

GALA: a unified landmark-free framework for coarse-to-fine spatial alignment across resolutions and modalities in spatial transcriptomics

Tao Ding and Pengcheng Zeng*

Institute of Mathematical Sciences, ShanghaiTech University, 393 Middle Huaxia Road, Pudong New Area, Shanghai, 201210, China

*Corresponding author. Institute of Mathematical Sciences, ShanghaiTech University, 393 Middle Huaxia Road, Pudong New Area, Shanghai, 201210, China. E-mail: 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 optimization. Its modality-aware rasterization harmonizes 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

Introduction

Spatial transcriptomics (ST) technologies enable high-throughput mapping of gene expression within intact tissue sections, providing detailed insights into cellular organization, tissue architecture, and disease-associated microenvironments [1–3]. Contemporary platforms, including 10× 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 [7–9]. 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 [10–12]. Correcting these transformations is essential to preserve tissue morphology and establish accurate spatial correspondence, enabling downstream applications, such as cross-sample integration, 3D reconstruction, deconvolution, and spatial domain annotation [8, 13–15]. For example, modelling cell–cell regulatory interactions or perturbation effects [16–18] relies on coherent spatial coordinates to interpret functional relationships. Geometric alignment therefore

constitutes a foundational step for spatially informed biological inference.

Beyond geometric distortion, ST datasets often differ in spatial resolution, gene coverage, and molecular sensitivity across platforms [3, 19]. These discrepancies complicate the integration of measurements across resolutions, from single-cell transcriptomes to multi-cellular spot profiles, as well as across modalities, including 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 [20], and GPSA [21], 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

Received: December 30, 2025. **Revised:** March 2, 2026. **Accepted:** March 24, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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 [22] combine global affine (GA) 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 [23, 24], though methods like SLAT [25] rely on sparse correspondences and have limited scalability. Tissue morphology, e.g. H&E images, can aid alignment [26]; GPSA [21] and PASTE2 [27] incorporate image features but remain spot-level. Cross-modality alignment, such as transcriptomics-to-histology [12, 28], 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 [29], as most methods assume full overlap, with only limited probabilistic handling in PASTE2 [27].

Despite rapid methodological advances in alignments, existing approaches remain fragmented across problem settings. High-resolution frameworks such as CAST [30] often suffer from substantial memory consumption due to dense point-wise matching. Other approaches, including SPACEL [31] and SPIRAL [32], depend heavily on pre-existing cell annotations or cluster-aware information, limiting applicability in fully unsupervised scenarios. Moreover, many methods assume substantial tissue overlap or matched resolutions, reducing robustness in cross-platform or partially overlapping datasets [7, 8, 33]. 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 (**Genetic Algorithm-guided Large Deformation Alignment**), a general and fully automated framework for multimodal spatial alignment. GALA combines a modality-aware rasterization module that projects heterogeneous transcriptomic and histological inputs onto a shared spatial grid. It then employs a coupled optimization strategy to jointly estimates GA 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.

Materials and methods

Overview of the GALA framework

GALA is a unified, landmark-free framework for ST alignment across resolutions and modalities. It jointly corrects large-scale geometric distortions and fine-scale local variations, while flexibly integrating transcriptomic and histological information when available. As illustrated in Fig. 1a, the workflow comprises four components: Input, Multimodal rasterization, 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 (e.g. H&E staining, assumed to be available).

To enable efficient and unified processing, heterogeneous transcriptomic measurements from informative genes and histological images are converted into structured, image-like grids, providing a common representation for efficient joint processing. The coupled alignment process consists of three elements: a GA operator A that captures large-scale linear transformations such as rotation, translation, scaling, and reflection, 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. Figure 1b summarizes the overall objective structure and optimization flow, while Fig. 1c highlights the range of alignment scenarios supported by GALA, including matched- and mismatched-resolution alignment, cross-modality alignment, and partial-overlap cases.

Data preprocessing and rasterization

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. Both the raw and preprocessed datasets, together with access links, are described in Section Data availability and are also provided in the accompanying GALA code package.

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

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

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 Section S1.1). A sensitivity analysis of dx is provided in Supplementary Section S2.1.3. 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 rasterized onto grids with comparable effective resolutions, enabling cross-resolution alignment (e.g. cell-to-spot). Moreover, the rasterized 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 $dx = 30 \sim \mu\text{m}$ for cell-level data, balancing spatial fidelity, and computational efficiency. To stabilize regions with weak or noisy expression, an additional channel recording the total expression across all genes is included, yielding the final transcriptomic tensor $I_{st} \in \mathbb{R}^{(p+1) \times H \times W}$.

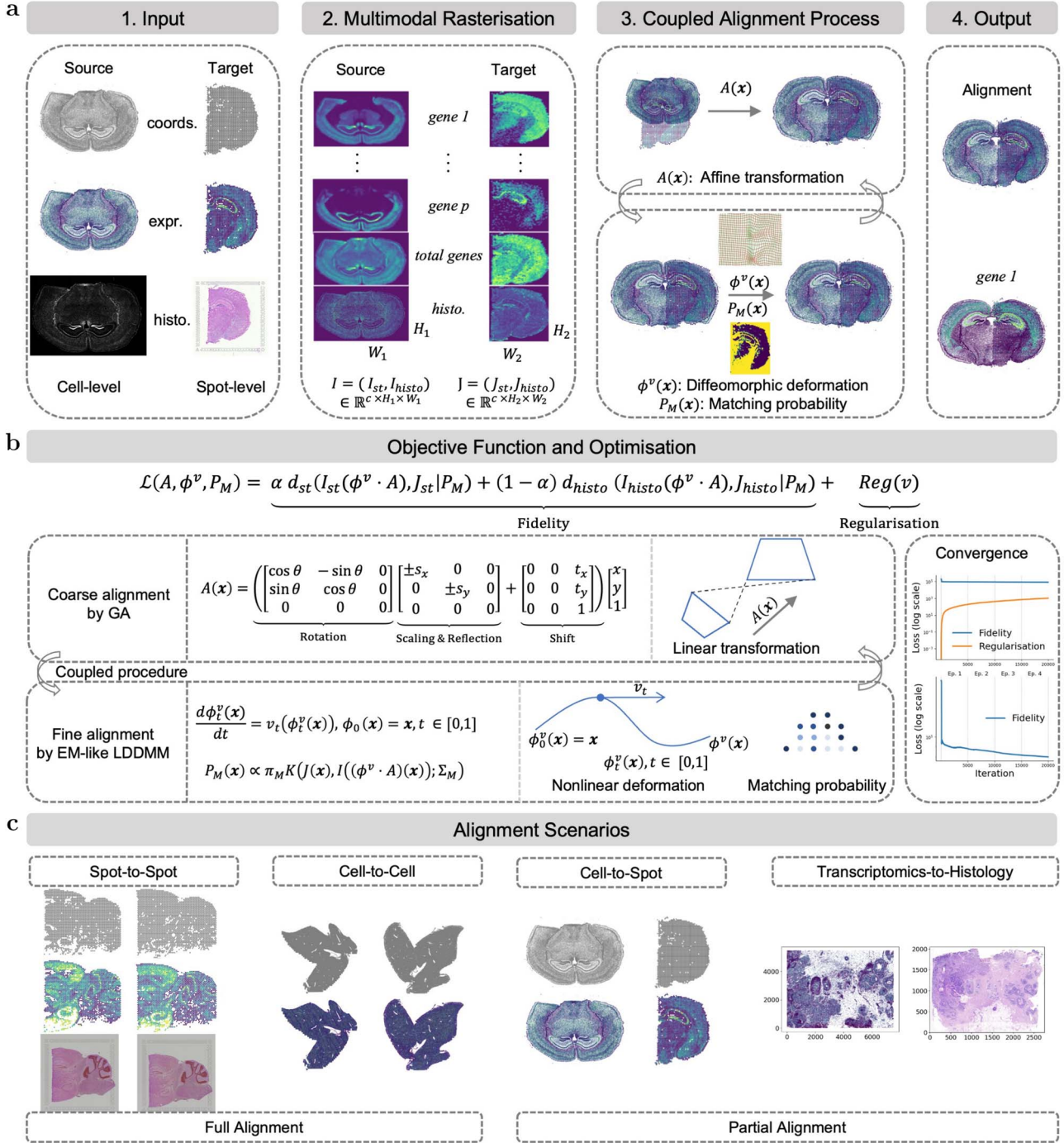


Figure 1 Overview of the GALA framework and its application scenarios. (a) Workflow of GALA. The multimodal source (to be aligned) and target datasets, comprising spatial coordinates (coords.), gene expression (expr.), and histological images (histo.), potentially at different resolutions, are rasterized into multi-channel tensors, $I = (I_{st}, I_{histo})$ and $J = (J_{st}, J_{histo})$, by concatenating the transcriptomic and histology channels. The coupled alignment jointly estimates a GA 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 optimization. The objective combines transcriptomic and histological fidelity terms (weighted by α and $1 - \alpha$) with a regularization on v . Fidelity is measured by a weighted squared error $d(I, J | P_M) = \sum_{x \in \Omega} P_M \|I - J\|^2$, where the matching probability P_M modulates the contribution of each spatial location. 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 regularization across global–local iterations is shown. (The bottom one only focuses the fidelity to better visualize its convergence trend). (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.

When histological images are available, they are rasterized 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 [34]). 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 optimization is performed.

Alignment objective and optimization

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

Building on the GA alignment that estimates the overlapping region between source and target, GALA restricts subsequent refinement to this overlap domain. Within the estimated overlapping region, a probabilistic matching model is introduced to distinguish matched (reliably aligned) areas from unmatched (background, missing, or artefactual) areas. Specifically, for each grid location $\mathbf{x} \in \Omega$, a latent label indicates whether it belongs to the matched or unmatched class. Let $I(\mathbf{x}), J(\mathbf{x}) \in \mathbb{R}^c$ denote the raster intensities of source and target at location $\mathbf{x}$. Under the composite transformation $(\phi^v \cdot A)(\mathbf{x})$, we model the observed target intensity as

$$J(\mathbf{x}) = \begin{cases} I((\phi^v \cdot A)(\mathbf{x})) + \epsilon_M, & \mathbf{x} \text{ is matched,} \\ \mu_N + \epsilon_N, & \mathbf{x} \text{ is unmatched,} \end{cases}$$

where the residuals ϵ are defined as deviations from their respective mean intensities. To quantify similarity, we evaluate these residuals using Gaussian kernels $K(J, \mu, \Sigma) = \exp(-\frac{1}{2}(J - \mu)^\top \Sigma^{-1}(J - \mu))$ with mean μ and covariance Σ . μ_N denotes the mean intensity of unmatched areas, estimated iteratively. This Gaussian form induces a quadratic data term and enables closed-form Expectation–Maximization (EM) updates [12, 35]. Therefore, given class priors π_M and π_N ($\pi_M + \pi_N = 1$), the posterior probability that location $\mathbf{x}$ belongs to the matched class is computed as

$$P_M(\mathbf{x}) = \frac{\pi_M K(J(\mathbf{x}), I((\phi^v \cdot A)(\mathbf{x})), \Sigma_M)}{\pi_M K(J(\mathbf{x}), I((\phi^v \cdot A)(\mathbf{x})), \Sigma_M) + \pi_N K(J(\mathbf{x}), \mu_N, \Sigma_N)},$$

where $\Sigma_M = \sigma_M^2 I_c$ and $\Sigma_N = \sigma_N^2 I_c$ are the covariance matrices of ϵ_M and ϵ_N , respectively, I_c is the $c \times c$ identity matrix. A high value of

$P_M(\mathbf{x})$ indicates strong correspondence, whereas a low value suggests background or artefactual areas.

To incorporate the probabilistic matching model into the alignment procedure, the posterior probability $P_M(\mathbf{x})$ is used to weight the contribution of each spatial location in the optimization. The multimodal alignment objective combines transcriptomic and histological fidelity terms (weighted by α and $1 - \alpha$) with a regularization on the deformation field v :

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

where σ_{M_1} , σ_{M_2} , and σ_R control feature matching and deformation smoothness. The first two terms minimize transcriptomic and morphological discrepancies between the iteratively aligned source and target, weighted by the matching probability $P_M(\mathbf{x})$. The final term regularizes v_t in a reproducing kernel Hilbert space, ensuring smooth and invertible mappings [36]. The selected $p + 1$ gene channels and histological image channel act as alignment ‘landmarks’, guiding the optimization—principally differing from STalign, which does not utilize this information. Additional details are provided in Supplementary Section S1.2.

We adopt a coupled optimization strategy that jointly refines GA alignment and local diffeomorphic deformation within a unified objective. In the global stage, a genetic algorithm (GA) [37] conducts a population-based search to estimate affine parameters and overlapping regions, providing robust initialization that avoids local minima. Subsequently, the local stage resolves residual nonlinear discrepancies using large deformation diffeomorphic metric mapping (LDDMM) [36], modelling smooth, invertible deformations suitable for fine-grained spatial adjustments. To reconcile local deformations with modality-specific information, an EM-like LDDMM scheme is implemented: in the E-step, matching probabilities P_M are updated from current residual kernels; in the M-step, the velocity field v_t is optimized by minimizing the weighted quadratic registration energy under the LDDMM framework, while μ_N is updated accordingly. The pseudo-code of GALA is provided in Supplementary Algorithm S1.

Unlike conventional pipelines that perform global (coarse) and local (fine) alignment independently, our coordinated coarse-to-fine scheme iteratively reinforces both stages. 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. A full mathematical formulation and optimization procedure are provided in Supplementary Section 1.3. Empirically, under default parameter settings (Supplementary Table S1), only two global–local episodes suffice to achieve stable alignment across most datasets (Supplementary Fig. S3).

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 [21], which ranks genes by their predictability from local spatial context. Specifically, for each gene, a k -nearest neighbour (k NN) 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 [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. Notably, GALA does not require these informative genes to provide full anatomical coverage or be spatially complementary, as the inclusion of a total expression channel provides a consistent global structural scaffold. The balance parameter α was observed to yield stable performance across its range (0, 1); we set $\alpha = 0.5$ when histological image data are available, and $\alpha = 1$ otherwise. Detailed robustness and sensitivity analyses, as well as spatial coverage analysis of informative genes, are provided in [Supplementary Sections S2.1.1, S2.1.2, and S2.1.4](#).

Evaluation metrics

To evaluate alignment performance across diverse experimental scenarios, we employ complementary metrics capturing spatial accuracy, biological coherence, and molecular consistency. Metrics are selected based on dataset characteristics, such as availability of anatomical annotations, spatial resolution, and modality, to ensure biologically meaningful and context-appropriate assessment.

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 and, more specifically, the preservation of biologically defined regions [8]. Second, biological coherence is evaluated via clustering-based metrics: a joint Gaussian mixture module (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 adjusted Rand index (ARI) and normalized mutual information (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 quantify preservation of continuous spatial gene expression gradients [40, 41]. For each gene, correlations between transcriptomic profiles in aligned source and target slices are computed with Gaussian-kernel weights based on inter-spot distances. Averaging across genes provides a global measure of spatial expression consistency, capturing local gradient preservation without requiring labels. Further details on evaluation metrics are provided in [Supplementary Section S1.4](#).

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 stabilize 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 mean absolute error (MAE) of annotated landmark coordinates:

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

where $i = 1, \dots, n$ indexes landmarks, and $\mathbf{x}_i, \mathbf{x}'_i \in \mathbb{R}^2$ denote the coordinates of the i th landmark in the target and aligned source data, respectively. Lower MAE indicates better spatial correspondence and provides an evaluation that is independent of gene expression or clustering assumptions, making it compatible with landmark-free alignment methods.

Overall, each metric is selected to match the characteristics of the corresponding dataset and alignment task, capturing structural correspondence (label consistency, MAE), transcriptional coherence (ARI, NMI, and pseudo-spot similarity), or preservation of spatial expression gradients (cross-correlation). These quantitative measures are complemented by visual overlays of aligned slices, providing intuitive verification of spot/cell correspondence, gene expression patterns, and regional morphology. Together, they offer a comprehensive assessment of GALA's performance across diverse ST datasets.

Implementation techniques for all alignment scenarios

GALA is a versatile alignment framework applicable to widespread application scenarios (see [Fig. 1c](#)). 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. In the following, we detail how GALA is applied across these scenarios.

First, because rasterization 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 Data preprocessing and rasterization). No additional algorithmic modifications are required: once rasterized, all datasets are processed identically by the optimization 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 ST 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 rasterization tensors, allowing both to be treated as image-like modalities. Alignment is then performed under the multimodal objective:

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

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 GA 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 Alignment objective and optimization). This mechanism confers robustness in scenarios involving incomplete, distorted, or physically damaged tissue.

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

Results

We systematically evaluate GALA across a wide range of ST alignment tasks (Fig. 1c) and compare its performance with state-of-the-art methods using established quantitative metrics (Section Evaluation metrics). Because existing alignment methods are typically designed for specific settings, the set of baseline methods differs across alignment scenarios.

Spot-to-spot alignment within spatial transcriptomics technologies

Human dorsolateral prefrontal cortex data by Visium

The first example used to evaluate GALA is the human dorsolateral prefrontal cortex (DLPFC) ST data generated with the 10× Genomics Visium platform [29]. 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 3000–4000 spots and ~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, 20, 42, 43]. 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, ARI, and NMI to quantitatively assess biological coherence after alignment (see Section Evaluation metrics).

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 GMM analysis yielding a mean ARI of 0.396 and NMI of 0.508, outperforming STAligner (0.385 and 0.491, respectively; see Supplementary Table S5 and Fig. S10a 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. S10b and 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*, and *MOBP*) exhibit increased overlap after alignment, indicating enhanced layer preservation and improved biological coherence. Using pairwise alignment, 3D visualization was performed by setting slice 1 as the template and aligning other slices to a common coordinate system, shown in Fig. 2e. Rather than a dedicated 3D reconstruction method, GALA achieves 3D structural consistency, i.e. comparable to specialized 3D integration approaches, including STAligner, GPSA, and ST-GEARS. In particular, cross-slice layer transitions and anatomical boundaries remain well preserved across adjacent sections. Quantitative pairwise label consistency results further support this observation, showing that GALA (0.847) performs competitively and differs only marginally from GPSA (0.850), consistent with the subtle visual differences observed in the 3D rendering.

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. 3a (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 Selection of informative genes and hyper-parameters, the fusion weight α is set to 0.5 by default when histology is available. Additional experiments (line chart in Fig. 3a) 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. Figure 3b shows label consistency accuracy for raw data, PASTE2, and GALA when the left section is retained; results for the bottom section are in Supplementary Fig. S11a. 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. S11c) confirms robustness consistent with Fig. 3a.

Finally, we illustrated why GALA performs well in partial alignment by visualizing learned parameters for a 60% overlap (first type of partial overlap; second type shown in Supplementary Fig. S7b). The left panel of Fig. 3c and Supplementary Fig. S11b show the GA 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

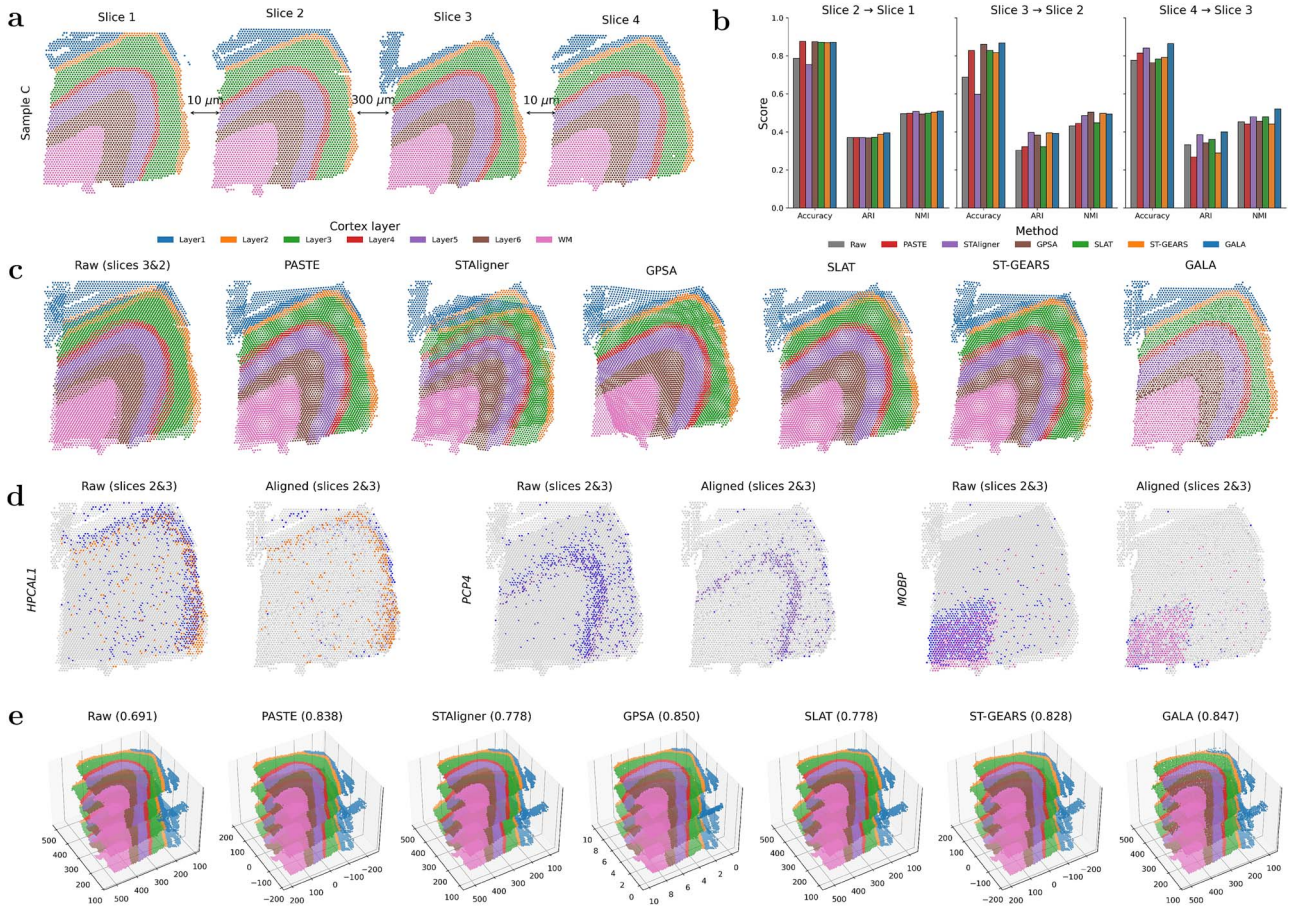


Figure 2 Alignment evaluation on Sample C of DLPFC data across GALA and baselines. (a) Four raw tissue slices from Sample C, with adjacent distances of 10 μm , 300 μm , and 10 μm . Spots are annotated with six neocortical layers (Layers 1–6) and WM. (b) Quantitative comparison of alignment performance for three pairwise slice alignments (2 $\rightarrow$ 1, 3 $\rightarrow$ 2, 4 $\rightarrow$ 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'). The differently sized white background dots result from the stacked two regular grids following affine transformation, generating a Moiré-like pattern. (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 highlighted, while low-expression spots are shown in grey. In the source and aligned source slices, high-expression spots are displayed according to their annotated layer in (a), while low-expression spots are shown in grey. (e) 3D visualization of all four slices reconstructed via sequential pairwise alignment using GALA and baselines. The accuracy values shown in parentheses after each method denote the mean pairwise label consistency accuracy.

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 GA transformation and local diffeomorphic mapping are iteratively optimized under the objective function, explaining GALA's effectiveness for partial ST alignment.

Mouse brain sagittal-posterior sections by Visium

The second example of spot-to-spot alignment involves mouse brain sagittal-posterior (MBSP) sections produced by 10 $\times$ Genomics Visium. The two sections contain 2805 and 2840 spots, respectively, each profiled for 32 285 genes, and are accompanied by H&E-stained histological images (left panel of Fig. 4a).

We first examined the role of spatially informative genes in GALA, defined as genes ranked by the R^2 -based strategy from GPSA [21], which orders genes according to how well their expression can be

predicted from local spatial context (see Section Selection of informative genes and hyper-parameters). The top three genes automatically selected were *Pcp2*, *Mbp*, and *Nrgn*. These genes exhibit complementary anatomical specificity (Fig. 4a, right panels): *Pcp2* is restricted to cerebellar Purkinje cells and selected cortical layers; *Mbp* marks oligodendrocyte-enriched WM tracts; *Nrgn* is expressed in excitatory neurons within cortical and hippocampal layers. Their distinct localization 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 (see Section Evaluation metrics). 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. 4b 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 ,

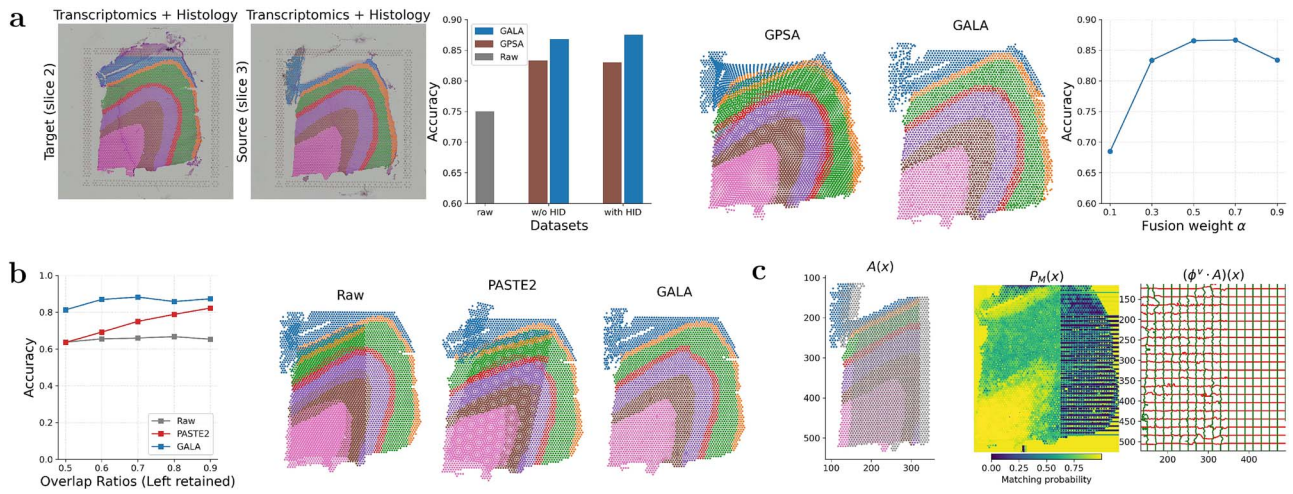


Figure 3 Multimodal alignment and robustness to partial overlap using GALA. Spots are coloured according to their annotated layer in Fig. 2(a). (a) Comparison of alignment between slices 3 and 2 when histological image data are included, across GPSA and GALA. The rightmost panel shows sensitivity to the relative weight α . 'w/o HID' indicates no histology. (b) 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%. (c) Illustration of the GALA mechanism for 'left 60% retained': GA transformation $A(x)$, matching probabilities $P_M(x)$, and local diffeomorphic mapping $(\phi^V \cdot A)(x)$.

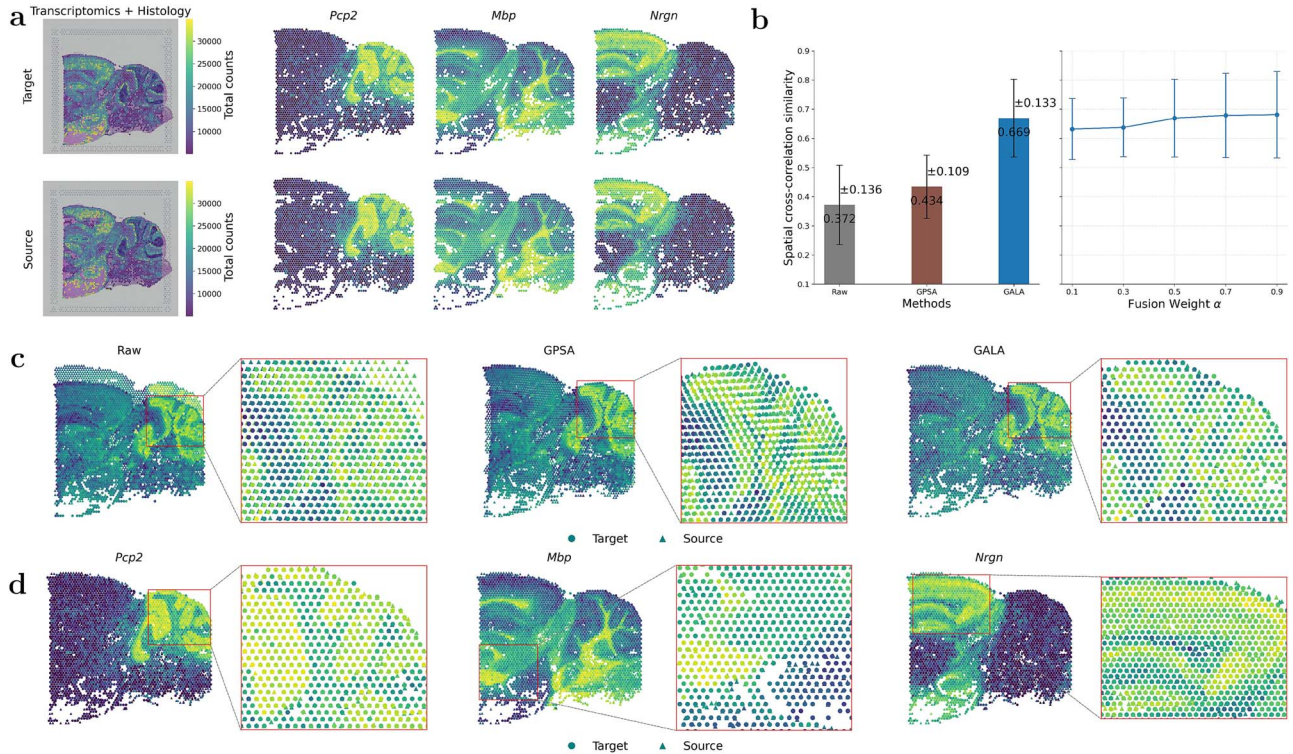


Figure 4 Alignment of two MBSP sections. (a) Raw slices overlaid on H&E images and spatial expression of three top informative genes (*Pcp2*, *Mbp*, and *Nrgn*). Brighter spots indicate higher gene expression. (b) Left: mean spatial cross-similarity scores ($\pm$ s.d.) across the top 135 informative genes for raw slices (Raw) and slices aligned by GPSA and GALA ($\alpha = 0.5$). Right: cross-similarity $\pm$ SD 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.

an improvement of over 23% relative to GPSA (0.434 ± 0.109). Visual comparison of stacked target and aligned source slices (Fig. 4c) 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. 4d). Compared with their raw, unaligned spatial patterns (see Supplementary Fig. S12), all three genes display markedly improved localization 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

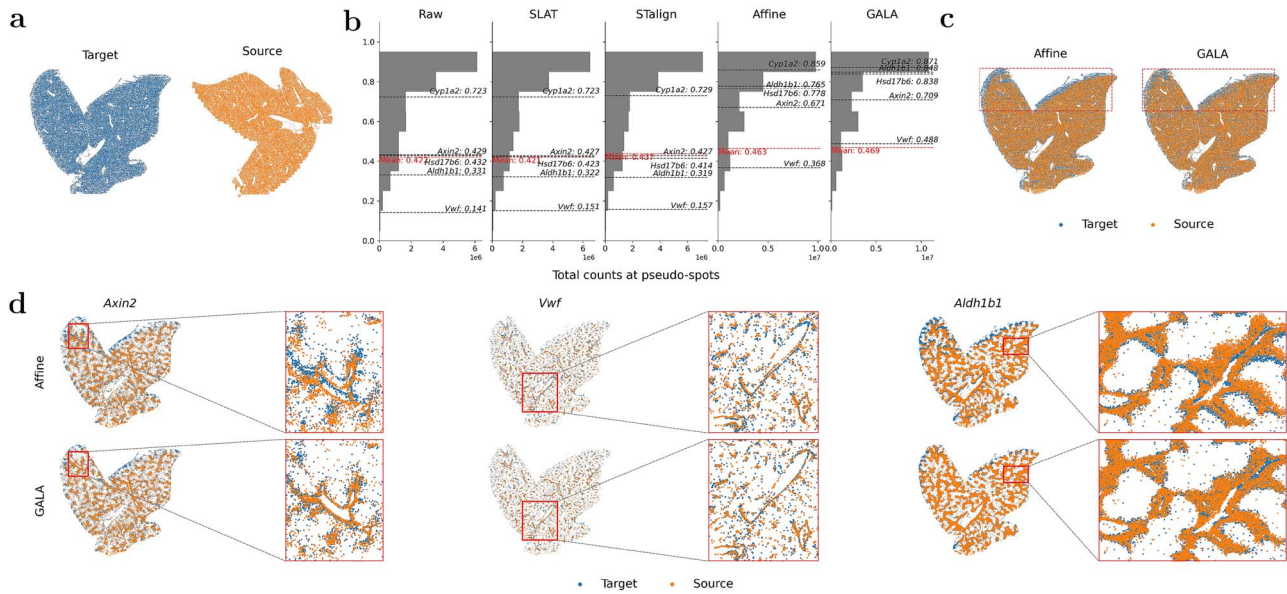


Figure 5 Alignment of two MERFISH mouse liver replicates. (a) Raw slices for source and target data are shown for visualization. Scaled expression slices are shown in [Supplementary Fig. 13a](#). (b) Distribution of cosine similarity scores for 134 liver marker genes and 3 informative genes against the total gene expression at pseudo-spots (aggregated into 30 μm) after alignment by Raw, SLAT, STalign, Affine, and GALA. Dashed lines indicate the mean and representative genes. (c) Visual comparison of whole-slice alignment results by Affine and GALA. Cells in target and aligned source data are distinguished by their origin (source or target). The top area highlighted by a rectangular box indicates 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.

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. 4b](#).

Cell-to-cell alignment within spatial transcriptomics 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 $\sim 300\,000$ cells with expression profiles for 347 genes, serving as target and source data, respectively. [Figure 5a](#) 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. S13a](#)), 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 [44] and top-ranked informative genes. Five representative genes, including canonical markers (*Axin2* [45] and *Vwf* [46]) and top informative genes (*Aldh1b1*, *Hsd17b6*, and *Cyp12a*), were analysed in detail.

First, we quantitatively assessed alignment across all baselines and GALA. [Figure 5b](#) 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 mean 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 mean similarity to 0.463, and the full GALA framework further raised it to 0.469, demonstrating that both GA and local diffeomorphic components contribute to alignment. Representative gene trends exemplify this: raw alignment produced scores of 0.429 (*Axin2*), 0.141 (*Vwf*), 0.331 (*Aldh1b1*), 0.432 (*Hsd17b6*), and 0.723 (*Cyp12a*); Affine improved these to 0.671, 0.368, 0.765, 0.778, and 0.859; full GALA further increased them to 0.709, 0.488, 0.848, 0.838, and 0.871.

Second, we visually compared alignment results for Affine and GALA. The right panels of [Fig. 5c](#) 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. [Figure 5d](#) 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. S13b, c](#) illustrates the corresponding gene expression values.)

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

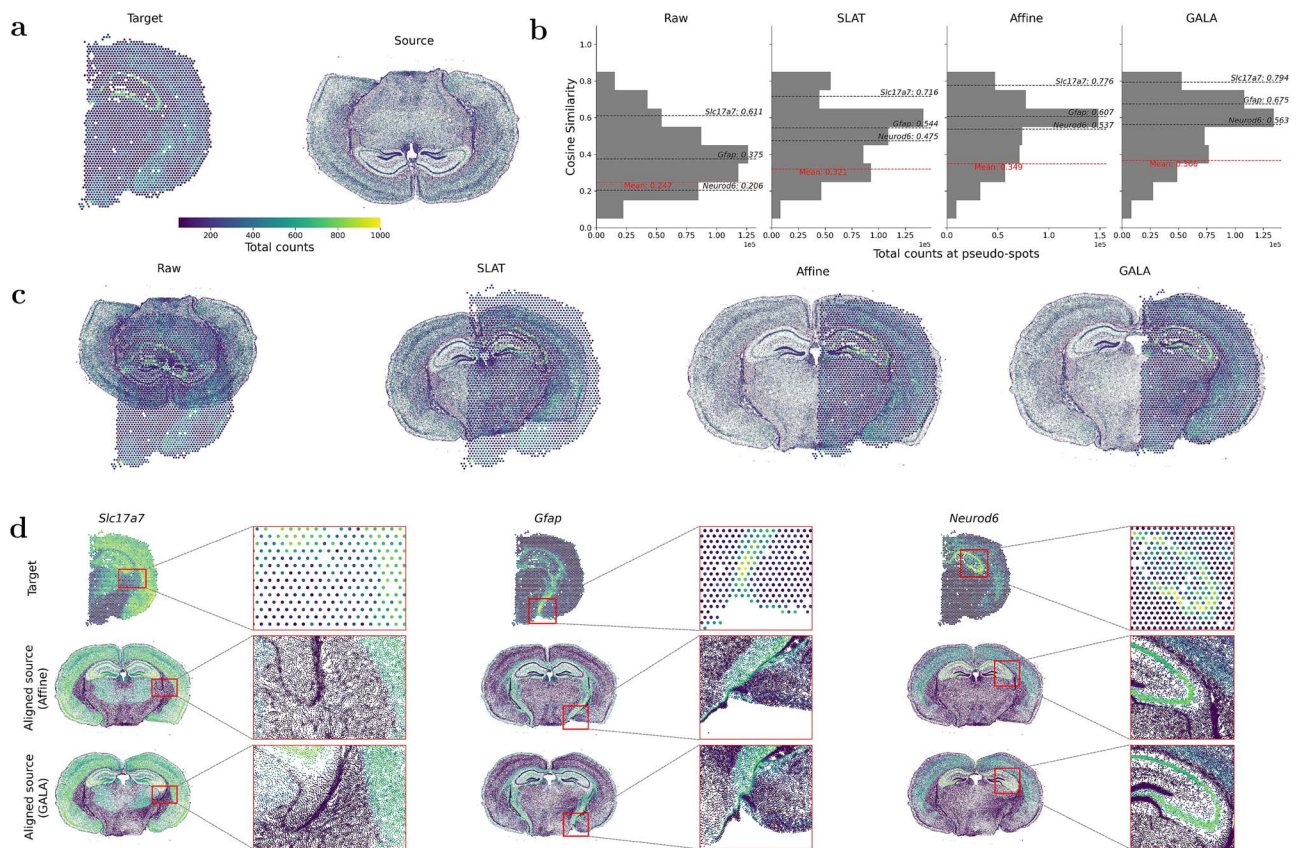


Figure 6 Alignment of mouse coronal sections from Visium and Xenium. (a) Visualization of Visium (spot-level, target) and Xenium (cell-level, source) slices. Brighter colours indicate higher gene expression. (b) Distribution of cosine similarities between gene expression in Visium target versus aligned Xenium source at matched spots/pseudo-spots. Dashed lines represent the mean similarity and three informative genes, respectively. (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 highlighted by rectangles.

Cell-to-spot alignment across spatial transcriptomics 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 [20, 24, 47]. 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 2634 spots with low-resolution expression profiles. As shown in Fig. 6a, 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. 6b). Mean 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). Figure 6b demonstrates that, compared with raw similarities of 0.611, 0.375, and 0.206; GALA increased these to 0.794, 0.675, and 0.563; outperforming Affine-aligned results of 0.776, 0.607, and 0.537; and SLAT-aligned results of 0.716, 0.544, and 0.475.

We further visually evaluated alignment across methods. As shown in Fig. 6c, SLAT performs poorly in this cross-resolution setting, primarily due to its reliance on one-to-one adversarial graph matching followed by a GA 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. S14 provides additional comparisons, showing target data as registered H&E images. Figure 6d 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 GA and local diffeomorphic refinement contribute substantially to the

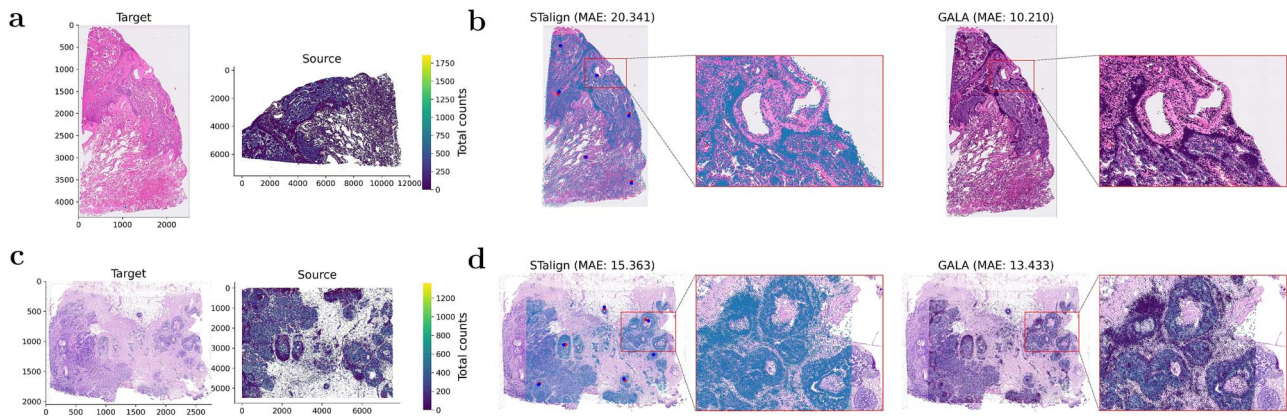


Figure 7 Alignment of Xenium single-cell RNA data to paired H&E images. (a) H&E histological image (target) and paired Xenium single-cell data (source) from a human lung cancer section, where brighter colours indicate higher total gene expression. (b) Alignment results of (a) by landmark-based STalign and landmark-free GALA, with MAE indicated. Dots in the STalign panel denote manually annotated landmarks in source/target data, respectively. As STalign ignores gene expression, its aligned data are visualized as an overlay on the target H&E image. Adjacent zoom-in panels highlight regions marked by rectangles. (c) H&E image (target) and paired single-cell 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 highlighted by rectangles.

improved alignment. These conclusions are consistent with observations in the cell-to-cell scenario in Section Cell-to-cell alignment within ST technologies.

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, 28]. 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.

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). Figure 7a and c shows 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. 7c 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 MAE of landmark coordinates to quantitatively assess alignment (see Section Evaluation metrics). Figure 7b shows the lung cancer dataset results: using six representative landmarks (compared with the four landmarks used for pre-alignment in STalign), STalign achieved an MAE of 20.341, whereas GALA achieved a more precise alignment with MAE of 10.210. Figure 7d 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. S15). In contrast, GALA does not rely on manual landmarks during optimization, eliminating user bias, and ensuring fully data-driven alignment. Landmarks are used only for independent evaluation, with a larger set of points than those provided to STalign, allowing an objective and fair comparison of geometric consistency across methods. This separation between optimization and evaluation allows GALA to remain fully automated and reproducible, while still enabling an objective assessment of alignment quality.

Computational efficiency

All methods were benchmarked under consistent hardware conditions: if a baseline supports GPU computation, it was run on the same GPU as GALA; otherwise, it was executed on the same CPU. Details of the operating system and device configuration are provided in Supplementary Section S2.3. Across multiple benchmarks, GALA demonstrated superior computational efficiency, achieving both faster runtime and substantially lower memory usage compared with existing methods. It consistently outperformed other nonlinear methods, being substantially faster than GPSA and ST-GEARS (averaging 135.22 s compared with 245.61 and 154.69 s, respectively, on slices 2 and 3 of Sample C), while also maintaining a minimal memory footprint. While specialized rigid methods like PASTE, STaligner, and SLAT were faster (all under 30 s), they required substantially more memory and 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 s with minimal memory usage, compared with STalign's 512.06 s and substantially higher memory consumption, due to GALA's streamlined optimization and reduced parameters (see Section Alignment objective and optimization).

GALA also 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, low memory usage, high-fidelity alignment of large-scale ST data. The details of comparison of computational efficiency with baseline methods are provided at [Supplementary Table S4](#).

Discussion and Conclusion

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 multi-modal rasterization strategy, which converts diverse transcriptomic and histological data into co-registered raster tensors. This standardized representation harmonizes 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 optimization couples a GA transformation, which identifies overlapping regions for initialization, with an EM-like LDDMM refinement capable of large-scale deformations. The EM-style probabilistic matching focuses alignment on confidently corresponding regions while downweighting background or artefacts, making the coupled strategy robust across resolutions, modalities, and partial tissue overlaps. Furthermore, the invertibility of the diffeomorphic transformation ensures directionally consistent alignment. In within-resolution tasks, source-to-target and target-to-source alignments yield highly similar performance ([Supplementary Section S2.2.1](#), [Fig. S8](#)). For cross-resolution or cross-modality alignment, the same property allows the high-resolution modality to be stably aligned to a lower-resolution reference, while the inverse transformation naturally provides the corresponding reverse mapping ([Supplementary Fig. S16](#)).

Compared with existing approaches ([Table 1](#)), GALA offers broader applicability and improved robustness. Spot-level methods such as PASTE and STaligner emphasize 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 specialized 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. 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.

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 [48]. Morphologically similar cells may exhibit distinct molecular profiles, and sharp molecular boundaries may lack histological counterparts [47]. Such discordances can introduce noise, highlighting the need for adaptive, biologically informed fusion strategies that weight each

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	✓	-	✓	✓	-	-	✓	✓	✓	-	✓	-	✓	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.

modality based on local context [3]. Furthermore, GALA focuses on geometric spatial alignment and does not explicitly correct gene expression batch effects. In datasets with substantial donor variability or platform differences, additional downstream integration methods may be required after alignment to harmonize molecular signals. Finally, as with all LDDMM-based frameworks, GALA assumes smooth, diffeomorphic deformations. Extremely irregular, discontinuous, or topology-disrupting distortions may violate this assumption and thus fall outside the model's representational capacity, leading to alignment failure (Supplementary Section S2.2.2).

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. For example, GALA ensures spatial correspondence across samples, while network-based models (e.g. Perturb-STNet [17]) capture cell-cell interactions and perturbation effects within the aligned coordinate system, enabling integrative analyses that combine geometric alignment with functional modelling. Beyond transcriptomics, GALA is readily extendable to other spatially resolved omics modalities, such as spatial proteomics or metabolomics [49, 50]. By aggregating molecular measurements onto a common spatial grid through rasterization, multi-omics data can be aligned within a unified framework. These representations may further be embedded into shared latent spaces using representation learning approaches, facilitating cross-modality comparison of spatial patterns while preserving modality-specific variation. Future work will explore joint multimodal alignment while addressing cross-modality consistency and heterogeneous noise. Taken together, integrating precise spatial alignment with downstream biological modelling represents 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.

Key points

- GALA (Genetic Algorithm-guided Large Deformation Alignment) targets spatial alignment across heterogeneous spatial transcriptomics (ST) datasets, supporting practical application scenarios including cross-platform and cross-resolution alignment, cross-modality integration, and partial tissue alignment.
- GALA implements a fully automated, landmark-free alignment framework that jointly models both global affine transformation and local diffeomorphic deformation within a unified multimodal representation.
- GALA demonstrates robust and accurate alignment across diverse alignment scenarios encountered in real-world ST applications, while scaling to large tissue sections and facilitating downstream integrative analyses.

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.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

None declared.

Funding

This work was supported by the High-Performance Computing (HPC) platform at ShanghaiTech University.

Code and data 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.

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://zenodo.org/records/18803459>) to allow full reproduction of the analyses and figures presented.

References

1. Asp M, Giacomello S, Larsson L *et al*. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* 2019;**179**:1647–1660.e19. <https://doi.org/10.1016/j.cell.2019.11.025>
2. Marx V. Method of the year: spatially resolved transcriptomics. *Nat Methods* 2021;**18**:9–14. <https://doi.org/10.1038/s41592-020-01033-y>
3. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 2022;**19**:534–46. <https://doi.org/10.1038/s41592-022-01409-2>
4. Ståhl PL, Salmen F, Vickovic S *et al*. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82. <https://doi.org/10.1126/science.aaf2403>
5. Janesick A, Shelansky R, Gottscho A *et al*. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. *Nat Commun* 2023;**14**:8353. <https://doi.org/10.1038/s41467-023-43458-x>
6. Chen KH, Boettiger AN, Moffitt JR *et al*. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;**348**:aaa6090. <https://doi.org/10.1126/science.aaa6090>
7. Khan M, Arslanturk S, Draghici S. A comprehensive review of spatial transcriptomics data alignment and integration. *Nucleic Acids Res* 2025;**53**:gkaf536. <https://doi.org/10.1093/nar/gkaf536>
8. Liu Y, Yang C. Computational methods for alignment and integration of spatially resolved transcriptomics data. *Comput Struct Biotechnol J* 2024;**23**:1094–105. <https://doi.org/10.1016/j.csbj.2024.03.002>
9. Lee J, Yoo M, Choi J. Recent advances in spatially resolved transcriptomics: challenges and opportunities. *BMB Rep* 2022;**55**:113–24. <https://doi.org/10.5483/BMBRep.2022.55.3.014>
10. Tward D, Li X, Huo B *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*, 162–73. Cham, Switzerland: Springer, 2019, https://doi.org/10.1007/978-3-030-33226-6_18.
11. Ekvall M, Bergenstråhle L, Andersson A *et al*. Spatial landmark detection and tissue registration with deep learning. *Nat Methods* 2024;**21**:673–9. <https://doi.org/10.1038/s41592-024-02199-5>
12. Clifton K, Anant M, Aihara G *et al*. STalign: alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat Commun* 2023;**14**:8123.
13. Zeira R, Land M, Strzalkowski A *et al*. Alignment and integration of spatial transcriptomics data. *Nat Methods* 2022;**19**:567–75. <https://doi.org/10.1038/s41592-022-01459-6>
14. Biancalani T, Scalia G, Buffoni L *et al*. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods* 2021;**18**:1352–62. <https://doi.org/10.1038/s41592-021-01264-7>
15. Kang L, Zhang Q, Qian F *et al*. Benchmarking computational methods for detecting spatial domains and domain-specific spatially variable genes from spatial transcriptomics data. *Nucleic Acids Res* 2025;**53**:gkaf303. <https://doi.org/10.1093/nar/gkaf303>
16. Cang Z, Zhao Y, Almet AA *et al*. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* 2023;**20**:218–28. <https://doi.org/10.1038/s41592-022-01728-4>
17. Egbon OA, Hickey JW, Anchang B. Fusion of spatiotemporal and network models to prioritize multiscale effects in single-cell perturbations. *Brief Bioinform* 2025;**26**:bbaf277. <https://doi.org/10.1093/bib/bbaf277>
18. Cesaro G, Nagai JS, Gnoato N *et al*. Advances and challenges in cell–cell communication inference: a comprehensive review of tools, resources, and future directions. *Brief Bioinform* 2025;**26**:bbaf280.
19. Chen T-Y, You L, Hardillo JAU *et al*. Spatial transcriptomic technologies. *Cells*. 2023;**12**:2042. <https://doi.org/10.3390/cells12162042>
20. Zhou X, Dong K, Zhang S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nat Comput Sci* 2023;**3**:894–906. <https://doi.org/10.1038/s43588-023-00528-w>
21. Andrew Jones F, Townes W, Li D *et al*. Alignment of spatial genomics data using deep gaussian processes. *Nat Methods* 2023;**20**:1379–87. <https://doi.org/10.1038/s41592-023-01972-2>
22. Xia T, Luni H, Zuo L *et al*. ST-GEARS: advancing 3d downstream research through accurate spatial information recovery. *Nat Commun* 2024;**15**:7806. <https://doi.org/10.1038/s41467-024-51935-0>
23. Zhou Z, Zhong Y, Zhang Z *et al*. Spatial transcriptomics deconvolution at single-cell resolution using redeconve. *Nat Commun* 2023;**14**:7930. <https://doi.org/10.1038/s41467-023-43600-9>
24. Wan X, Xiao J, Tam SST *et al*. Integrating spatial and single-cell transcriptomics data using deep generative models with spatialscope. *Nat Commun* 2023;**14**:7848. <https://doi.org/10.1038/s41467-023-43629-w>
25. Xia C-R, Cao Z-J, Xin-Ming T *et al*. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nat Commun* 2023;**14**:7236. <https://doi.org/10.1038/s41467-023-43105-5>
26. Jian H, Schroeder A, Coleman K *et al*. Statistical and machine learning methods for spatially resolved transcriptomics with histology. *Comput Struct Biotechnol J* 2021;**19**:3829–41. <https://doi.org/10.1016/j.csbj.2021.06.052>
27. Liu X, Zeira R, Raphael BJ. PASTE2: partial alignment of multi-slice spatially resolved transcriptomics data. *Genome Research* 2023;**33**:1124–32. <https://doi.org/10.1101/gr.277670.123>
28. Shi Z, Zhu F, Min W. Inferring multi-slice spatially resolved gene expression from H&E-stained histology images with STMCL. *Methods* 2025;**234**:187–95. <https://doi.org/10.1016/j.ymeth.2024.11.016>
29. Maynard KR, Collado-Torres L, Weber LM *et al*. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 2021;**24**:425–36. <https://doi.org/10.1038/s41593-020-00787-0>
30. Tang Z, Shuchen Luo H, Zeng JH *et al*. Search and match across spatial omics samples at single-cell resolution. *Nat Methods* 2024;**21**:1818–29. <https://doi.org/10.1038/s41592-024-02410-7>
31. Hao X, Wang S, Fang M *et al*. SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nat Commun* 2023;**14**:7603.
32. Guo T, Yuan Z, Pan Y *et al*. SPIRAL: integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biol* 2023;**24**:241.
33. Yunfei H, Xie M, Li Y *et al*. Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. *Genome Biol* 2024;**25**:212. <https://doi.org/10.1186/s13059-024-03361-0>
34. 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*, pp. 31–5. Piscataway, NJ, USA: IEEE, 2009. <https://doi.org/10.1109/IASP.2009.5054605>
35. Tward D, Brown T, Kageyama Y *et al*. Diffeomorphic registration with intensity transformation and missing data: application to 3D digital pathology of alzheimer's disease. *Front Neurosci* 2020;**14**:52. <https://doi.org/10.3389/fnins.2020.00052>
36. Faisal Beg M, Miller MI, Trounev A *et al*. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vis* 2005;**61**:139–57. <https://doi.org/10.1023/B:VISI.0000043755.93987.a>
37. Ahmed Fawzy Gad. PyGAD: an intuitive genetic algorithm python library. *Multimed Tools Appl* 2024;**83**:58029–42.

38. Edsgård D, Johnsson P, Sandberg R. Identification of spatial expression trends in single-cell gene expression data. *Nat Methods* 2018;**15**:339–42. <https://doi.org/10.1038/nmeth.4634>
39. Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods* 2020;**17**:193–200. <https://doi.org/10.1038/s41592-019-0701-7>
40. Chen Y. A new methodology of spatial cross-correlation analysis. *PLoS One* 2015;**10**:e0126158. <https://doi.org/10.1371/journal.pone.0126158>
41. Tan Y, Wang Z, Wang A *et al.* Integrative cross-sample alignment and spatially differential gene analysis for spatial transcriptomics. *bioRxiv*. 2025;2025. <https://doi.org/10.1101/2025.06.05.653933>
42. Li H, Lin Y, He W *et al.* SANTO: a coarse-to-fine alignment and stitching method for spatial omics. *Nat Commun* 2024;**15**:6048. <https://doi.org/10.1038/s41467-024-50308-x>
43. Shi X, Yang Y, Ma X *et al.* Probabilistic cell/domain-type assignment of spatial transcriptomics data with spatialanno. *Nucleic Acids Res* 2023;**51**:e115. <https://doi.org/10.1093/nar/gkad1023>
44. Yanfang L, Liu M, Yang J *et al.* Spatial transcriptome profiling by MERFISH reveals fetal liver hematopoietic stem cell niche architecture. *Cell Discovery* 2021;**7**:47. <https://doi.org/10.1038/s41421-021-00266-1>
45. Sun T, Pikiolek M, Orsini V *et al.* AXIN2+ pericentral hepatocytes have limited contributions to liver homeostasis and regeneration. *Cell Stem Cell* 2020;**26**:97–107.e6. <https://doi.org/10.1016/j.stem.2019.10.011>
46. Nightingale T, Cutler D. The secretion of von Willebrand factor from endothelial cells; an increasingly complicated story. *J Thromb Haemost* 2013;**11**:192–201. <https://doi.org/10.1111/jth.12225>
47. Kleshchevnikov V, Shmatko A, Dann E *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* 2022;**40**:661–71. <https://doi.org/10.1038/s41587-021-01139-4>
48. Chelebian E, Avenel C, Wählby C. Combining spatial transcriptomics with tissue morphology. *Nat Commun* 2025;**16**:1–13. <https://doi.org/10.1038/s41467-025-58989-8>
49. Tran T, Wee F, Joseph CR *et al.* An integrated approach for analyzing spatially resolved multi-omics datasets from the same tissue section. *Front Mol Biosci* 2025;**12**:1614288. <https://doi.org/10.3389/fmolb.2025.1614288>
50. Tian R, Xue Z, Chen Y *et al.* Integrating cross-sample and cross-modal data for spatial transcriptomics and metabolomics with spatialmeta. *Nat Commun* 2025;**16**:8855. <https://doi.org/10.1038/s41467-025-63915-z>