two liver slices, L2S1 and L2S2 from Animal 2, each comprises approximately 300,000 cells with expression profiles for 347 genes, serving as target and source data, respectively. Fig. 4a shows the raw slices, displaying only the positions of cells with detected gene expression (the corresponding total gene expression maps are provided in Supplementary Fig. S9a), highlighting substantial rotation. Baseline methods include SLAT, STalign, and the affine component of GALA (denoted as Affine).

To quantify alignment quality, cells were aggregated into uniform 30 μm pseudo-spots to mitigate local sampling density variation, following STalign's protocol. Cosine similarity of expression was computed between matched pseudo-spots, capturing local correspondence in both gene expression magnitude and spatial structure. Evaluation considered 134 known liver marker genes[32] and top-ranked informative genes. Five representative genes, including canonical markers (*Axin2*[33], *Vwf*[34]) and top informative genes (*Aldh1b1*, *Hsd17b6*, *Cpy12a*), were analysed in detail.

First, we quantitatively assessed alignment across all baselines and GALA. Fig. 4b shows the distribution of cosine similarity scores for 134 liver marker genes against the total gene expression at pseudo-spots after alignment by these methods (including Raw data). The median similarity score and the scores for the five representative genes are indicated by red dashed line and black dashed lines, respectively. SLAT yielded negligible improvement: its one-to-one matching is sensitive to variations in cell density and can introduce spurious correspondences. STalign delivered only marginal gains because it depends on manually specified priors to guide deformation, which limits performance when prior correspondence is imprecise. The affine-only variant of GALA (Affine) modestly increased the median similarity to 0.41, and the full GALA framework further raised it to 0.43, demonstrating that both global affine and local diffeomorphic components contribute to alignment. Representative gene trends exemplify this: raw alignment produced scores of 0.43 (*Axin2*), 0.14 (*Vwf*), 0.33 (*Aldh1b1*), 0.43 (*Hsd17b6*), and 0.72 (*Cyp12a*); Affine improved these to 0.67, 0.37, 0.77, 0.78, and 0.86; full GALA further increased them to 0.71, 0.49, 0.85, 0.84, and 0.87.

Second, we visually compared alignment results for Affine and GALA. The right panels of Fig. 4c show that both methods achieve global alignment, but GALA produces better local alignment, with the top region (red box) showing more coherent overlap and reduced spatial drift. Fig. 4d illustrates alignment for three

representative genes, including zoomed-in regions, showing that GALA more accurately aligns source and target cells for canonical markers *Axin2*, *Vwf* and informative gene *Aldh1b1* (Supplementary Fig. S9c illustrate the corresponding gene expression values).

These results demonstrate that GALA's coupled coarse-to-fine optimisation improves global alignment accuracy while preserving local gene-level patterns in highly resolved single-cell spatial transcriptomic datasets, even without H&E guidance or manual landmarks.

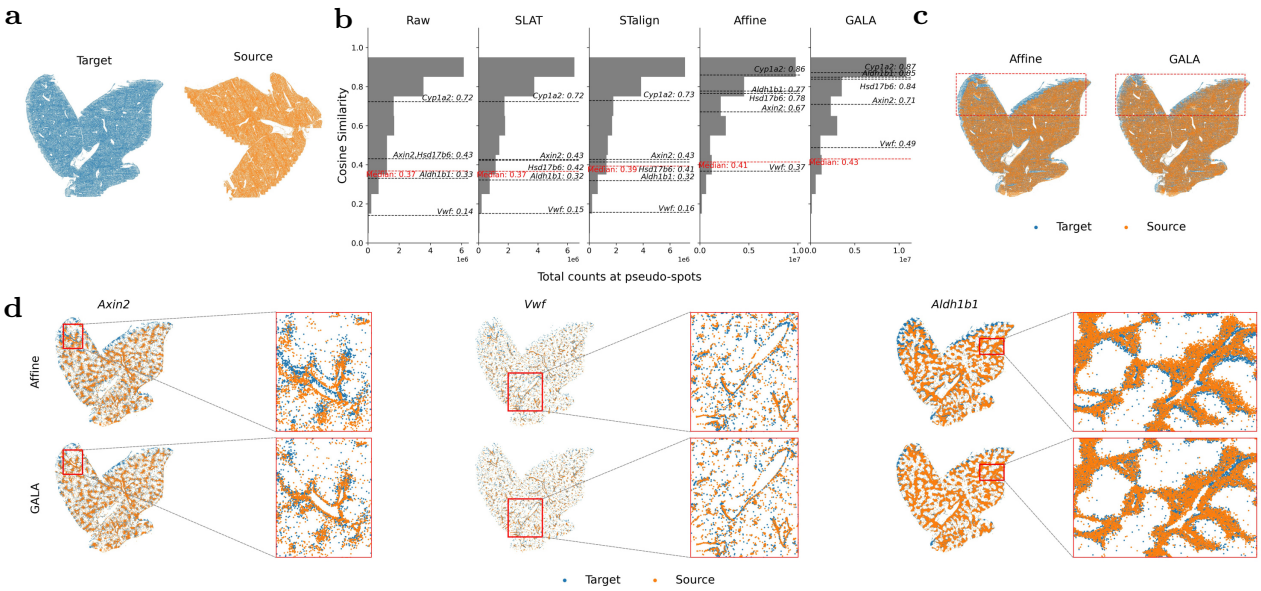

Figure 4: **Alignment of two MERFISH mouse liver replicates.** **a** Raw slices for source and target data. **b** Distribution of cosine similarity scores for 134 liver marker genes and 3 informative genes against the total gene expression at pseudo-spots (aggregated into 30 μm) after alignment by Raw, SLAT, STalign, Affine, and GALA. Red and black dashed lines indicate the median and representative genes, respectively. **c** Visual comparison of whole-slice alignment results by Affine and GALA. Cells in target and aligned source data are indicated as blue and orange, respectively. The top area in red rectangular highlights the significant improvement of local transformation from GALA. **d** Alignment of three representative genes at whole-slice and zoomed-in regions, comparing Affine and GALA.

2.4 Cell-to-spot alignment across ST technologies

GALA also supports mismatched resolution alignment across technologies, i.e., aligning single-cell resolved ST data (e.g. from MERFISH, Xenium) to multi-cellular resolved ST data (e.g. from Visium). Such alignment enables mapping of cell types and gene expression domains across resolution scales, supporting comparative studies between experiments and modalities[17,21,35]. In the fourth example, the source and target data are Fresh Frozen mouse brain replicate 2 from the Xenium platform and a mouse coronal section from the Visium platform, respectively. The Xenium dataset contains 149,292 single cells with high-resolution gene expression, whereas the Visium data comprise 2,634 spots with low-resolution expression profiles. As shown in Fig. 5a, the Xenium slice spans a full coronal brain section, while the Visium slice captures only cortical and subcortical structures in the left hemisphere, making this a partial alignment across resolutions. Baseline methods include SLAT and Affine. STalign was not considered because it cannot directly perform cell-to-spot alignment; instead, it aligns single-cell ST data to the registered H&E image corresponding to the multi-cellular ST data, requiring manually selected landmarks.

To quantify alignment, we computed cosine similarity of gene expression vectors aggregated into 30 μm pseudo-spots for all common genes (Fig. 5b). Median similarity scores show that GALA and Affine substantially outperform Raw and SLAT. We also examined alignment of three spatially informative marker genes: *Slc17a7* (excitatory cortical neurons), *Gfap* (astrocytes near the ventricular zone and midline), and *Neurod6* (hippocampus and olfactory bulb). Fig. 5b demonstrates that, compared with raw similarities of 0.61, 0.38,

and 0.21, GALA increased these to 0.79, 0.67, and 0.56, outperforming Affine-aligned results of 0.78, 0.61, and 0.54, and SLAT-aligned results of 0.72, 0.54, and 0.48.

We further visually evaluated alignment across methods. As shown in Fig. 5c, SLAT performs poorly in this cross-resolution setting, primarily due to its reliance on one-to-one adversarial graph matching followed by a global affine transformation, which introduces noise and fails to capture local nonlinear deformations. In contrast, GALA and Affine achieve markedly improved global alignment, with GALA producing finer geometric correspondence and better preservation of fine-scale structural coherence. Supplementary Fig. S11 provides additional comparisons, showing target data as registered H&E images. Fig. 5d demonstrates alignment at the molecular level for GALA and Affine. Regions highlighted by red boxes exemplify alignment differences, with GALA achieving more coherent correspondence between aligned source and target data.

These results indicate that in the cell-to-spot alignment scenario, GALA outperforms baseline methods, and both components of the global affine and local diffeomorphic refinement contribute substantially to the improved alignment. These conclusions are consistent with observations in the cell-to-cell scenario in Section 2.3.

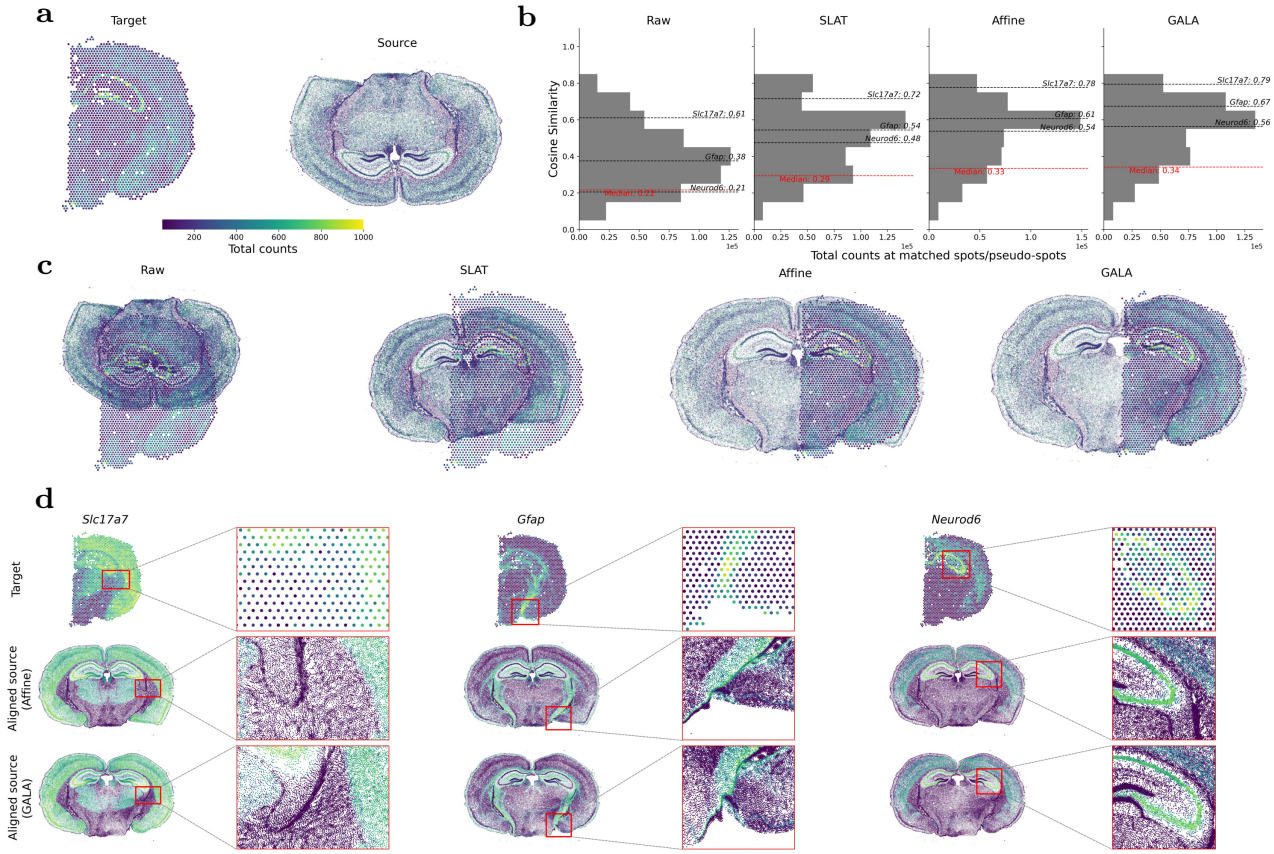

Figure 5: **Alignment of mouse coronal sections from Visium and Xenium.** **a** Visualisation of Visium (spot-level, target) and Xenium (cell-level, source) slices. Brighter colours indicate higher gene expression. **b** Distribution of cosine similarities between gene expression in Visium target versus aligned Xenium source at matched spots/pseudo-spots. Red dashed lines indicate median similarity; black dashed lines highlight three informative genes. **c** Comparison of unaligned slices with slices aligned by SLAT, Affine, and GALA. **d** Spatial expression of three informative genes, *Slc17a7*, *Gfap*, and *Neurod6*, in Visium (top), Affine-aligned (middle), and GALA-aligned Xenium data (bottom), with magnified regions indicated by red rectangles.

2.5 Cell-level transcriptomics to histological image alignment within the same tissue section

Lastly, we evaluated GALA for aligning ST data to paired H&E images within the same tissue section and experimental platform. Such registration is necessary to correct physical and technical misalignments arising from the automated instrument workflow, rather than biological variation[12,25]. Misalignments are primarily rigid-body transformations (translation and rotation) introduced during mechanical handling, with minor scaling effects possible due to optical calibration. Accurate alignment under this scenario bridges the gap between the abstract molecular map and interpretable tissue morphology.

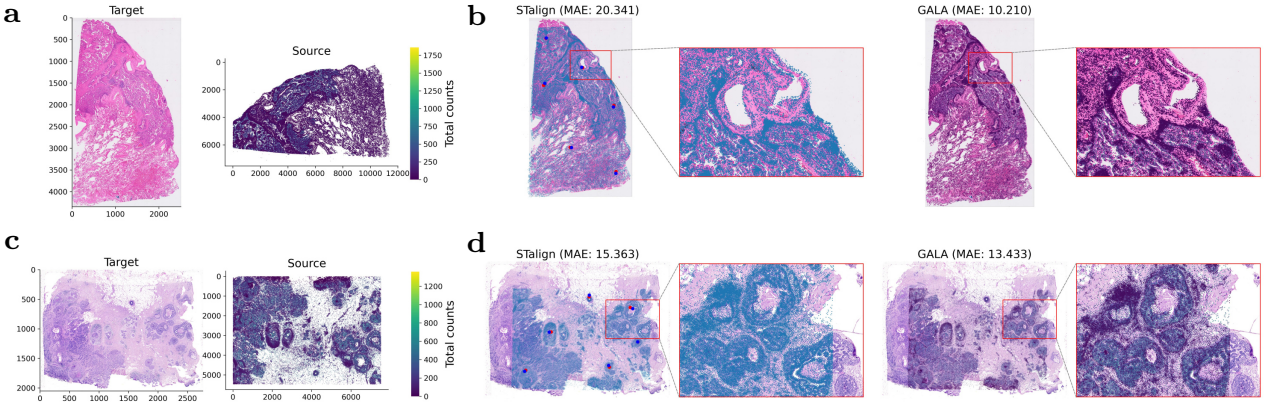

Figure 6: **Alignment of Xenium single-cell RNA data to paired H&E images.** **a** H&E histological image (target) and paired Xenium single-cell RNA data (source) from a human lung cancer section, where brighter colours indicate higher total gene expression. **b** Alignment results of **a** by landmark-based STalign and landmark-free GALA, with MAE indicated. Blue/red dots in the STalign panel denote manually annotated landmarks in source/target data. As STalign ignores gene expression, its aligned data are visualised as a blue overlay on the target H&E image. Adjacent zoom-in panels highlight regions marked by red boxes. **c** H&E image (target) and paired single-cell RNA data (source) from a human breast cancer section, with brighter colours indicating higher expression. **d** Alignment results of **c** by STalign and GALA, with MAE indicated and magnified regions within red rectangles.

We applied GALA to two datasets: an FFPE human lung cancer section (Experiment 1, Xenium In Situ) and an FFPE human breast cancer section (Replicate 1, Xenium In Situ Sample 1). Fig. 6a and c show the H&E image (left; target) and the corresponding high-resolution single-cell expression data (right; source). Clear differences in scale, rotation, and translation are visible, while Fig. 6c also shows a reflection between the expression and histology coordinates, likely arising from slide orientation or scanning artefacts.

To our knowledge, besides GALA, only STalign addresses the problem of transcriptomics-to-histology alignment. However, STalign requires manually annotated landmarks for affine transformation without utilizing gene expression, whereas GALA is fully landmark-free and incorporates gene expression. We compared STalign and GALA using the mean absolute error (MAE) of landmark coordinates to quantitatively assess alignment (see Section 4.4). Fig. 6b shows the lung cancer dataset results: using six representative landmarks, STalign achieved an MAE of 20.341, whereas GALA achieved a more precise alignment with MAE of 10.210. Fig. 6d further demonstrates GALA's superior performance on the breast cancer dataset, with MAEs of 15.363 and 13.433 for STalign and GALA, respectively.

Landmark selection is inherently subjective and prone to user bias. Different experts may select different landmarks, leading to inconsistent alignments. For instance, when landmarks are sparse or lack anatomical specificity, STalign's performance degrades markedly (Supplementary Fig. S13). By contrast, GALA removes this human-in-the-loop variability, making the alignment fully automated, reproducible, and objective—a critical advantage for large-scale studies and ensuring consistency across research groups.

2.6 Computational efficiency

Across multiple benchmarks, GALA demonstrated superior computational efficiency, striking an optimal balance between speed and the accuracy of its alignment. It consistently outperformed other nonlinear methods, being substantially faster than GPSA and ST-GEARS (averaging 135.22 seconds compared with 245.61 and 154.69 seconds, respectively, on slices 2 and 3 of Sample C). While specialized rigid methods like PASTE, STAligner, and SLAT were faster (all under 30 seconds), they yielded inferior accuracy. Furthermore, GALA also outperformed STalign, its primary competitor built on the same LDDMM core, across multiple datasets. For example, on the human breast cancer dataset, GALA achieved an average runtime of 74.16 seconds compared with STalign's 512.06 seconds, due to GALA's streamlined optimisation (see Section 4.2) and GPU acceleration. Crucially, GALA excelled at scale where many methods failed due to excessive memory demands (e.g. 500 GB, MERFISH mouse liver data); it successfully aligned datasets of hundreds of thousands of cells in under a minute, far outpacing SLAT's lengthy cell-matching phase. This efficiency extended to multimodal and mismatched-resolution scenarios, solidifying GALA's advantage for rapid, high-fidelity alignment of large-scale ST data. The details of comparison of computational efficiency with baseline methods are provided at Supplementary Notes S2.2.

3 Discussion

In this study, we introduced GALA, a unified, landmark-free spatial alignment framework capable of addressing a wide range of challenges in ST alignment. A central strength of GALA lies in its multimodal rasterisation strategy, which converts diverse transcriptomic and histological data into co-registered raster tensors. This standardised representation harmonises heterogeneous resolutions, modalities, and sequencing platforms, and naturally scales from spot-level to dense single-cell data, enabling the aggregation of millions of cells onto a user-defined grid where each pixel encodes local features. By design, GALA jointly leverages transcriptomic and histological information with tunable weights, enhancing both alignment accuracy and interpretability. Its coarse-to-fine optimisation couples a global affine transformation (GA) with local diffeomorphic deformation (LDDMM), providing flexible adaptation to cross-resolution, cross-modality, and partial tissue overlap scenarios: the GA rapidly identifies overlapping regions for initialisation, while LDDMM achieves precise, topology-preserving alignment.

Compared with existing approaches (Tab. 1), GALA offers broader applicability and improved robustness. Spot-level methods such as PASTE and STAligner emphasise rigid alignment, whereas STalign and GPSA provide non-rigid deformations but with limited modality support or scalability. SLAT relies on one-to-one graph matching, which restricts its flexibility in heterogeneous tissue contexts. Coarse-to-fine frameworks such as ST-GEARS are specialised for particular modalities or task types, and most baselines, except for PASTE2, cannot handle partial tissue overlaps. In contrast, GALA uniquely integrates rigid and non-rigid deformations within a coupled framework and supports spot–to-spot, cell–to-cell, cell–to-spot, and transcriptomic–to-histology alignment across both full and partial tissue overlaps. Importantly, GALA can align Source–Target and Target–Source interchangeably (Supplementary Figs. S10,S12), ensuring that results are independent of the arbitrary designation of source or target and that inferred spatial relationships are reproducible rather than algorithmic artefacts. While STalign also employs LDDMM, it cannot leverage gene expression or H&E image data, does not support spot-to-spot or direct cell-to-spot alignment, and relies on manually defined landmarks.

Table 1: Summary of all baselines and GALA

Methods	Strategies		Application scenarios				Modalities			Regions		Landmark-free	Techniques
	Rigid	Non-rigid	spot-to-spot	cell-to-cell	cell-to-spot	st.-to-histo.	coords.	expr.	histo.	full	partial		
PASTE	✓		✓				✓	✓		✓		✓	Optimal transport
PASTE2	✓		✓				✓	✓	✓	✓	✓	✓	Optimal transport+probability matching
STAligner	✓		✓	✓			✓	✓		✓		✓	Iterative closet point (ICP) on landmark domain
STalign		✓		✓		✓	✓			✓	✓		LDDMM
GPSA		✓	✓				✓	✓	✓	✓		✓	Gaussian process-based deformation
SLAT	One-to-One		✓	✓	✓		✓	✓				✓	Graph adversarial matching+ICP
ST-GEARS	✓	✓	✓	✓			✓	✓		✓		✓	Procrustes + elastic registration
GALA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	GA+LDDMM (Coupled)

Note: ✓ indicates full support for the corresponding feature.

Despite its strengths, several limitations of GALA should be considered. Multimodal fusion of transcriptomic and histological data does not always improve alignment, reflecting both methodological constraints

and the inherent biological reality that molecular and morphological states are not perfectly concordant[36]. Morphologically similar cells may exhibit distinct molecular profiles, and sharp molecular boundaries may lack histological counterparts[35]. Such discordances can introduce noise, highlighting the need for adaptive, biologically informed fusion strategies that weight each modality based on local context[3]. Furthermore, while GALA demonstrates robustness across adjacent slices and conditions, it focuses solely on spatial alignment and does not explicitly address downstream tasks such as cell-type annotation or correction for batch effects from sample preparation. Future iterations could incorporate batch-aware modelling to enhance cross-sample generalisation.

Nonetheless, by providing accurate and flexible spatial alignment, GALA establishes a critical foundation for downstream applications, including cross-sample atlas construction, consistent cell-type mapping, and trajectory inference. Beyond transcriptomics, the framework can be naturally extended to other spatially resolved omics modalities, such as spatial proteomics or metabolomics, where multimodal alignment across molecular layers can enable integrated analyses of tissue architecture, functional states, and molecular interactions[? ?]. Looking forward, the integration of precise alignment with subsequent biological analyses offers a promising direction. We anticipate that GALA will serve as a versatile component within broader spatial omics pipelines, potentially coupled with modules for spatially informed gene expression modelling, or used to infer conserved spatial programmes across individuals, conditions, and technologies.

4 Methods

4.1 Data preprocessing and rasterisation

All datasets used in this study are preprocessed according to quantitative quality-control criteria, including standard procedures such as cell filtering, gene selection, and log-transformation. Detailed preprocessing steps for each dataset are provided in the accompanying `GALA` code package.

To integrate transcriptomic and histological information within a unified optimisation framework, GALA represents all data modalities as co-registered raster tensors. Each dataset is projected onto a regular two-dimensional grid Ω with spacing dx and dimensions $H \times W$, onto which transcriptomic or histological measurements are rasterised. Let $\boldsymbol{x}_i \in \mathbb{R}^2$ denote the coordinates of spatial anchors (i.e., cells or spots) for $i = 1, \ldots, n$ and $\boldsymbol{y}_i \in \mathbb{R}^p$ their corresponding gene expression vectors. The rasterised transcriptomic tensor $I_{\text{st}}(\boldsymbol{x})$ is defined as

$$I_{\text{st}}(\boldsymbol{x}) = \sum_{i=1}^{n} \boldsymbol{y}_i \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{2\sigma^2}\right) \in \mathbb{R}^{p \times H \times W},$$

which smooths expression vectors over spatial coordinates using Gaussian kernels, thereby encoding both molecular intensity and spatial proximity. The Gaussian width σ is set to $2dx$ by default and can be adjusted to match the spatial characteristics of each dataset: larger values suppress noise through stronger smoothing, whereas smaller values preserve fine-grained spatial patterns (see examples in Supplementary Notes 1.1). When datasets differ in spatial scale (e.g., cell-level versus spot-level), the Gaussian kernel widths can be tuned independently. This ensures that datasets with mismatched resolutions are rasterised onto grids with comparable effective resolutions, enabling cross-resolution alignment (e.g., cell-to-spot). Moreover, the rasterised representation substantially reduces computational cost by operating on downsampled raster signals rather than high-dimensional molecular matrices. In this study, we typically set $dx \approx 1$ unit for spot-level data (corresponding to low-resolution scale) and $dx = 30$ μm for cell-level data, balancing spatial fidelity and computational efficiency. To stabilise regions with weak or noisy expression, an additional channel recording the total expression across all genes is included, yielding the final transcriptomic tensor $I_{\text{st}} \in \mathbb{R}^{(p+1) \times H \times W}$.

When histological images (e.g., H&E staining) are available, they are rasterised in parallel onto the same grid used for transcriptomics. Because each image pixel is spatially registered to the transcriptomics, the resulting histological raster is naturally co-aligned with the transcriptomic tensor, enabling a unified multimodal representation for subsequent alignment. Specifically, raw RGB images are first converted to greyscale, and structural contrast is enhanced using gradient magnitude filters (e.g., the Sobel operator[37]). Gaussian smoothing is then applied to generate a single-channel histological raster tensor, $I_{\text{histo}} \in \mathbb{R}^{1 \times H \times W}$. Alternatively, RGB images can be processed channel-wise to obtain $I_{\text{histo}} \in \mathbb{R}^{3 \times H \times W}$; however, gradient-based greyscale maps are used by default due to their robust structural representation and reduced sensitivity to staining variation.

Thus, each dataset is ultimately represented by a raster tensor $I \in \mathbb{R}^{c\times H\times W}$, where $I = I_{\text{st}}$ when only transcriptomic data are used ($c = p+1$), and $I = (I_{\text{st}}, I_{\text{histo}})_{\text{channel}}$ when histology is included ($c = p+2$), with concatenation performed along the channel dimension. These unified rasters provide a continuous, resolution-adaptive representation on which all subsequent optimisation is performed.

4.2 Alignment objective and optimisation

GALA models the spatial correspondence between source and target raster tensors through a composite transformation $(\phi^v \cdot A)(\boldsymbol{x}) = \phi^v(A(\boldsymbol{x}))$, which combines a global affine matrix A to correct large-scale differences (rotation, translation, scaling, and reflection) with a diffeomorphic deformation ϕ^v parameterised by a time-varying velocity field v_t ($t \in [0, 1]$) for each spatial location $\boldsymbol{x} \in \mathbb{R}^2$ (see Fig. 1b and Supplementary Notes S1.2).

Building on the coarse affine alignment that estimates overlapping regions between source and target, GALA refines local correspondences via a probabilistic matching model inspired by EM-like LDDMM frameworks[12,29], while simplifying the latent classification to two classes: Matching and Non-Matching (e.g. background, missing, artefacts). Each grid location $\boldsymbol{x} \in \Omega$ is therefore assigned a latent label indicating whether it belongs to a reliably aligned region or to background/artefactual areas lacking biological correspondence. Let $I(\boldsymbol{x}), J(\boldsymbol{x}) \in \mathbb{R}^{c\times H\times W}$ denote the source and target raster tensors. Under affine transformation $A(\boldsymbol{x})$ and diffeomorphic mapping $\phi^v(\boldsymbol{x})$, the target intensity $J(\boldsymbol{x})$ follows a Gaussian mixture model:

$$J(\boldsymbol{x}) \sim \begin{cases} \mathcal{N}(I((\phi^v \cdot A)(\boldsymbol{x})), \Sigma_M), & \boldsymbol{x} \text{ is matched}, \\ \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N), & \boldsymbol{x} \text{ is non-matched}, \end{cases}$$

where $\Sigma_M = \sigma_M^2 I_c$ and $\Sigma_N = \sigma_N^2 I_c$ denote covariances for matched and non-matched components, respectively, I_c is the c-dimensional identity matrix, and $\boldsymbol{\mu}_N$ represents the mean intensity of the non-matching region, estimated iteratively during model fitting. Given class priors π_M and π_N ($\pi_M + \pi_N = 1$), the posterior probability that location $\boldsymbol{x}$ belongs to the matched class is

$$P_M(\boldsymbol{x}) = \frac{\pi_M \mathcal{N}(J(\boldsymbol{x}); I((\phi^v \cdot A)(\boldsymbol{x})), \Sigma_M)}{\pi_M \mathcal{N}(J(\boldsymbol{x}); I((\phi^v \cdot A)(\boldsymbol{x})), \Sigma_M) + \pi_N \mathcal{N}(J(\boldsymbol{x}); \boldsymbol{\mu}_N, \Sigma_N)},$$

where $\mathcal{N}(\cdot; \cdot, \cdot)$ denotes the multivariate Gaussian probability density function. High $P_M(\boldsymbol{x})$ values indicate strong source–target correspondence, whereas low values highlight background or missing areas. During optimisation, $P_M(\boldsymbol{x})$ is updated iteratively from local intensity residuals in an EM-like fashion, enabling joint refinement of the diffeomorphic deformation and probabilistic matching. This mechanism allows the alignment to adaptively focus on confidently matched regions while downweighting unreliable or artefactual areas.

The multimodal objective function jointly estimates the affine matrix A, local deformation ϕ^v, and matching weights P_M:

$$\begin{aligned} \mathcal{L}(A, \phi^v, P_M) &= \alpha d_{\text{st}}(I_{\text{st}}(\phi^v \cdot A), J_{\text{st}} \mid P_M) + (1-\alpha) d_{\text{histo}}(I_{\text{histo}}(\phi^v \cdot A), J_{\text{histo}} \mid P_M) + \text{Reg}(v) \\ &= \alpha \frac{1}{2\sigma_{M_1}^2} \sum_{g=1}^{p+1} \sum_{\boldsymbol{x}} P_M(\boldsymbol{x}) \, \| I_{\text{st}}^g((\phi^v \cdot A)(\boldsymbol{x})) - J_{\text{st}}^g(\boldsymbol{x}) \|_2^2 \\ &+ (1-\alpha) \frac{1}{2\sigma_{M_2}^2} \sum_{\boldsymbol{x}} P_M(\boldsymbol{x}) \, \| I_{\text{histo}}((\phi^v \cdot A)(\boldsymbol{x})) - J_{\text{histo}}(\boldsymbol{x}) \|_2^2 + \frac{1}{2\sigma_R^2} \int_0^1 \| v_t \|_V^2 \, dt \end{aligned} \tag{1}$$

where σ_{M_1}, σ_{M_2}, and σ_R control feature matching and deformation smoothness, and $\alpha \in (0, 1)$ balances transcriptomic and histological contributions. The final term regularises v_t in a reproducing kernel Hilbert space (RKHS), ensuring smooth and invertible mappings[28]. The selected $p + 1$ gene channels and histological image channel act as alignment "landmarks", guiding the optimisation—principally differing from STalign, which does not utilise this information. Additional details are provided in Supplementary Notes S1.2.

We introduce a coupled optimisation framework that unifies global and local alignment within a single objective, rather than treating them sequentially. Unlike conventional pipelines that perform coarse alignment and fine registration independently, our coordinated coarse-to-fine scheme iteratively reinforces both stages. In the global stage, a GA[27] conducts a population-based search to estimate affine parameters and overlapping regions, providing robust initialisation that avoids local minima. Subsequently, the local stage resolves residual nonlinear discrepancies using LDDMM[28], modelling smooth, invertible deformations suitable for fine-grained spatial adjustments. To reconcile local deformations with modality-specific information, an EM-like

mechanism iteratively updates both the spatial weighting function P_M and the time-dependent velocity field v_t based on the current global alignment. LDDMM captures flexible local shape variations, while EM-like updates ensure refinements respect spatial overlap and modality-specific correspondence[12,29]. This bidirectional coupling, where coarse estimates guide local refinements and local updates feedback to adjust global parameters, enables more stable convergence than purely hierarchical or decoupled approaches. Empirically, under default parameter settings (Supplementary Tab. S1), only two global–local episodes suffice to achieve stable alignment across most datasets (Supplementary Notes S1.3).

4.3 Selection of informative genes and hyper-parameters

GALA requires two key hyper-parameters: the number of selected informative genes p, which guide alignment, and the relative weight α that balances contributions from transcriptomics profiles versus histological images.

For gene selection, we adopted the R^2-based strategy from GPSA[18], which ranks genes by their predictability from local spatial context. Specifically, for each gene, a k-nearest neighbour (kNN) regression model is trained to predict expression levels from the spatial coordinates of neighbouring spots or cells. The coefficient of determination (R^2) quantifies spatial predictability, and the top p genes with the highest R^2 values from the source data are retained. Alternative approaches, such as identifying spatially variable genes (SVG)[38,39] or using known marker genes, are also viable; we chose the R^2-based method for its simplicity, computational efficiency, and robust performance across diverse datasets, as previously validated by GPSA. Systematic sensitivity analysis indicated that $p = 3$ (top three informative genes) suffices even under partial alignment scenarios, and this value is set as default. The balance parameter α was observed to yield stable performance across its range $(0, 1)$; we set $\alpha = 0.5$ when histological image data is available, and $\alpha = 1$ otherwise. Detailed robustness and sensitivity analyses are provided in Supplementary Notes S2.1.

4.4 Evaluation metrics

To evaluate alignment performance across diverse experimental scenarios, we employ complementary metrics capturing spatial accuracy, biological coherence, and molecular consistency. The choice of metric depends on the availability of anatomical annotations and the spatial resolution of the data.

When anatomical labels or ground truth are available, we assess alignment fidelity using two strategies. First, label consistency accuracy identifies mutual nearest-neighbour pairs between aligned slices and reports the proportion sharing the same anatomical label, directly quantifying structural correspondence[8] (Supplementary Notes S1.4). Second, biological coherence is evaluated via clustering-based metrics: a joint GMM is fitted to combined spatial and gene expression features from both aligned source and target slices, and cluster assignments are compared with ground-truth labels using ARI and NMI. These metrics assess how well alignment preserves region-specific transcriptional and anatomical structure.

For spot-based datasets lacking reliable anatomical annotations, spatial cross-correlation metrics are employed[40,41]. Given aligned source and target slices, a Gaussian-kernel weight matrix based on inter-spot distances is computed, and for each gene, the correlation between transcriptomic profiles is evaluated, weighted by spatial proximity. Scores are normalised and averaged across genes to provide a global measure, capturing preservation of spatially coherent expression patterns without requiring labels.

In cell-based datasets, direct cell-to-cell comparison is often unreliable due to biological heterogeneity and measurement sparsity. Following STalign[12], neighbouring cells are aggregated into pseudo-spots to stabilise expression signals. Cosine similarity between corresponding pseudo-spots across aligned slices then quantifies local molecular agreement, mitigating single-cell noise and providing an effective fidelity measure for high-resolution single-cell datasets without histological or anatomical ground truth.

For transcriptomics-to-histology alignment, we use the MAE of annotated landmark coordinates:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\boldsymbol{x}_i - \boldsymbol{x}'_i|,$$

where $i = 1, \ldots, n$ indexes landmarks, and $\boldsymbol{x}_i, \boldsymbol{x}'_i \in \mathbb{R}^2$ denote the coordinates of the i-th landmark in the target and aligned source data, respectively. Lower MAE indicates better alignment, with zero representing perfect correspondence. Importantly, this evaluation is independent of the alignment procedure, making it compatible with any landmark-free method.

4.5 Implementation techniques for all alignment scenarios

GALA is a versatile alignment framework applicable to widespread application scenarios. For example, whether the source and target data i) belong to the same modality or not, ii) are consistent in spatial resolution or not, and iii) are fully or partially overlapped (see Fig. 1b). In the following, we detail how GALA is applied across these scenarios.

First, because rasterisation maps each dataset onto a user-defined grid, same-modality alignment under both matched and mismatched resolutions, including spot-to-spot, cell-to-cell, and cell-to-spot scenarios, is handled directly through the choice of grid spacing dx (see Section 4.1). No additional algorithmic modifications are required: once rasterised, all datasets are processed identically by the optimisation pipeline. This design enables resolution-aware and platform-agnostic alignment across Visium, Xenium, MERFISH, and related technologies.

Second, GALA supports cross-modality alignment, such as registering spatial transcriptomic measurements to histological images from the same tissue section. Histology images are converted to greyscale and transformed into structural contrast maps, whereas transcriptomic measurements are encoded as rasterisation tensors, allowing both to be treated as image-like modalities. Alignment is then performed under the multimodal objective:

$$\mathcal{L}(A, \phi^v, P_M) = \frac{1}{2\sigma_M^2} \sum_{\boldsymbol{x}} P_M(\boldsymbol{x}) \left\| I_{\text{st}}((\phi^v \cdot A)(\boldsymbol{x})) - J_{\text{histo}}(\boldsymbol{x}) \right\|_2^2 + \frac{1}{2\sigma_R^2} \int_0^1 \|v_t\|_V^2 dt,$$

where $I_{\text{st}} \in \mathbb{R}^{1\times H_1 \times W_1}$ and $J_{\text{histo}} \in \mathbb{R}^{1\times H_2 \times W_2}$. This variant enables cross-modality alignment without requiring landmarks or manual correspondence, facilitating integration across different types of measurements in ST studies.

Third, GALA explicitly accommodates partial alignment. The framework first performs a population-based affine parameter search to approximate the overlapping region between the two datasets. This global affine estimate then guides the subsequent LDDMM-based refinement. During refinement, an EM-like procedure updates pixel-wise correspondence weights, allowing the model to concentrate on structurally matched areas while downweighting unmatched or noisy areas (see Section 4.2). This mechanism confers robustness in scenarios involving incomplete, distorted, or physically damaged tissue.

Finally, GALA's multimodal rasterisation and flexible objective function allow natural extension to other omics modalities beyond transcriptomics. For instance, metabolic or proteomic imaging data can be rasterised similarly, enabling alignment of metabolite distributions or protein expression patterns to transcriptomic or histological references. This generalisability makes GALA a broadly applicable framework for multi-omics spatial alignment, supporting integrative studies across diverse molecular modalities and imaging platforms.

5 Data Availability

All datasets used in this study are publicly available from the original sources as cited below and were employed to benchmark GALA across various alignment tasks: The **DLPFC dataset**, used for full and partial spot-based alignment, was obtained from the Lieber Institute (https://github.com/LieberInstitute/HumanPilot/tree/master/10X); preprocessed data were downloaded from Zenodo (https://zenodo.org/records/6334774). The **mouse brain sagittal–posterior dataset** (Sections 1 and 2), used to evaluate multimodal alignment and histology integration at the spot level, was downloaded from 10x Genomics: https://www.10xgenomics.com/datasets/mouse-brain-serial-section-1-sagittal-posterior-1-standard-1-1-0 and https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard-1-1-0. The **MERFISH mouse liver dataset** (L2S1 and L2S2 from Animal 2), used for cell-level alignment without landmarks, was obtained from Vizgen: https://info.vizgen.com/mouse-liver-data. The **Visium mouse brain coronal section dataset**, used for cross-platform spot-to-cell alignment, was downloaded from: https://www.10xgenomics.com/datasets/adult-mouse-brain-coronal-section-fresh-frozen-1-standard. The **Xenium mouse brain dataset** (Replicate 2), used for cross-platform and cross-modality alignment with Visium: https://www.10xgenomics.com/datasets/fresh-frozen-mouse-brain-replicates-1-standard. The **Xenium human lung cancer FFPE dataset** (Experiment 1), used for evaluating image-to-cell alignment and multimodal registration in human tissue: https://www.10xgenomics.com/datasets/xenium-human-lung-cancer-post-xenium-technote. The **Xenium human breast cancer FFPE dataset** (Replicate 1), used for image-to-cell alignment and multimodal registration in human tissue: https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast.

All preprocessed ST datasets and corresponding alignment outputs generated in this study are publicly available via Zenodo (https://doi.org/10.5281/zenodo.17576488) to allow full reproduction of the analyses and figures presented.

6 Code Availability

The source code implementing GALA is openly available at https://github.com/TaoDing2/GALA under an open-source licence. The repository provides detailed instructions for installation, data preparation, and reproduction of all analyses. In addition, the Zenodo repository contains scripts and processed data necessary to replicate all experiments and figures reported in this study.

7 Competing interests

No competing interest is declared.

8 Author contributions

T.D. and P.Z. conceived the study. T.D. developed the GALA algorithm, performed data analyses, and drafted the manuscript. P.Z. supervised the project, provided conceptual guidance, and contributed to manuscript revisions. All authors read and approved the final manuscript.

References

[1] Asp, M. *et al.* A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 (2019).

[2] Marx, V. Method of the year: spatially resolved transcriptomics. *Nature methods* **18**, 9–14 (2021).

[3] Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nature methods* **19**, 534–546 (2022).

[4] Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

[5] Ke, R., Jaffe, A. E. *et al.* High-plex spatially resolved rna profiling in formalin-fixed, paraffin-embedded tissue using the 10x genomics xenium in situ platform. *Nature Biotechnology* **42**, 219–229 (2024).

[6] Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed rna profiling in single cells. *Science* **348**, aaa6090 (2015).

[7] Khan, M., Arslanturk, S. & Draghici, S. A comprehensive review of spatial transcriptomics data alignment and integration. *Nucleic Acids Research* **53**, gkaf536 (2025).

[8] Liu, Y. & Yang, C. Computational methods for alignment and integration of spatially resolved transcriptomics data. *Computational and Structural Biotechnology Journal* (2024).

[9] Lee, J., Yoo, M. & Choi, J. Recent advances in spatially resolved transcriptomics: challenges and opportunities. *BMB reports* **55**, 113 (2022).

[10] Tward, D. *et al.* 3d mapping of serial histology sections with anomalies using a novel robust deformable registration algorithm. In *International Workshop on Multimodal Brain Image Analysis*, 162–173 (Springer, 2019).

[11] Ekvall, M. *et al.* Spatial landmark detection and tissue registration with deep learning. *Nature Methods* **21**, 673–679 (2024).

[12] Clifton, K. *et al.* Stalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nature communications* **14**, 8123 (2023).

[13] Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nature Methods* **19**, 567–575 (2022).

[14] Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature methods* **18**, 1352–1362 (2021).

[15] Kang, L., Zhang, Q., Qian, F., Liang, J. & Wu, X. Benchmarking computational methods for detecting spatial domains and domain-specific spatially variable genes from spatial transcriptomics data. *Nucleic Acids Research* **53**, gkaf303 (2025).

[16] Chen, T.-Y., You, L., Hardillo, J. A. U. & Chien, M.-P. Spatial transcriptomic technologies. *Cells* **12**, 2042 (2023).

[17] Zhou, X., Dong, K. & Zhang, S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science* **3**, 894–906 (2023).

[18] Jones, A., Townes, F. W., Li, D. & Engelhardt, B. E. Alignment of spatial genomics data using deep gaussian processes. *Nature Methods* **20**, 1379–1387 (2023).

[19] Xia, T. *et al.* St-gears: Advancing 3d downstream research through accurate spatial information recovery. *Nature Communications* **15**, 7806 (2024).

[20] Zhou, Z., Zhong, Y., Zhang, Z. & Ren, X. Spatial transcriptomics deconvolution at single-cell resolution using redeconve. *Nature Communications* **14**, 7930 (2023).

[21] Wan, X. *et al.* Integrating spatial and single-cell transcriptomics data using deep generative models with spatialscope. *Nature Communications* **14**, 7848 (2023).

[22] Xia, C.-R., Cao, Z.-J., Tu, X.-M. & Gao, G. Spatial-linked alignment tool (slat) for aligning heterogenous slices. *Nature Communications* **14**, 7236 (2023).

[23] Hu, J. *et al.* Statistical and machine learning methods for spatially resolved transcriptomics with histology. *Computational and structural biotechnology journal* **19**, 3829–3841 (2021).

[24] Liu, X., Zeira, R. & Raphael, B. J. Paste2: partial alignment of multi-slice spatially resolved transcriptomics data. *bioRxiv* (2023).

[25] Shi, Z., Zhu, F. & Min, W. Inferring multi-slice spatially resolved gene expression from h&e-stained histology images with stmcl. *Methods* **234**, 187–195 (2025).

[26] Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience* **24**, 425–436 (2021).

[27] Gad, A. F. Pygad: An intuitive genetic algorithm python library. *Multimedia tools and applications* **83**, 58029–58042 (2024).

[28] Beg, M. F., Miller, M. I., Trouvé, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision* **61**, 139–157 (2005).

[29] Tward, D. *et al.* Diffeomorphic registration with intensity transformation and missing data: Application to 3d digital pathology of alzheimer's disease. *Frontiers in neuroscience* **14**, 52 (2020).

[30] Li, H. *et al.* Santo: a coarse-to-fine alignment and stitching method for spatial omics. *Nature Communications* **15**, 6048 (2024).

[31] Shi, X. *et al.* Probabilistic cell/domain-type assignment of spatial transcriptomics data with spatialanno. *Nucleic acids research* **51**, e115–e115 (2023).

[32] Lu, Y. *et al.* Spatial transcriptome profiling by merfish reveals fetal liver hematopoietic stem cell niche architecture. *Cell Discovery* **7**, 47 (2021).

[33] Sun, T. *et al.* Axin2+ pericentral hepatocytes have limited contributions to liver homeostasis and regeneration. *Cell stem cell* **26**, 97–107 (2020).

[34] Nightingale, T. & Cutler, D. The secretion of von w illebrand factor from endothelial cells; an increasingly complicated story. *Journal of Thrombosis and Haemostasis* **11**, 192–201 (2013).

[35] Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology* **40**, 661–671 (2022).

[36] Chelebian, E., Avenel, C. & Wählby, C. Combining spatial transcriptomics with tissue morphology. *Nature Communications* **16**, 1–13 (2025).

[37] Jin-Yu, Z., Yan, C. & Xian-Xiang, H. Edge detection of images based on improved sobel operator and genetic algorithms. In *2009 International Conference on Image Analysis and Signal Processing*, 31–35 (IEEE, 2009).

[38] Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nature methods* **15**, 339–342 (2018).

[39] Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nature methods* **17**, 193–200 (2020).

[40] Chen, Y. A new methodology of spatial cross-correlation analysis. *PloS one* **10**, e0126158 (2015).

[41] Tan, Y. *et al.* Integrative cross-sample alignment and spatially differential gene analysis for spatial transcriptomics. *bioRxiv* 2025–06 (2025).