Spatial-linked alignment tool (SLAT) for aligning heterogenous slices properly

Chen-Rui Xia[1,2†], Zhi-Jie Cao[1,2†*], Xin-Ming Tu[1,#], and Ge Gao[1,2*]

[1] State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Biomedical Pioneering Innovative Center (BIOPIC) and Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), Peking University, Beijing 100871, China

[2] Changping Laboratory, Beijing 102206, China.

[†] These authors contributed equally to this work.

[*] Authors to whom correspondence should be addressed: gaog@mail.cbi.pku.edu.cn (for G.G.), or caozj@mail.cbi.pku.edu.cn (for Z.J.C)

[#] Current address: Department of Computer Science, University of Washington, Seattle, WA 98195, United States

Abstract

Spatially resolved omics technologies reveal the spatial organization of cells in various biological systems. Integrative and comparative analyses of spatial omics data depend on proper slice alignment, which should take both omics profiles and spatial context into account. Here we propose SLAT (Spatially-Linked Alignment Tool), a graph-based algorithm for efficient and effective alignment of spatial omics data. Adopting a graph adversarial matching strategy, SLAT is the first algorithm capable of aligning heterogenous spatial data across distinct technologies and modalities. Systematic benchmarks demonstrate SLAT's superior precision, robustness, and speed vis à vis existing methods. Applications to multiple real-world datasets further show SLAT's utility in enhancing cell-typing resolution, integrating multiple modalities for regulatory inference, and mapping fine-scale spatial-temporal changes during development. The full SLAT package is available at https://github.com/gao-lab/SLAT.

Introduction

Recently emerging spatial omics technologies enable profiling the location, intercommunication, and functional cooperation of native cells through fluorescence *in situ* hybridization (seqFISH[1], MERFISH[2], seqFISH+[3] and Xenium[4]) and spatial barcoding (10x Visium[5], HDST[6], Slide-seqV2[7], Stereo-seq[8] and spatial-ATAC-seq[9]) from multiple tissue "slices", revealing tissue structure heterogeneity and shedding light on the underlying physiological and pathological mechanisms[8,10].

Properly aligning multiple slices, especially these generated from distinct sources, is critical for their integrative analysis. E.g., inter-technology alignment effectively bridges technologies complementary in spatial resolution and omics coverage[4], and the alignment across various experimental conditions helps reveal alterations in spatial patterns relevant to disease or perturbation[11]. However, current spatial alignment algorithms[12–14] are mostly designed for homogeneous alignments (e.g., three-dimensional reconstruction from consecutive slices[15,16]), and can hardly handle heterogeneous slices which often involve complex non-rigid deformations, uneven spatial resolutions as well as complex batch effects.

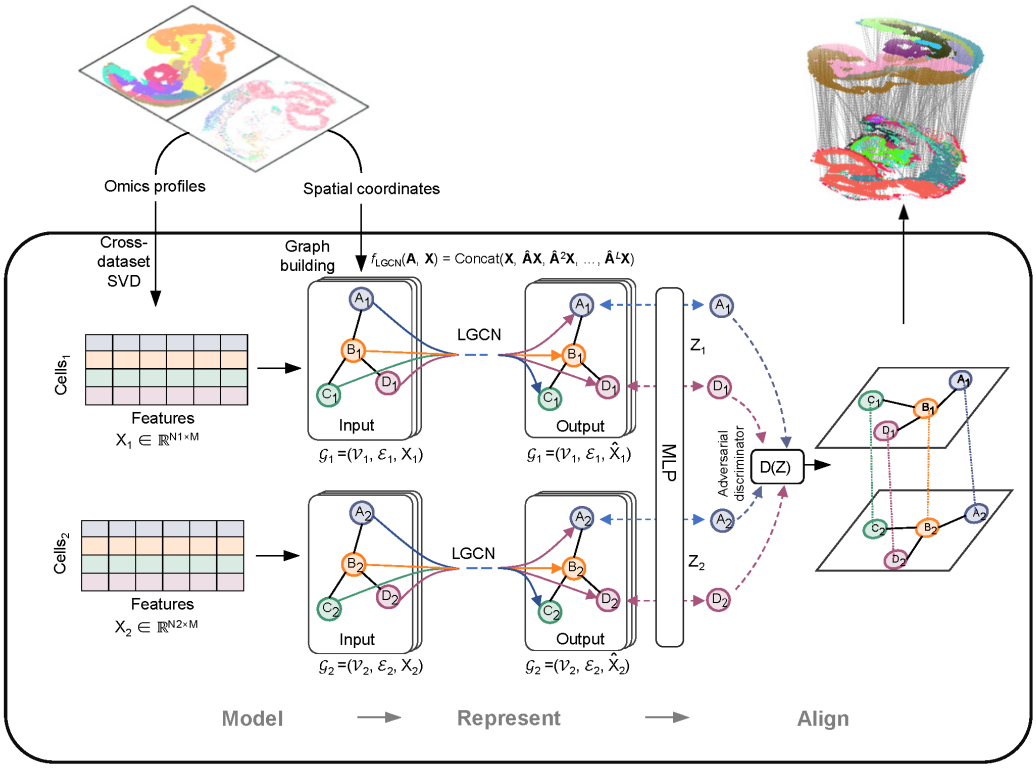
Here, we introduce SLAT (Spatially-Linked Alignment Tool), a unified framework for aligning both homogenous and heterogeneous single-cell spatial datasets. By modeling the intercellular relationship as a spatial graph, SLAT adopts graph neural networks and adversarial matching for robustly aligning spatial slices. In addition to its superior performance revealed by systematic benchmarks, as the first algorithm capable of aligning heterogenous spatial data, SLAT introduces a wide range of application scenarios including alignment across distinct technologies and experimental conditions. SLAT is publicly accessible at https://github.com/gao-lab/SLAT and will be continuously updated as spatial omics technologies evolve.

Results

Align heterogenous spatial omics data via graph adversarial matching

By modeling the spatial topology per slice as a spatial graph where each cell is connected to its nearest neighbors by edges, we reformulate the slice-alignment task as a graph-matching problem. In efforts to correct potential cross-dataset batch effect, SLAT employed an SVD-based matrix decomposition strategy to project omics profiles of the cells into a shared low-dimensional space (**Methods**), which in turn serves as node features of the spatial graphs.

Multilayer lightweight graph convolutional networks are incorporated to propagate and aggregate information between cells and their neighbors via stepwise concatenations, generating a holistic representation with information at multiple scales from individual cells to local niches as well as global positions. Then, SLAT solves a minimum-cost bipartite matching problem between the spatial graphs through a dedicated adversarial component[17] to align cells from different slices (Fig. 1, **Methods**).



SLAT: Spatially-Linked Alignment Tool

Fig. 1 Architecture of SLAT framework.

The SLAT algorithm can be divided into three main steps: model, represent and align. We model omics data from two single-cell spatial slices as low dimensional representations $\mathbf{X}_1 \in \mathbb{R}^{N_1 \times M}, \mathbf{X}_2 \in \mathbb{R}^{N_2 \times M}$ obtained via SVD-based cross-dataset decomposition, where N_1, N_2 are cell numbers of each slice and M is the SVD dimensionality. We use spatial coordinates of the cells to build K-neighbor spatial graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1, \mathbf{X}_1)$, $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2, \mathbf{X}_2)$ in each slice, respectively. The neighbor size can be set dynamically to adjust for uneven spatial resolution across technologies. To encode local and global spatial information, SLAT uses L-layer lightweight graph convolution networks to propagate and aggregate information across the spatial graph into cell embeddings $\widehat{\mathbf{X}}_1, \widehat{\mathbf{X}}_2$, which are then projected into alignment space $\mathbf{Z}_1, \mathbf{Z}_2$ with a multi-layer perceptron trained adversarially against a discriminator $f_D(\mathbf{Z})$ to align the two graphs by minimizing their Wasserstein distance (see **Methods**). Node labels with the same letters (e.g., A_1, A_2) represent corresponding cells in different slices.

Notably, apart from congruent regions that can be well aligned across slices, heterogeneous alignment often involves distinct regions that reflect biologically relevant spatial alterations. To avoid artificially aligning such regions (i.e., over-alignment), SLAT adopts an adaptive clipping strategy during adversarial matching to retain only the closest cells between slices, which represent the most reliable anchors to guide the alignment. Furthermore, SLAT calculates a similarity score for each aligned cell pair, making it possible to pinpoint spatially discrepant regions (see below and **Methods** for more details).

Systematic benchmarks suggest SLAT is accurate robust and fast

To evaluate SLAT's performance over existing algorithms which are designed for homogeneous alignment[13,15], we first ran systematic benchmarks based on consecutive slices from the same tissue generated by three representative technologies with a wide range of throughput, resolution, and technological routes: 10x Visium[5], MERFISH[2], and Stereo-seq[8] (Fig. 2a, Supplementary Data Table 1). All slices were randomly rotated before feeding to the alignment methods, which in turn attempt to recover the correct cell matching (as visualized by vertical lines connecting the two slices in Fig. 2b).

a

	Visium	MERFISH	Stereo-seq
Resolution	50μm	subcellular	0.2μm
Cells/Spots	~3,000	~6,500	~100,000
Features	>20,000 genes	151 genes	>20,000 genes
Tissue	Human DLPFC	Mouse hypothalamic preoptic	Whole E15.5 mouse embryo

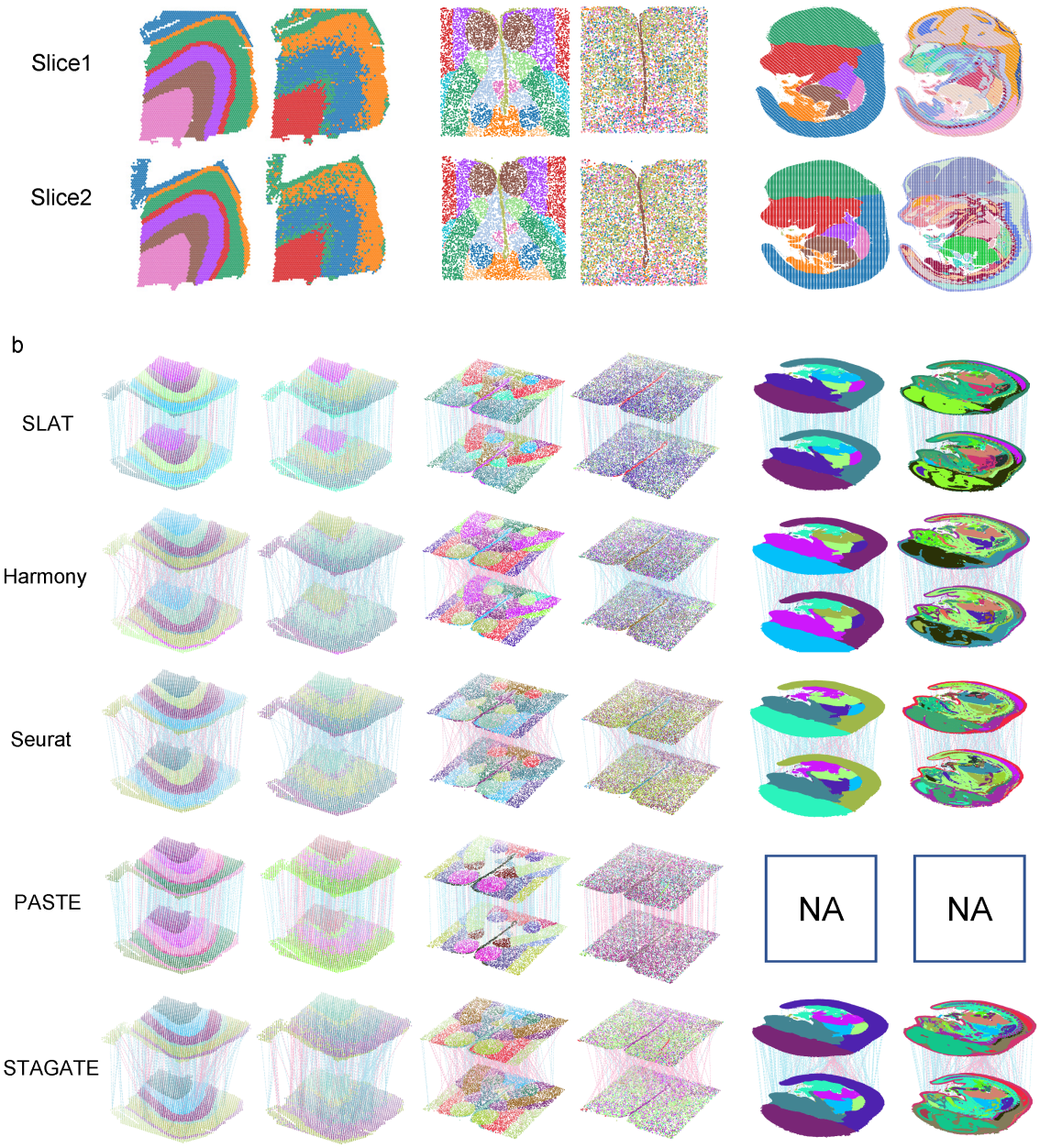


Fig. 2 Visualization of homogeneous evaluation on homogeneous spatial alignment. **a**, Summary of the three benchmark datasets (table above) and the slice used (panels below). For each dataset, two slices were used for alignment (one on each row), which are colored by spatial regions (left panels) and cell types (right panels), respectively, see Supplementary Fig. 1 for detailed color legends. **b**, Visualization of alignment results of different methods on the benchmark datasets in **a**. Vertical lines connect aligned cell pairs (subsampled to 300 alignment pairs for clear visualization). Blue lines indicate correct alignments while red lines indicate incorrect alignments. PASTE failed to run on the Stereo-seq dataset due to GPU memory overflow (capping at 80 GB).

An accurate spatial alignment method should match cells with maximal similarity in terms of both gene expression and spatial location, which can be quantified by the matching accuracy in expert curated cell types and spatial regions, respectively. Thus, we introduced a composite metric the Cell type and Region matching Index (CRI), which is defined as (see **Methods** for more details).

$$\mathrm{CRI} = \frac{1}{|\mathcal{M}|} \sum_{(v_i, v_j) \in \mathcal{M}} I(i,j),$$

$$I(i,j) = \begin{cases} 1, & c_1^{(i)} = c_2^{(j)} \text{ and } r_1^{(i)} = r_2^{(j)} \\ 0, & \text{otherwise.} \end{cases}$$

where i, j are indices into the two aligned slices, $\mathcal{M}$ is the set of aligned cell pairs, $c_1^{(i)}, c_2^{(j)}$ and $r_1^{(i)}, r_2^{(j)}$ are cell type and spatial region annotations of the aligned cells.

Given the fact that CRI relies on high-level annotations and lacks single cell resolution, we also reported the average Euclidean distance between matched cells as measured in the original coordinates before rotation. Of note, the Euclidean distance metric implicitly assumes that the nearest cells across slices have equivalent identities. This assumption is reasonable in "cluster" scenarios where cells of the same type occupy continuous spatial regions, but can be violated in "patchwork" scenarios where cell types are spatially interlaced (Extended Data Fig. 1a).

SLAT outperformed current spatially aware algorithms PASTE[15] and STAGATE[13] in all cases, except in the MERFISH slices where PASTE registered the best Euclidean distance (Fig. 3a), which is not unexpected as it explicitly uses Euclidean distance as the target function[15]. Nonetheless, we noticed that the Euclidean metric is unreliable for this particular dataset as it falls into the "patchwork" category (Extended Data Fig. 1a and Supplementary Fig. 2). The very same PASTE alignment exhibited particularly low cell type accuracy (Fig. 3b), indicating that it is in fact incorrect. Although STAGATE also employs graph neural networks for modeling spatial omics data, it produced suboptimal accuracies. An additional comparison using split slices free of batch effects (**Methods**, Supplementary Fig. 3) found increased accuracy of STAGATE (Fig. 3c), suggesting that lack of proper batch effect correction could be the reason for inferior accuracy in real alignment tasks.

3D reconstruction from consecutive slices is a common application of homogenous alignment. Thus, we also evaluated the performance in this scenario using three groups of consecutive brain slices (Supplementary Fig. 4) quantified by region matching accuracy (**Methods**), i.e., the same dataset and metric used in PASTE[15]. SLAT remained the top performer in average accuracy (0.780), slightly higher than the second-best method PASTE (0.754), and also exhibited favorable stability (Fig. 3d).

As expected, SLAT also achieved significantly higher alignment accuracy than the conventional spatially unaware algorithms Seurat[18] and Harmony[19] (Fig. 3a). In particular, although these algorithms matched cell types reasonably well, their alignment largely disarranged spatial regions (Fig. 3b). By combining spatial context with transcriptome profiles, SLAT is better equipped to distinguish transcriptionally similar but spatially distinct cell groups, thus boosting alignment accuracy.

SLAT's lightweight graph convolutional component effectively improves model robustness. Evaluation with subsampled data of various sizes showed that SLAT consistently provides the best results, even with as few as 200 cells (Fig. 3e, **Methods**). Further inspection showed that the performance of SLAT remains highly robust to a wide range of hyperparameter settings and against random corruption of the spatial graph (Extended Data Fig. 2b, **Methods**).

As technologies continue to evolve, the throughput of spatial single-cell experiments is constantly increasing[20]. Implemented as a neural network and optimized for parallelism, SLAT is highly scalable: benchmarks showed that SLAT is consistently the fastest method across all data sizes and aligns slices each with over 100,000 cells in just 3 minutes (Fig. 3f). We also noticed that PASTE fails to run as cell number exceeds 25,600, partly due to its memory intensive implementation for optimal transport.

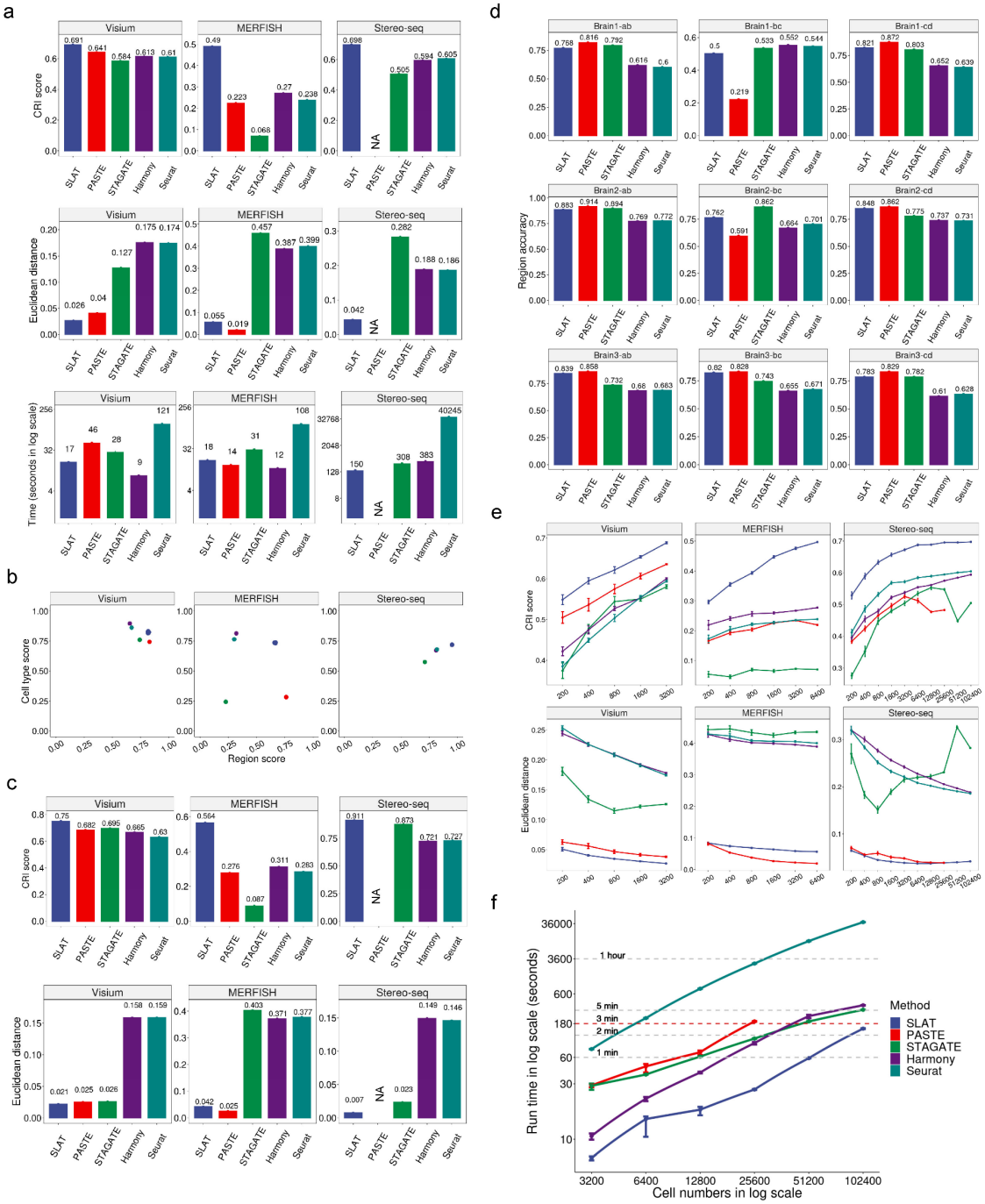


Fig. 3 Quantitative performance evaluation in homogenous spatial alignment. **a**, Alignment accuracy of different methods on the three representative technologies in Fig. 2. From top to bottom are different metrics (CRI, Euclidean distance, and running time). $n = 8$ repeats with different model random seeds. PASTE failed to run on the Stereo-seq dataset due to GPU memory overflow (capping at 80 GB). **b**, Cell-type matching accuracy (y-axis) and region-matching accuracy (x-axis) in CRI plotted separately. **c**, Alignment accuracy of different methods on split slices of the three datasets. $n = 8$ repeats with different model random seeds. **d**, Region matching accuracy of different methods in 3D reconstruction. $n = 8$ repeats with different model random seeds. Dataset contains three groups, each with three consecutive slices (a, b and c, more details are in **Methods**). **e**, Changes in CRI (top) and Euclidean distance (bottom) on the three datasets with subsamples of varying sizes. $n = 8$ repeats with different subsampling random seeds. **f**, Running time of each method on subsampled datasets of varying sizes. $n = 8$ repeats with different subsampling random seeds. Error bars indicate mean ± s.d.

169 **Matching heterogeneous datasets across distinct technologies and modalities**

170 Spatial alignment across multiple technologies and modalities enables complementary information
171 towards a comprehensive *in situ* view of cellular states[21]. Benefiting from its novel design, SLAT
172 provides reliable alignment across heterogeneous datasets, which is largely beyond the capability of
173 current alignment algorithms.

174

175 We first used SLAT to align spatial transcriptomic datasets generated by two distinct technologies
176 with varying scale and detectability[8,22]. Specifically, the seqFISH dataset contained over 10,000 cells
177 with only 350 genes detected in total [3,22], whereas the Stereo-seq dataset contains about 5,000 cells
178 covering over 20,000 genes[8]. Meanwhile, the cell types in the two datasets were annotated at
179 different resolutions: the seqFISH dataset has finer-grained annotation (21 cell types, Extended Data
180 Fig. 3a) than the Stereo-seq dataset (11 cell types, Extended Data Fig. 3b). The high-quality
181 alignment produced by SLAT (Fig. 4a, also see Extended Data Fig. 4 for a more quantitative
182 assessment based on paired Visium and Xenium slices) enables an accurate label transfer for cell
183 typing improvement (Fig. 4b). For example, cells labeled as "Neural crest" in the Stereo-seq dataset
184 were aligned to seqFISH regions with four cell types, refining cell typing which was further
185 validated by known marker genes (Fig. 4c and Extended Data Fig. 3c and 3d). Notably, slices in the
186 two datasets also differ in spatial structure with substantial non-rigid deformations (Extended Data
187 Fig. 3a and 3b). Such challenging scenario effectively failed methods other than SLAT (Extended
188 Data Fig. 5).

189

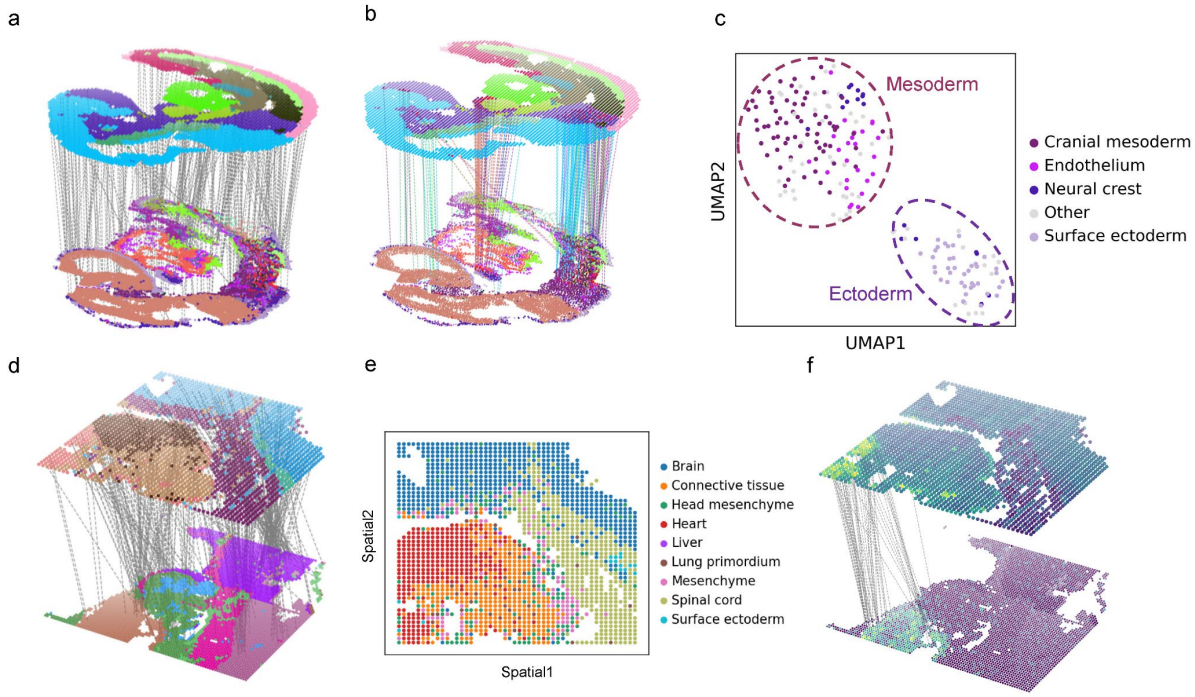


Fig. 4 Spatial alignment across distinct technologies and modalities.
a, Visualization of the alignment of E8.75 seqFISH mouse embryo and E9.5 Stereo-seq mouse embryo datasets, colored by cell types (subsampled to 300 alignment pairs for clear visualization). **b**, Highlighting the alignment of cells labeled as "Neural crest" in Stereo-seq (top) to seqFISH (bottom), alignment lines are colored by cell type of aligned cells in seqFISH. **c**, UMAP visualization of cells labeled as "Neural crest" in Stereo-seq, colored by cell type of aligned cells in seqFISH, cell types with fractions less than 5% are labeled as "Other". Circles indicate manual annotation by marker genes (corresponding to Extended Data Fig. 3c and 3d). **d**, Visualization of the alignment of E11.5 spatial-ATAC-seq mouse embryo and E11.5 Stereo-seq mouse embryo datasets, with spatial-ATAC-seq dataset colored by clusters (see Extended Data Fig. 6e) and Stereo-seq colored by cell types, respectively (subsampled to 300 alignment pairs for clear visualization). **e**, Cell type labels of spatial-ATAC-seq transferred from Stereo-seq via SLAT alignment. **f**, Showing chromatin accessibility score and gene expression pattern of heart marker *Tnnt2* on SLAT alignment.

More challenging is the cross-modality spatial alignment, which is partly due to the disjoint feature spaces where invalidate the use of our SVD-based cross dataset matrix decomposition strategy as well as other canonical batch correction methods. Benefitting from the modular design of SLAT (Fig. 1), we replaced the SVD-based strategy with GLUE[23] to project cells of different modalities into a shared embedding space before feeding them into the LGCNs (**Methods**). With this extension, SLAT successfully produced a spatial alignment across RNA (Stereo-seq) and ATAC (spatial-ATAC-seq) slices. While the different modalities featured drastically different spatial resolutions (0.2 μm for Stereo-seq but 20 μm for spatial-ATAC-seq), SLAT managed to align them well (Fig. 4d, Extended Data Fig. 6a). Cell-type labels transferred from Stereo-seq to spatial-ATAC-seq based on the

212 alignment were consistent with anatomical features and with the accessibility of tissue-specific genes
213 (Fig. 4e and 4f, Extended Data Fig. 6d). Joint regulatory analysis using the aligned cell pairs and
214 transferred labels effectively identified various key regulators in the heart, such as *Jund* and
215 *Ctnnb1*[24,25] (**Methods**, Extended Data Fig. 6b). We also experimented with other spatial alignment
216 methods by feeding GLUE embeddings as input, but found suboptimal results (Supplementary Fig.
217 6).
218

219 **Mapping fine-scale spatial-temporal transitions by developmental alignment**

220 Embryonic development is a highly dynamic process with extensive spatial-temporal transitions
221 involving the generation, maturation, and functional alteration of organs and tissues at specific
222 timepoints. To probe spatial-temporal dynamics during early development, we used SLAT to align
223 two spatial atlases of mouse embryonic development at E11.5 and E12.5[8] (Fig. 5a).
224

225 Most cells in the brain, heart, and liver were well aligned with high alignment credibility scores,
226 consistent with their spatial-temporal conservation at the two timepoints (Fig. 5b, Extended Data Fig.
227 7a and 7b). In addition, several regions were enriched with less-aligned cells of lower alignment
228 credibility scores, which may be attributed to both biological development (e.g., the newly emerged
229 organs kidney and ovary at region I[8,26–28], the rapid enlargement of lung primordium and its
230 displacement from the upper part of the heart to the lower part at region II[29] as shown in Extended
231 Data Fig. 7c, and the disappearance of branchial arch at corresponding position of region III in
232 E11.5[8]) and technical variation (e.g., tissue loss at corresponding position of region IV in E11.5).
233

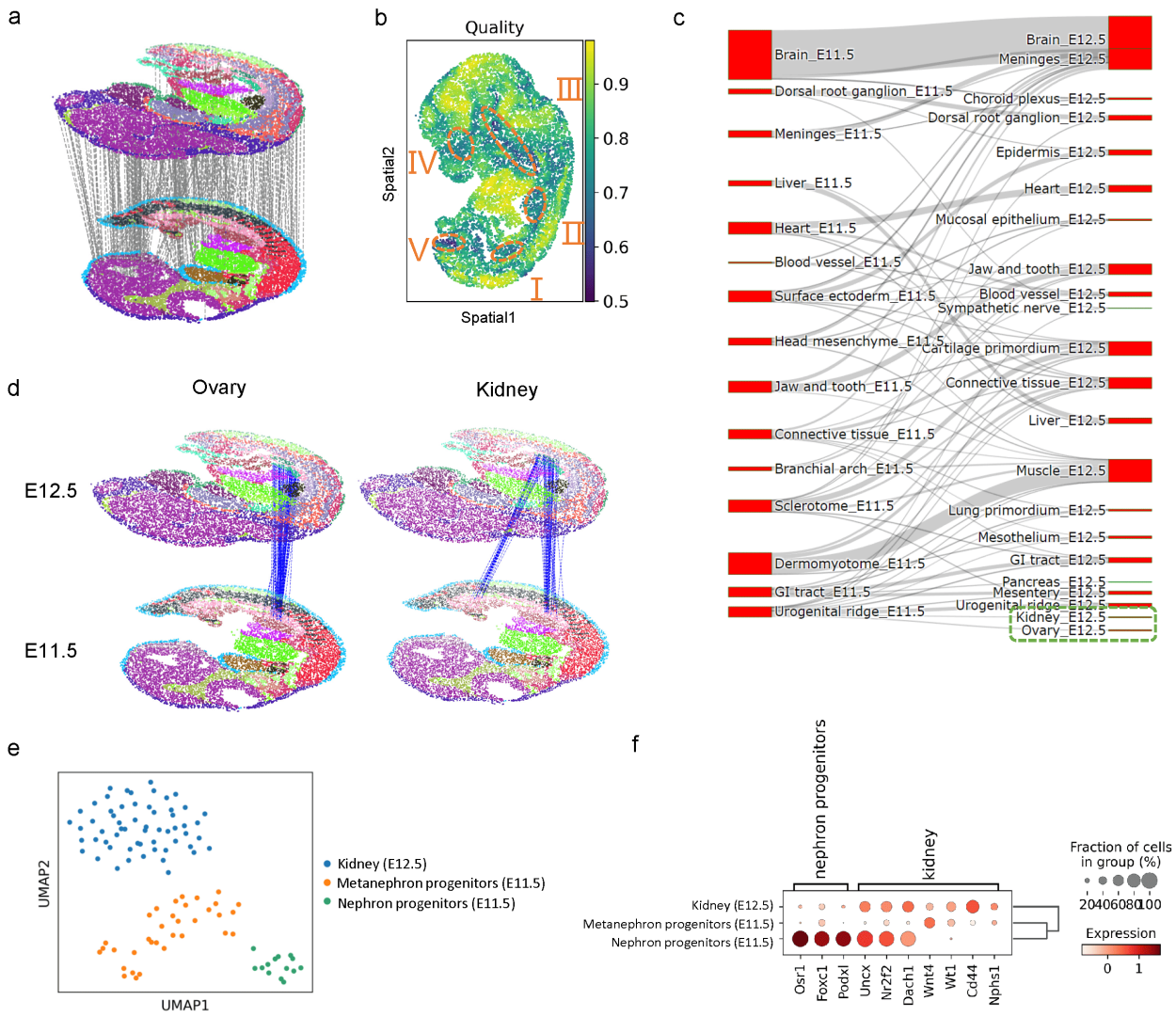


Fig. 5 Developmental alignment of mouse embryo.

a, Visualization of the alignment of E11.5 and E12.5 mouse embryo Stereo-seq datasets, colored by cell types (subsampled to 300 alignment pairs for clear visualization). **b**, Similarity score of the alignment. Higher scores indicate higher alignment confidence. Circles highlight five regions with low similarity scores. **c**, Sanky plot showing cell type correspondence of SLAT alignment between the two datasets. The green box highlights cells labeled as "Kidney" and "Ovary" in E12.5. **d**, Alignment visualization highlighting cells labeled as "Kidney" (left) and "Ovary" (right) in E12.5 and their aligned cells in E11.5, respectively. **e**, UMAP visualization of elaborate cell type annotation for cells labeled as "Kidney" in E12.5 and their aligned cells in E11.5. **f**, Dot plot showing marker gene expression of cell types in **e**.

We next followed up on kidney and ovary, two newly emerging organs at around E11.5–12.5[8,26–28]. SLAT accurately identified them both as developing from the "Urogenital ridge" at E11.5 (highlighted by the green box in Fig. 5c). Consistent with previous reports, SLAT alignment showed precisely that the ovary develops directly from a single area[30] (left panel of Fig. 5d) while the kidney develops from two separate areas corresponding to the mesonephros and metanephros structures in

early kidney development[26,31,32] (Fig. 5d, Extended Data Fig. 7d). In addition, based on the clustering analysis of aligned cells, we further identified a group of rare nephron progenitor cells located in the mesonephros, likely corresponding to an ephemeral mesonephric tubule during mesonephros development[33] (Fig. 5e and 5f). Interestingly, the results show that the mesonephros is spatially adjacent to the ovary, which is consistent with the well-documented developmental cascade whereby a part of the mesonephros develops into the fallopian tube after degeneration (Fig. 5d). Our findings revealed the unique value of heterogeneous spatial alignment for interrogating the spatial-temporal process of organogenesis.

We also attempted the same alignment task with other methods (Extended Data Fig. 8). The spatially unaware algorithms (Seurat, Harmony) simply missed mesonephros, while the spatially aware algorithms (PASTE, STAGATE) align them to incorrect cells, further demonstrating the unique capability of SLAT on heterogeneous spatial alignment.

262 # Discussion

263 One of the essential challenges in spatial omics alignment is to appropriately model spatial context.
264 Early methods such as Splotch[14] and Eggplant[34] model slices as rigid bodies and require manual
265 annotation of landmark spots to guide the alignment. PASTE[15] eliminates the need for landmark
266 annotation by considering gene expression of all spatial spots, but its reliance on exact spatial
267 distance impedes application to heterogeneous alignments involving complex non-rigid structural
268 alterations. By combining spatial graph convolution and adversarial matching, SLAT achieves
269 reliable spatial alignment for both homogeneous and heterogeneous slices in an unsupervised, data-
270 oriented manner.
271
272 In addition to pairwise alignments as demonstrated in the 3D reconstruction benchmark, SLAT also
273 generalizes to the alignment of more than two slices. As an example, Extended Data Fig. 9 shows a
274 simultaneous alignment across a consecutive stack of four slices from the same E15.5 mouse
275 embryo. While similar 3D stacking can also be achieved with PASTE[15], SLAT is better equipped to
276 account for non-rigid structural alterations along the z-axis, which allows building more accurate 3D
277 models corrected for deformational artifacts.
278
279 SLAT's unique ability to conduct heterogeneous spatial alignments promises a wide range of
280 biological applications. E.g., the alignment of tumor samples across different stages could help
281 pinpoint key alterations of spatial patterns relevant to tumor progression. Aligning the same type of
282 tissue across multiple species could provide insight into the evolution of cell types and their
283 interactions, as well as clues for their functional contributions. Additionally, such spatial alignments
284 also enable identifying spatially-resolved changes in gene expression, analogous to cell-type specific
285 differential expression analyses commonly used in conventional single-cell studies.
286
287 SLAT is fast. In fact, generating the input cell embeddings is the most time-consuming step, while
288 training and inferring with the SLAT core model takes only about 10 seconds for 10^6 cells. Once the

input embeddings are ready, aligning millions of cells can be completed in near real-time, enabling an efficient search of a massive database within an affordable timeframe. SLAT's blazing speed sets the stage for some exciting applications such as inferring spatially dependent causality of perturbations by systematically comparing multiple Perturb-map slices[11], or constructing whole organ 3D atlases involving thousands of slices[35].

Last but not least, designed as a flexible framework, SLAT can be readily adapted and extended. For instance, additional information such as tissue imaging or expert curation may be incorporated into the coordinate matching module to help distinguish symmetric structures and polish the final alignment (**Methods**). Meanwhile, the spatial graph modeling technique employed in SLAT may also be adapted to address other problems, such as identifying novel cell-cell interactions by integrating ligand-receptor pairing information[36].

Overall, SLAT provides a unified framework for various spatial integration scenarios. To promote its application by the research community, the SLAT package, along with detailed tutorials and demo cases, is available online at https://github.com/gao-lab/SLAT.

Acknowledgments

We thank Mr. H. Yang for his helpful guidance on mouse embryonic development and anatomy, as well as Drs. Z. Zhang, L. Tao, F. Tang, X.S. Xie, C. Li, and J. Lu at Peking University for their helpful discussions and comments during the study. This work was supported by funds from the National Key Research and Development Program of China (2022ZD0115004), as well as the State Key Laboratory of Protein and Plant Gene Research, the Beijing Advanced Innovation Center for Genomics (ICG) at Peking University, the Changping Laboratory, and the Shaw Foundation Hong Kong Limited. The research of G.G. was supported in part by the National Program for Support of Top-Notch Young Professionals.

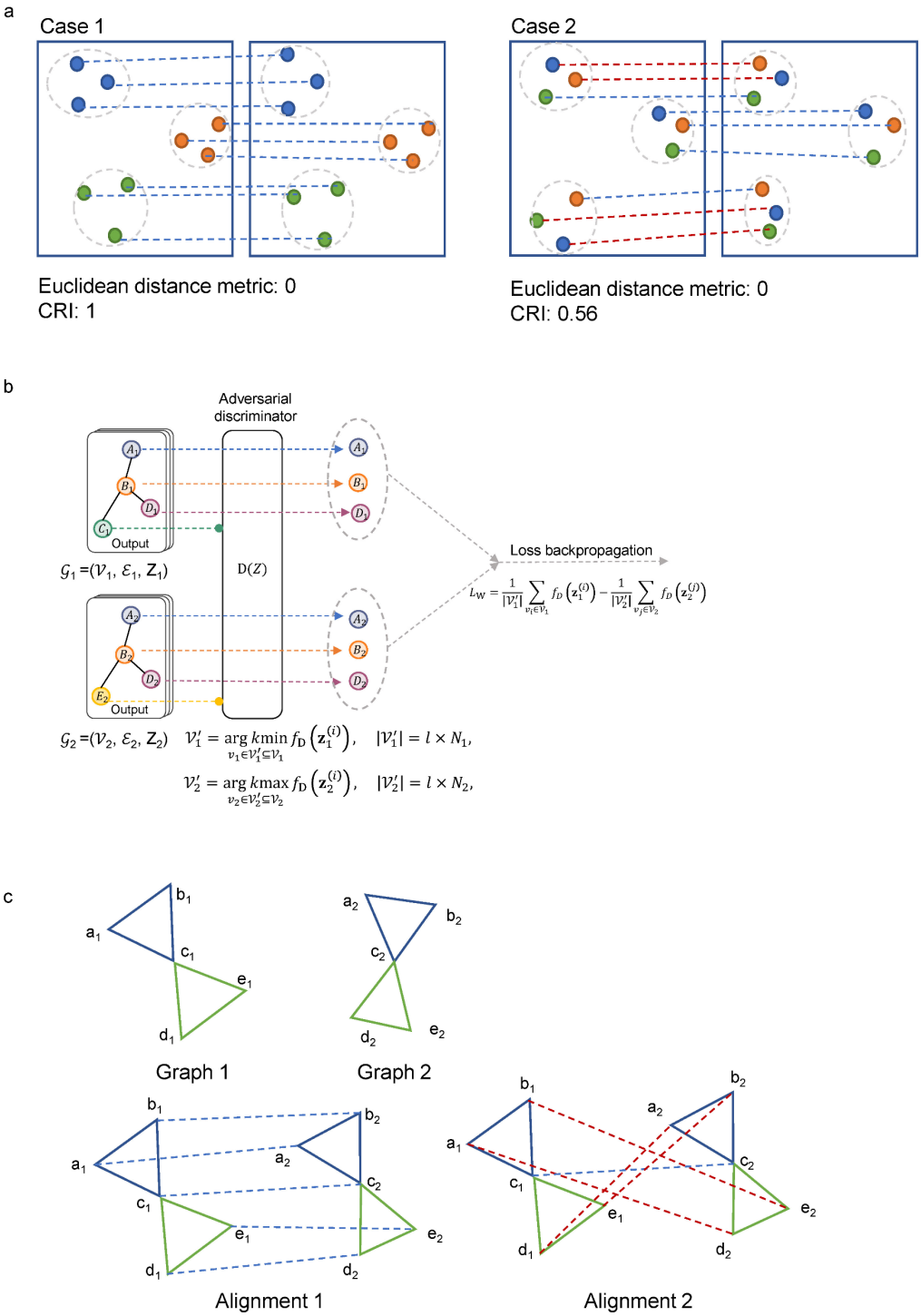
Author contributions

G.G. and Z.J.C conceived the study and supervised the research. C.R.X and Z.J.C. designed and implemented the computational framework and conducted benchmarks and case studies with guidance from G.G. X.M.T proposed and implemented the matrix decomposition-based batch correction strategy. C.R.X, Z.J.C., X.M.T and G.G. wrote the manuscript.

Competing interests

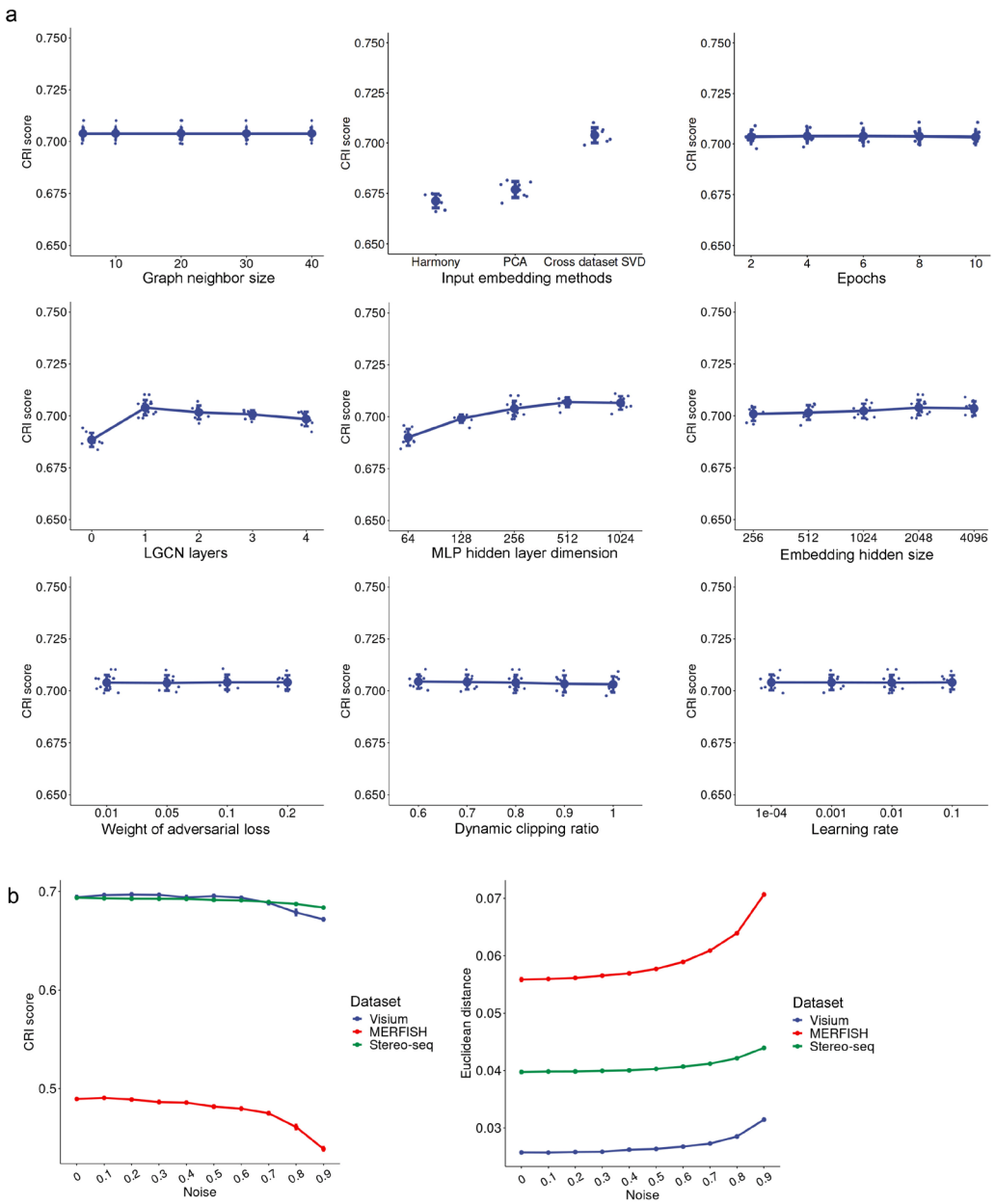
The authors declare no competing interests.

Extended Data Figures

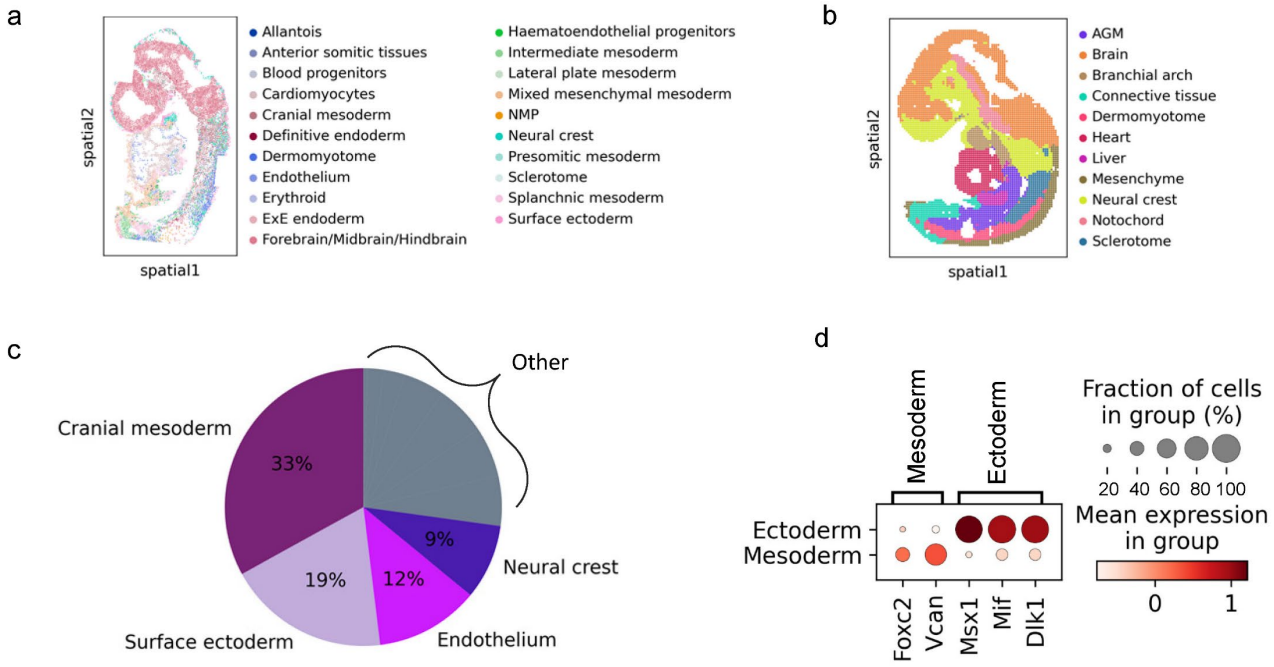


Extended Data Fig. 1 Illustration of dynamic clipping and alternative metrics.

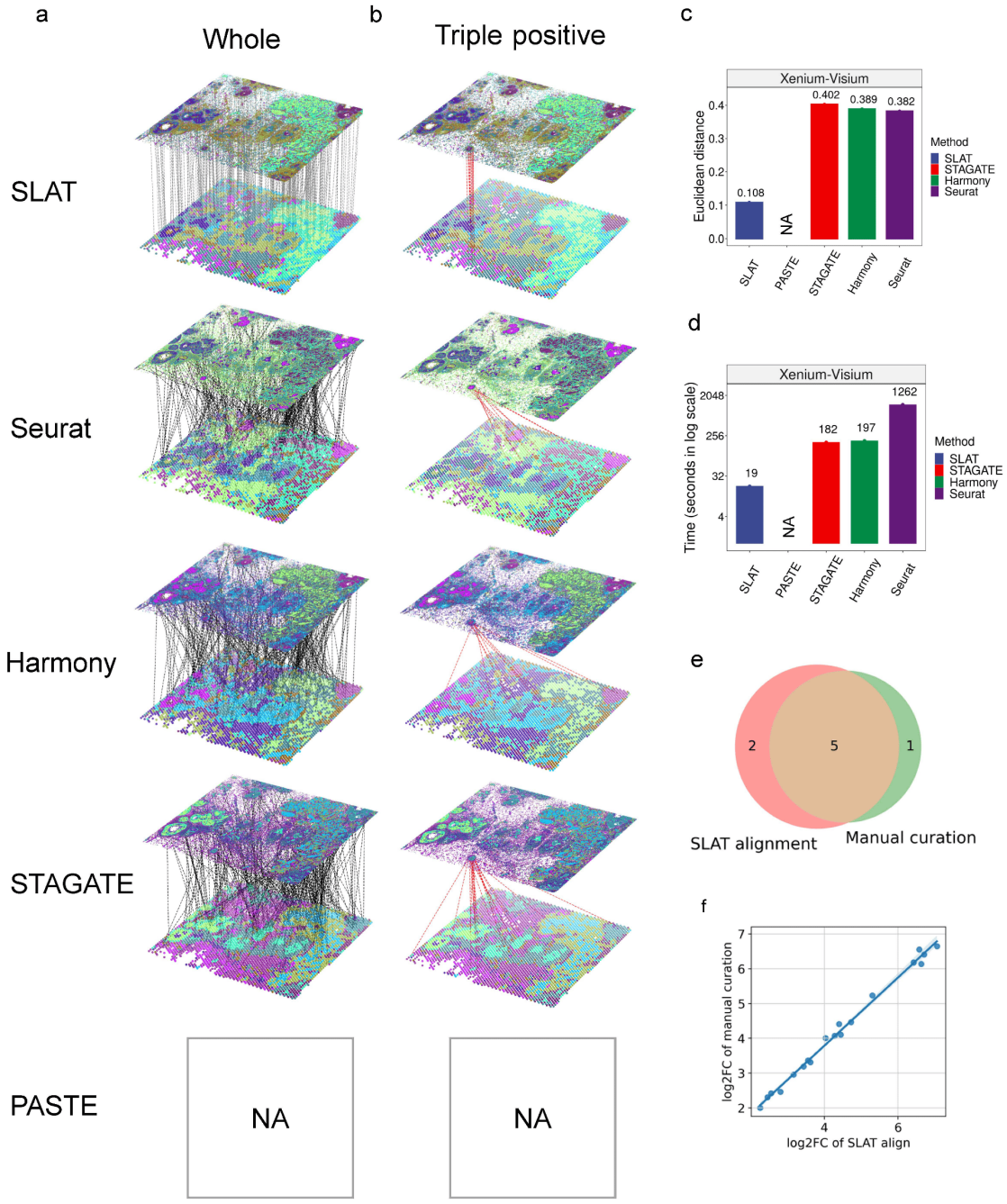
a, Illustration where the Euclidean distance metric is ineffective. Each point represents a cell and the color indicates cell type. The blue dashed circles indicate spatial regions. "Case 1" represents a "cluster" scenario where cells of the same types occupy continuous spatial regions. "Case 2" represents a "patchwork" scenario where cell types are spatially interlaced. Horizontal lines across two slices indicate cell pairs matched by shortest Euclidean distances, where blue lines indicate correct alignments and red lines indicate erroneous alignments. In "Case 1", both CRI and Euclidean distance are effective. In "Case 2", some cells with shortest distances are not of the same cell type, but Euclidean distance is cell type-unaware and gives the same score as in "Case 1", while CRI penalizes cell type mismatches correctly. **b**, Illustration of the dynamic clipping criteria used to select cells for adversarial training. Only a subset of cells from the two datasets with minimal Wasserstein distance are selected (see **Methods**). **c**, Illustration where the "edge score" metric is ineffective. "Graph 1" and "Graph 2" are two graphs with five nodes, where nodes with the same letters (e.g., a_1, a_2) represent ground truth node pairs. "Alignment 1" and "Alignment 2" are two graph-matchings: "Alignment 1" correctly matches all nodes between two graphs, while only one node is correctly matched in "Alignment 2". Nevertheless, both alignments give the same "edge score".



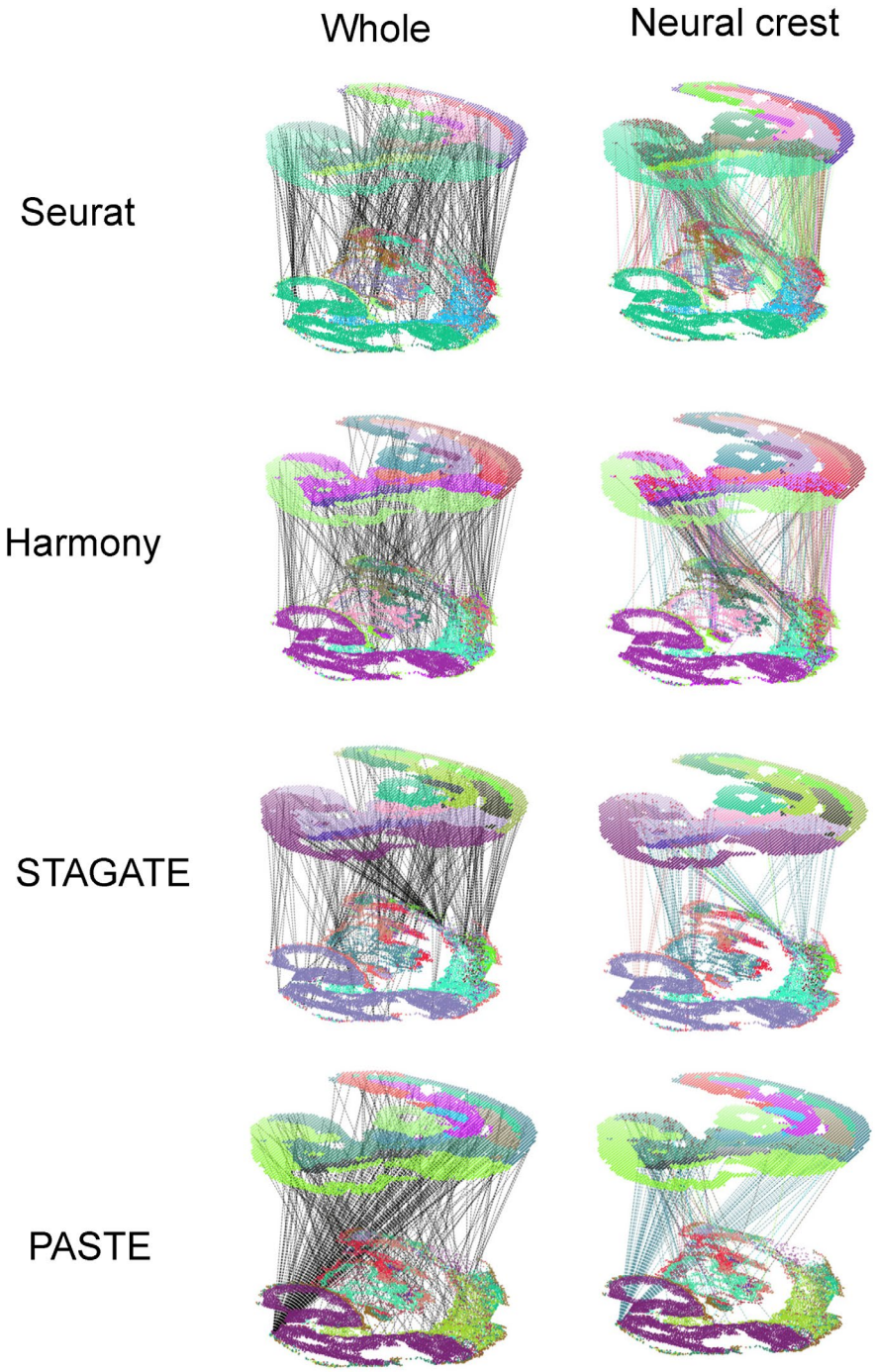
Extended Data Fig. 2 Effect of hyperparameter settings on alignment performance. **a**, CRI metric for SLAT with different hyperparameter settings including: graph neighbor size (K), input embedding methods, number of training epochs, number of LGCN layers (L), MLP hidden layer dimension, embedding hidden size (P), weight of the adversarial loss (α), dynamic clipping ratio (l), and learning rate. $n = 8$ repeats with different model random seeds. We scaled the y-axis to [0.65, 0.75] to show subtle changes. See **Methods** for detailed explanations of the hyperparameters. **b**, Changes in CRI metric (left) and Euclidean distance metric (right) at different spatial graph corruption rates. $n = 8$ repeats with different model random seeds. **c**, Ratio of nearest cell pairs across two slices that belong to the same cell type. **d**, Ratio of nearest cell pairs within one spatial slice that belong to the same cell type. The two slices of each technology are shown separately.



Extended Data Fig. 3 Refining cell type annotation in Stereo-seq via label transfer.
a-b, Original annotation of **a**, the seqFISH dataset and **b**, the Stereo-seq dataset. **c**, Proportions seqFISH transferred cell types in Stereo-seq cells originally labeled as "Neural crest". Cell types with proportions of less than 5% are labeled as "Other". **d**, Expression of cell type marker genes as partitioned in Fig. 5c.

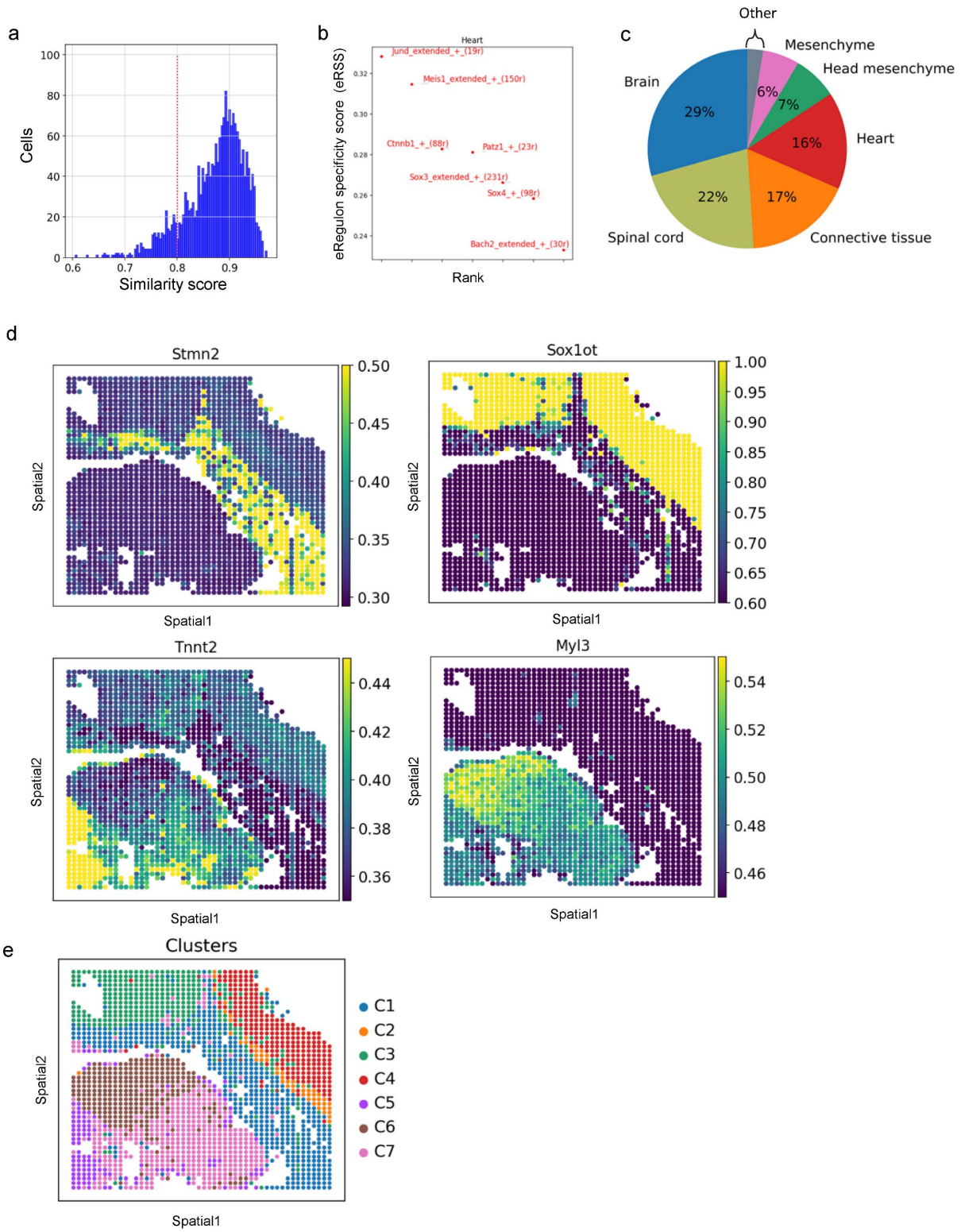

a
Whole
b
Triple positive
SLAT
Seurat
Harmony
STAGATE
PASTE
NA
NA
c
Xenium-Visium
Euclidean distance
0.108
NA
0.402
0.389
0.382
SLAT
PASTE
STAGATE
Harmony
Seurat
Method
SLAT
STAGATE
Harmony
Seurat
d
Xenium-Visium
Time (seconds in log scale)
19
NA
182
197
1262
SLAT
PASTE
STAGATE
Harmony
Seurat
Method
SLAT
STAGATE
Harmony
Seurat
e
2
5
1
SLAT alignment
Manual curation
f
log2FC of manual curation
log2FC of SLAT align

Extended Data Fig. 4 Alignment results of current methods on the Xenium and Visium slices. Xenium and Visium are spatial technologies designed by 10x with complementary capabilities[4]. Xenium achieves subcellular resolution but only detects a limited set of genes, while Visium profiles the complete transcriptome with coarse spatial resolution. Specifically, the Xenium slice covers more than 100,000 cells with 313 genes detected in total, and the Visium slice contains 3,841 spots with over 20,000 genes[4]. **a**, **b**, Visualization of the alignment results of two consecutive Xenium and Visium human breast cancer slices via different methods **(**subsampled to 300 alignment pairs for clear visualization), **b** Highlighting cells labeled as "Triple positive breast tumor cells" (ERBB2+/ESR1+/PGR+, Supplementary Fig. 5a and 5b) in Xenium and their aligned cells in Visium. PASTE failed due to GPU memory overflow (capping at 80 G). Of note, SLAT is the only algorithm that successfully aligned Xenium-resolved rare triple positive cells to Visium slice, even though they span less than 10 spots (Supplementary Fig. 5c and 5d). **c**, **d**, Euclidean distance metric (**c**) and run time (**d**) of different methods. **e**, Overlapping of spots between the SLAT aligned and manually curated triple positive spots. **f**, Log_2 fold change of shared top 25 marker genes in **e** for SLAT aligned compared to manually curated triple positive spots (Supplementary Fig. 5e).

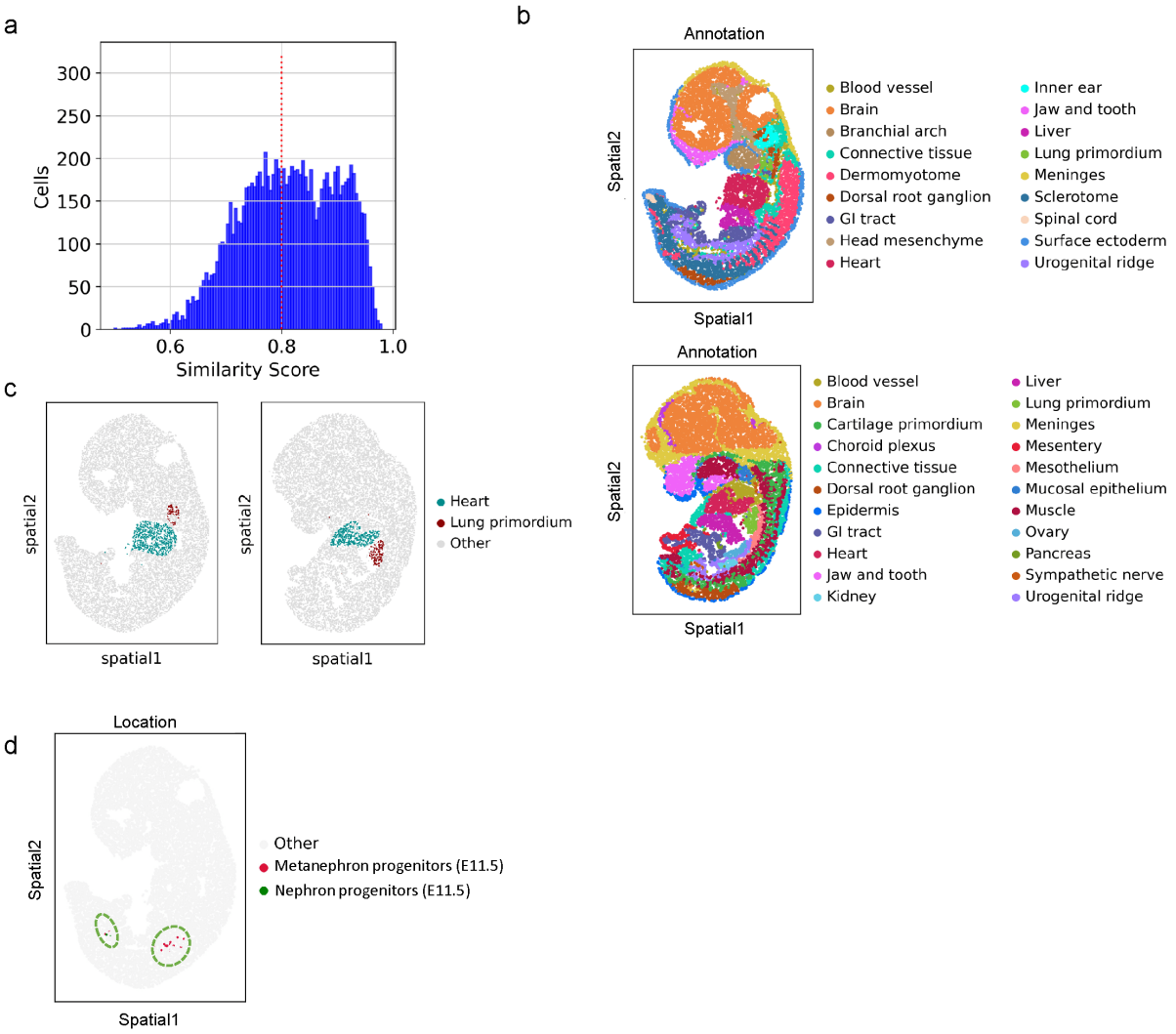


Extended Data Fig. 5 Alignment results of current methods on the seqFISH and Stereo-seq slices.

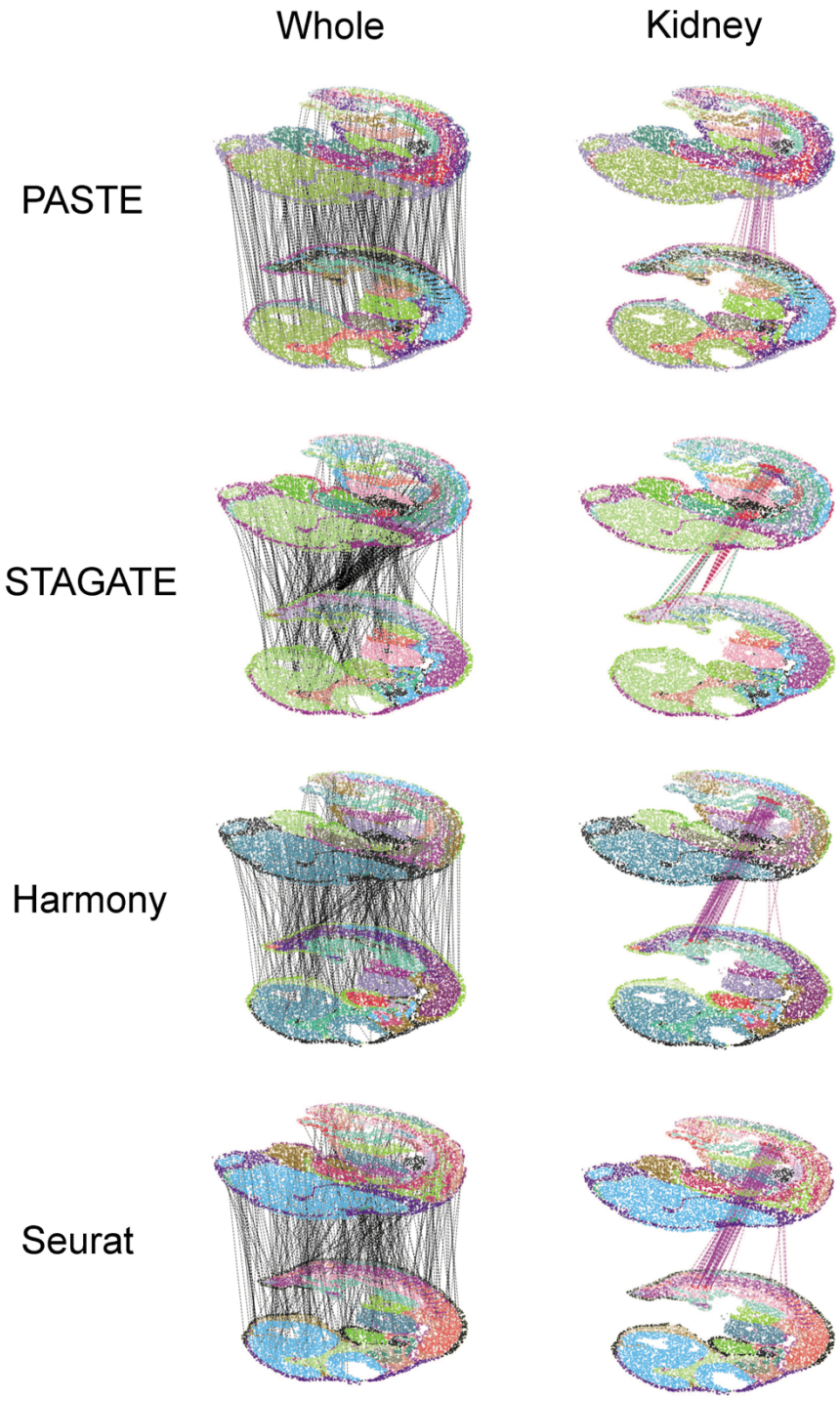
Visualization of the alignment results of seqFISH and Stereo-seq mouse embryo slices via current methods. The left panels show the complete alignment results (subsampled to 300 alignment pairs for clear visualization), while the right panels highlight cells labeled as "Neural crest" in Stereo-seq and their aligned cells in seqFISH.



Extended Data Fig. 6 Alignment-based cross-modality annotation and regulatory reference. **a**, Similarity score distribution in spatial-ATAC-seq and Stereo-seq alignment. **b**, Key heart regulators identified by SCENIC+ based on the SLAT alignment. **c**, Proportions of Stereo-seq transferred cell types in the spatial-ATAC-seq slice. Cell types with proportions of less than 5% are labeled as "Other". **d**, Visualization of chromatin accessibility score of cell type marker genes in the spatial-ATAC-seq slice. **e**, Leiden clustering results of the spatial-ATAC-seq slice.



Extended Data Fig. 7 Spatial-temporal dynamics in mouse embryonic development. **a**, Similarity score distribution in E11.5 and E12.5 mouse embryo Stereo-seq slice alignment. **b**, Original annotation of E11.5 (top) and E12.5 (bottom) mouse embryo datasets. **c**, Spatial location of "Heart" and "Lung primordium" cells in the E11.5 and E12.5 slice, respectively. **d**, Spatial location of "Kidney" aligned cells in E11.5 in the right panel of **d**, colored by cell type in **e**.



Extended Data Fig. 8 Alignment results of current methods in mouse embryonic development. Alignment of E11.5 and E12.5 mouse embryo slices using current methods. The left panels show the complete alignments (subsampled to 300 alignment pairs for clear visualization), while the right panels highlight cells labeled as "Kidney" in E12.5 and their aligned cells in E11.5.

Extended Data Fig. 9 Multi-slice SLAT alignment for 3D reconstruction.
Simultaneous SLAT alignment across four slices from the same E15.5 mouse embryo, ordered by vertical position (subsampled to 300 alignment pairs for clear visualization).

398 # References

399 1. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA
400 profiling by sequential hybridization. *Nat Methods* **11**, 360–361 (2014).
401 2. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly
402 multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
403 3. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+.
404 *Nature* **568**, 235–239 (2019).
405 4. Janesick, A. *et al.* High resolution mapping of the breast cancer tumor microenvironment using
406 integrated single cell, spatial and in situ analysis of FFPE tissue. *bioRxiv* 2022.10.06.510405
407 (2022) doi:10.1101/2022.10.06.510405.
408 5. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial
409 transcriptomics. *Science* **353**, 78–82 (2016).
410 6. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods*
411 **16**, 987–990 (2019).
412 7. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with
413 Slide-seqV2. *Nat Biotechnol* **39**, 313–319 (2021).
414 8. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-
415 patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).
416 9. Deng, Y. *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*
417 1–9 (2022) doi:10.1038/s41586-022-05094-1.

10. Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766–777 (2022).
11. Dhainaut, M. *et al.* Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* **185**, 1223-1239.e20 (2022).
12. Long, Y. *et al.* DeepST: A versatile graph contrastive learning framework for spatially informed clustering, integration, and deconvolution of spatial transcriptomics. 2022.08.02.502407 Preprint at https://doi.org/10.1101/2022.08.02.502407 (2022).
13. Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* **13**, 1739 (2022).
14. Äijö, T. *et al.* Splotch: Robust estimation of aligned spatial temporal gene expression data. 757096 Preprint at https://doi.org/10.1101/757096 (2019).
15. Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nat Methods* **19**, 567–575 (2022).
16. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).
17. Gao, J., Huang, X. & Li, J. Unsupervised Graph Alignment with Wasserstein Distance Discriminator. in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* 426–435 (ACM, 2021). doi:10.1145/3447548.3467332.
18. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
19. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289–1296 (2019).

439 20. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534–546 (2022).

440 21. Eisenstein, M. Seven technologies to watch in 2022. *Nature* **601**, 658–661 (2022).

441 22. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse

442 organogenesis. *Nat Biotechnol* **40**, 74–85 (2022).

443 23. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with

444 graph-linked embedding. *Nat Biotechnol* 1–9 (2022) doi:10.1038/s41587-022-01284-4.

445 24. Ricci, R. *et al.* Distinct functions of junD in cardiac hypertrophy and heart failure. *Genes Dev* **19**,

446 208–213 (2005).

447 25. Ombrato, L. *et al.* Generation of neighbor-labeling cells to study intercellular interactions in

448 vivo. *Nat Protoc* **16**, 872–892 (2021).

449 26. Development of Murine Urinary System. *GUDMAP*

450 https://www.gudmap.org/tutorials/urogenital-dev/urinary-system/.

451 27. Liu, C.-F., Liu, C. & Yao, H. H.-C. Building Pathways for Ovary Organogenesis in the Mouse

452 Embryo. *Curr Top Dev Biol* **90**, 263–290 (2010).

453 28. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**,

454 496–502 (2019).

455 29. Warburton, D. *et al.* Lung Organogenesis. *Curr Top Dev Biol* **90**, 73–158 (2010).

456 30. Wear, H. M., McPike, M. J. & Watanabe, K. H. From primordial germ cells to primordial

457 follicles: a review and visual representation of early ovarian development in mice. *Journal of*

458 *Ovarian Research* **9**, 36 (2016).

459 31. Bouchard, M. Transcriptional control of kidney development. *Differentiation* **72**, 295–306

(2004).

32. Reidy, K. J. & Rosenblum, N. D. Cell and molecular biology of kidney development. *Semin Nephrol* **29**, 321–337 (2009).

33. Miao, Z. *et al.* Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat Commun* **12**, 2277 (2021).

34. Andersson, A. *et al.* A Landmark-based Common Coordinate Framework for Spatial Transcriptomics Data. 2021.11.11.468178 Preprint at https://doi.org/10.1101/2021.11.11.468178 (2021).

35. Peng, H. *et al.* Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021).

36. Fischer, D. S., Schaar, A. C. & Theis, F. J. Learning cell communication from spatial graphs of cells. 2021.07.11.451750 Preprint at https://doi.org/10.1101/2021.07.11.451750 (2021).

37. Wu, F. *et al.* Simplifying Graph Convolutional Networks. Preprint at http://arxiv.org/abs/1902.07153 (2019).

38. Wei, R., Yin, H., Jia, J., Benson, A. R. & Li, P. Understanding Non-linearity in Graph Neural Networks from the Bayesian-Inference Perspective. Preprint at https://doi.org/10.48550/arXiv.2207.11311 (2022).

39. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein Generative Adversarial Networks. in *Proceedings of the 34th International Conference on Machine Learning* 214–223 (PMLR, 2017).

40. Marstal, K., Berendsen, F., Staring, M. & Klein, S. SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration. in *2016 IEEE Conference on Computer Vision and*

Pattern Recognition Workshops (CVPRW) 574–582 (2016). doi:10.1109/CVPRW.2016.78.

41. Besl, P. J. & McKay, N. D. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**, 239–256 (1992).
42. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425–436 (2021).
43. Douglas, J. *et al.* Metrics for Evaluating Network Alignment. 6.
44. Conte, D., Foggia, P., Sansone, C. & Vento, M. THIRTY YEARS OF GRAPH MATCHING IN PATTERN RECOGNITION. *Int J Pattern Recogn* **18**, 265–298 (2004).
45. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
46. González-Blas, C. B. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. 2022.08.19.504505 Preprint at https://doi.org/10.1101/2022.08.19.504505 (2022).

Methods

SLAT framework

Modeling spatial coordinates and omics feature of single cell

We denote a spatial omics dataset as $\mathcal{D} = \{(\mathbf{g}^{(i)}, \mathbf{s}^{(i)}), i = 1,2, \ldots, N\}$, where N is the number of spots or cells, $\mathbf{g}^{(i)} \in \mathbb{R}^G$, and $\mathbf{s}^{(i)} \in \mathbb{R}^2$ are the raw omics features (e.g., genes) and spatial coordinates of cell i, respectively, where G is the number of omics features. For datasets containing non-identical omics features, we use their overlapping features. For ease of notation, we denote the combination of omics features and spatial coordinates across all cells in a dataset as matrices $\mathbf{G} \in \mathbb{R}^{N\times G}$, $\mathbf{S} \in \mathbb{R}^{N\times 2}$, respectively. Subscripts such as $\mathbf{G}_1, \mathbf{G}_2$ and $\mathbf{S}_1, \mathbf{S}_2$ are added to distinguish two datasets being aligned.

In attempts to correct inter-sample batch effects, we employ an SVD-based cross-dataset matrix decomposition strategy as a preprocessing step. To begin with, we denote the log-normalized and scaled omics matrices of two spatial datasets as $\widetilde{\mathbf{G}}_1 \in \mathbb{R}^{N_1\times G}$ and $\widetilde{\mathbf{G}}_2 \in \mathbb{R}^{N_2\times G}$. We then apply SVD on their dot product as follows:

$$\widetilde{\mathbf{G}}_1 \widetilde{\mathbf{G}}_2{}^{\top} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top} \tag{1}$$

Using the decomposed matrices, we obtain the batch-corrected embeddings of the two datasets as:

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{U}_{1:M}\mathbf{\Sigma}_{1:M}{}^{\frac{1}{2}} \\ \mathbf{X}_2 &= \mathbf{V}_{1:M}\mathbf{\Sigma}_{1:M}{}^{\frac{1}{2}} \end{aligned} \tag{2}$$

where $\mathbf{U}_{1:M} \in \mathbb{R}^{N_1\times M}$, $\mathbf{V}_{1:M} \in \mathbb{R}^{N_2\times M}$, $\mathbf{\Sigma}_{1:M} \in \mathbb{R}^{M\times M}$ are truncated decompositions corresponding to the M-largest singular values. Notably, this preprocessing step is flexible and allows for modular input, as other dedicated embedding algorithms like GLUE[23] can be incorporated instead for cross-modality integration.

We model the spatial information of cells in $\mathcal{D}$ as a spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where each node $v_i \in \mathcal{V}$ corresponds to a cell with the batch corrected embedding $\mathbf{x}^{(i)} \in \mathbb{R}^M$ as its node attribute, and

the edges connect the K-nearest neighbors in the spatial space. We also denote the adjacency matrix of the graph as $\mathbf{A} \in \{0,1\}^{N \times N}$. $\mathbf{A}_{i,j} = 1$ if the edge $(v_i, v_j) \in \mathcal{E}$, otherwise $\mathbf{A}_{i,j} = 0$. In particular, for cross-technology alignment, different technologies may have distinct spatial resolutions, in which case we can select different K's for each slice according to its spatial resolution. SLAT also supports building the spatial graph by radius, where all cells located within a specific radius are taken as "neighbors".

SLAT formulates the alignment of two spatial datasets $\mathcal{D}_1 = \left\{\left(\mathbf{x}_1^{(i)}, \mathbf{s}_1^{(i)}\right), i = 1,2, \dots, N_1\right\}$ and $\mathcal{D}_2 = \left\{\left(\mathbf{x}_2^{(i)}, \mathbf{s}_2^{(i)}\right), i = 1,2, \dots, N_2\right\}$ as a minimum-cost bipartite matching problem of their corresponding spatial graphs $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1, \mathbf{X}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2, \mathbf{X}_2)$:

$$\min_{\mathcal{M}} \sum_{(v_i, v_j) \in \mathcal{M}} \left\| \mathbf{z}_1^{(i)} - \mathbf{z}_2^{(j)} \right\|, \quad \text{s.t. } \mathcal{M} \subset \mathcal{V}_1 \times \mathcal{V}_2, |\mathcal{M}| = m, \tag{3}$$

where $\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(j)} \in \mathbb{R}^P$ are node embeddings of v_i and v_j in $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively, $\mathcal{M}$ is a set of matches of fixed size, and P is the dimensionality of node embeddings. Similarly, we use matrices $\mathbf{Z}_1 \in \mathbb{R}^{N_1 \times P}$ and $\mathbf{Z}_2 \in \mathbb{R}^{N_2 \times P}$ to denote the combination of cell embeddings of all cells in the two datasets.

It has been demonstrated in a previous work, that the above matching problem is equivalent to minimizing the Wasserstein distance between node embeddings from different graphs[17]. SLAT follows the same approach with adaptations for spatial omics data. Below, we explain how node or cell embeddings $\mathbf{Z}_1, \mathbf{Z}_2$ can be obtained and optimized for spatial graph alignment.

Construction of holistic cell representations

An accurate alignment of spatial omics datasets should align cells that are similar in both the molecular modality and the spatial context. In particular, the spatial context can involve various resolutions, ranging from microenvironments to global positions within the tissue. Inspired by previous work[17,37,38], we first employ the lightweight graph-convolutional network (LGCN) to derive

a holistic cell representation with all such information integrated for each dataset. A LGCN propagates and aggregates information along the spatial graph through stepwise concatenations:

$$\widetilde{\mathbf{X}} = f_{\mathrm{LGCN}}(\mathbf{A}, \mathbf{X}) = \mathrm{Concat}\left(\mathbf{X}, \widehat{\mathbf{A}}\mathbf{X}, \widehat{\mathbf{A}}^2\mathbf{X}, \ldots, \widehat{\mathrm{A}}^L\mathbf{X}\right), \tag{4}$$

where $\widehat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}$, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\widetilde{\mathbf{D}}$ is the diagonal degree matrix of $\widetilde{\mathbf{A}}$, and L is the maximal number of steps. The resulting cell representation $\widetilde{\mathbf{X}} \in \mathbb{R}^{N\times(L+1)M}$ is a concatenation of multi-level information. The first M dimensions correspond to no graph propagation, which is simply a copy of the omics data $\mathbf{X}$. The second M dimensions correspond to one-step graph propagation, reflecting the composition of a cell's immediate neighbors, which form its microenvironment. The information coarsens as the number of steps increases, gradually becoming a representation of rough locations within the tissue. Thus, $\widetilde{\mathbf{X}}$ contains informative features for spatial alignment at multiple levels of spatial context.

The cell representation $\widetilde{\mathbf{X}}$ is constructed separately for each dataset by using dataset-specific omics profiles and adjacency matrices:

$$\begin{aligned} \widetilde{\mathbf{X}}_1 &= f_{\mathrm{LGCN}}(\mathbf{A}_1, \mathbf{X}_1), \\ \widetilde{\mathbf{X}}_2 &= f_{\mathrm{LGCN}}(\mathbf{A}_2, \mathbf{X}_2). \end{aligned} \tag{5}$$

Adversarial graph alignment

Based on the holistic cell representations $\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2$ described above, we use adversarial alignment to learn cell embeddings $\mathbf{Z}_1, \mathbf{Z}_2$ that minimize the Wasserstein distance for graph matching[17]. Specifically, we apply a multilayer perceptron denoted as f_{Z} to mitigate systematic bias in $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{X}}_2$ that may arise from differences in omics distribution or spatial topology across datasets:

$$\begin{aligned} \mathbf{Z}_1 &= f_{\mathrm{Z}}(\widetilde{\mathbf{X}}_1), \\ \mathbf{Z}_2 &= f_{\mathrm{Z}}(\widetilde{\mathbf{X}}_2). \end{aligned} \tag{6}$$

We then introduce the Wasserstein discriminator f_{D}, which uses $\mathbf{Z}_1, \mathbf{Z}_2$ as input and tries to

maximize the following Wasserstein loss L_{W} to estimate the Wasserstein distance:

$$L_{\mathrm{W}} = \frac{1}{|\mathcal{V}_1|}\sum_{v_i \in \mathcal{V}_1} f_D\left(\mathbf{z}_1^{(i)}\right) - \frac{1}{|\mathcal{V}_2|}\sum_{v_j \in \mathcal{V}_2} f_D\left(\mathbf{z}_2^{(j)}\right), \tag{7}$$

where $\mathbf{z}_1^{(i)}$ is a row in $\mathbf{Z}_1$ corresponding to cell v_i, and $\mathbf{z}_2^{(j)}$ is a row in $\mathbf{Z}_2$ corresponding to cell v_j. The transformation f_{Z} can then be adversarially trained to minimize (7), for aligning the distribution of cell embeddings in the two datasets properly. However, different single-cell spatial datasets may contain different cell-type proportions or distinct spatial regions. It is thus unreasonable to assume identical distribution of their cell embeddings as assumed in the standard scheme described above[39]. Inspired by a previous study[17], we use the output of the Wasserstein discriminator f_D as a dynamic clipping criterium to select $l \times N_1$ and $l \times N_2$ cells from the two datasets with minimum Wasserstein distance for adversarial training (Extended Data Fig. 1b):

$$\begin{aligned} \mathcal{V}_1' &= \underset{v_i \in \mathcal{V}_1' \subseteq \mathcal{V}_1}{\arg\, k\min} f_{\mathrm{D}}\left(\mathbf{z}_1^{(i)}\right), \quad |\mathcal{V}_1'| = l \times N_1, \\ \mathcal{V}_2' &= \underset{v_j \in \mathcal{V}_2' \subseteq \mathcal{V}_2}{\arg\, k\max} f_{\mathrm{D}}\left(\mathbf{z}_2^{(j)}\right), \quad |\mathcal{V}_2'| = l \times N_2, \end{aligned} \tag{8}$$

where l is a hyperparameter between 0 and 1. These cells correspond to the most reliable anchors to guide the alignment. The Wasserstein discriminator loss L_{W} is then modified accordingly as follows:

$$L_{\mathrm{W}} = \frac{1}{|\mathcal{V}_1{}'|}\sum_{v_i \in \mathcal{V}_1{}'} f_{\mathrm{D}}\left(\mathbf{z}_1^{(i)}\right) - \frac{1}{|\mathcal{V}_2{}'|}\sum_{v_j \in \mathcal{V}_2{}'} f_{\mathrm{D}}\left(\mathbf{z}_2^{(j)}\right). \tag{9}$$

This approach ensures that distinct regions across two spatial datasets will not be forcibly aligned. The results show that SLAT performs best when $l = 0.6$, although the performance depends only weakly on l (Extended Data Fig. 2a).

To avoid a degenerate solution where all embeddings collapse to a singular point, we adopt an additional reconstruction term to ensure that the embeddings have sufficient information to

reconstruct input, essentially enhancing model stability. We use a simple multilayer perceptron network denoted as f_{R} for data reconstruction, making the following reconstruction loss:

$$L_{\mathrm{R}} = \frac{1}{|\mathcal{V}_1|} \sum_{v_i \in \mathcal{V}_1} \left\| f_{\mathrm{R}}\left(\mathbf{z}_1^{(i)}\right) - \mathbf{x}_1^{(i)} \right\| + \frac{1}{|\mathcal{V}_2|} \sum_{v_j \in \mathcal{V}_2} \left\| f_{\mathrm{R}}\left(\mathbf{z}_2^{(j)}\right) - \mathbf{x}_2^{(j)} \right\|. \tag{10}$$

Overall objective

Finally, the overall training objective of SLAT can be summarized as follows:

$$\max_{f_{\mathrm{D}}} \alpha \cdot L_{\mathrm{W}}, \tag{11}$$

$$\min_{f_{\mathrm{Z}}, f_{\mathrm{R}}} \alpha \cdot L_{\mathrm{W}} + (1-\alpha) \cdot L_{\mathrm{R}}, \tag{12}$$

where L_{W} and L_{R} are defined by Eqs. (9) and (10), respectively, α is a hyperparameter balancing the contribution of adversarial alignment and data reconstruction. We use stochastic gradient descent (SGD) with the Adam optimizer to train the SLAT model.

Coordinate matching

Apart from the core model described above, we also provide options to use additional information to match the spatial coordinates $\mathbf{S}$ of two slices which can help distinguish symmetric structures (e.g., left and right hemispheres of the brain) and improve the final matching quality. The goal of coordinate matching is to roughly align different slices in terms of their overall direction by estimating an affine transformation matrix $\mathbf{M}$. The exact strategy depends on the type of information available.

First, guided by expert knowledge of tissue structures, $\mathbf{M}$ can be computed by combining the scaling, rotation, and translation operations required to obtain a rough alignment between the two slices. Second, if imaging data like H-E staining images are available, $\mathbf{M}$ can be estimated following our tutorial based on SimpleElastix[40], which is a state-of-the-art medical image registration

tool. Finally, if no other information is available, we also provide a default solution based on iterative closet point (ICP)[41], a point-cloud registration algorithm, where we treat the spatial datasets as point clouds on a two-dimensional plane and uses geometric features for registration. With the obtained $\mathbf{M}$ matrix, we consort the coordinates of the two datasets by the following transformation:

$$\mathbf{S}_2' = \mathbf{S}_2 \cdot \mathbf{M}. \tag{13}$$

Cell matching and quality assessment

With the cell embeddings $\mathbf{Z}_1, \mathbf{Z}_2$ learned by the SLAT core model and the matched spatial coordinates $\mathbf{S}_1$ and $\mathbf{S}_2'$, we match dataset $\mathcal{D}_2$ with $\mathcal{D}_1$ using the following strategy: For cell j in dataset $\mathcal{D}_2$, SLAT first selects the top K most similar $\mathcal{D}_1$ cells in the embedding space $\mathbf{Z}$ as a candidate set $\mathcal{C}_j$, then chooses the one with minimal distance in matched spatial coordinates $\mathbf{S}$ as the final match. K is a hyperparameter that can be adjusted according to the confidence in the coordinate matching result.

$$\mathcal{C}_j = \underset{v_i \in \mathcal{C}_j \subseteq \mathcal{V}_1}{\arg\, k\max} \cos\left(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(j)}\right), \left|\mathcal{C}_j\right| = K, \tag{14}$$

$$\mathrm{match}(v_j) = \underset{v_i \in \mathcal{C}_j}{\operatorname{argmin}} \left\| \mathbf{s}_1^{(i)} - \mathbf{s}_2'^{(j)} \right\|, \tag{15}$$

where $\mathbf{s}_1^{(i)}$ and $\mathbf{s}_2'^{(j)}$ are rows in the coordinate matrices $\mathbf{S}_1$ and $\mathbf{S}_2'$, respectively. The cosine similarity $\cos\left(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(j)}\right)$ is used to quantify the matching quality. We filter out unreliable matches by defining a cutoff for cosine similarity. Cell pairs with low matching quality suggest either biologically distinct changes or technological artifacts and should be carefully checked (see Results). Empirically, we introduce a default cutoff at 0.8 which can also be adjusted in a case-by-case manner.

Systematic benchmarks

Benchmark datasets

We selected 10x Visium, MERFISH and Stereo-seq as representative spatial technologies for benchmarking alignment methods. The 10x Visium dataset comes from consecutive slices of human dorsolateral prefrontal cortex, containing about 3,000 spots per slice, each spanning 50 μm with over 20,000 genes detected[42]. The MERFISH dataset comes from consecutive slices of mouse hypothalamic preoptic, with subcellular resolution but only 151 genes detected in total[22]. The Stereo-seq dataset comes from consecutive slices of an E15.5 mouse embryo, containing over 100,000 single cells per slice, divided into 25 cell types with complex spatial organization, and detects over 20,000 genes in total[8]. For each technology, we chose the first two slices if the dataset contains more than two slices. Cell type annotation and tissue region segmentation were obtained from the original authors whenever possible. Since the Stereo-seq dataset does not provide tissue segmentation, we segmented the most prominent regions in the embryo, including Brain, Jaw and face, Spinal cord, Heart, Lung, Liver, and Belly under the guidance of an expert in mouse anatomy. The spatial segmentation of the MERFISH dataset is provided in publication figures but not in raw data, so we re-segmented the data as guided by the figures.

Slice processing

For each technology, we first removed the cells/spots that are unannotated in both slices, then rotated the second slice with a random angle before feeding to the alignment methods.

Benchmarked methods

The benchmarked methods Harmony, PASTE, STAGATE, and SLAT were executed using the Python packages “harmonypy” (v0.0.6), “paste-bio” (v1.3.0), “STAGATE_pyG” (latest commit 8b9c8ef), and “scSLAT” (v0.2.0), respectively, in Python (v3.8). Seurat was executed using the R package “Seurat” (v4.1.1) in R (v4.1.3). For each method, we used the default hyperparameter settings and data preprocessing steps recommended by the original authors.

Evaluation metrics

Given that an accurate spatial alignment should align cells similar in both molecular modality and spatial context, we devised a specialized composite metric called the cell type and region matching index (CRI) that reflects alignment accuracy in both aspects. Suppose that we match dataset $\mathcal{D}_2$ to $\mathcal{D}_1$ and the datasets contain N_2 and N_1 cells, respectively. We denote the cell type and spatial region of cell i in $\mathcal{D}_1$ as $c_1^{(i)}$ and $r_1^{(i)}$, respectively, and denote the cell type and spatial region of cell j in $\mathcal{D}_2$ as $c_2^{(j)}$ and $r_2^{(j)}$, respectively. The CRI is defined as

$$\mathrm{CRI} = \frac{1}{|\mathcal{M}|} \sum_{(v_i, v_j) \in \mathcal{M}} I(i,j)\,, \tag{16}$$

$$I(i,j) = \begin{cases} 1, & c_1^{(i)} = c_2^{(j)} \text{ and } r_1^{(i)} = r_2^{(j)} \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

where $\mathcal{M} = \{(\mathrm{match}(v_j), v_j),\ j = 1,2,\dots,N_2\}$ is the set of matched cell pairs.

In evaluating 3D reconstruction, we used the region accuracy defined in PASTE[15]. Compared to CRI, it only considers spatial regions but not cell types:

$$\text{Region accuracy} = \frac{1}{|\mathcal{M}|} \sum_{(v_i, v_j) \in \mathcal{M}} I(i,j)\,, \tag{18}$$

$$I(i,j) = \begin{cases} 1, & r_1^{(i)} = r_2^{(j)} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Both CRI and region accuracy rely on high-level annotations and thus loses single-cell resolution. As an alternative metric that retains single-cell resolution, we used the Euclidean distance between matched cells as measured in the original coordinates before rotation and shifting:

$$\text{Euclidean distance} = \frac{1}{|\mathcal{M}|} \sum_{(v_i, v_j) \in \mathcal{M}} \left\| \mathbf{s}_1{}^{(i)} - \mathbf{s}_2{}^{(j)} \right\|, \tag{20}$$

Meanwhile, we also noticed that there exists other metrics commonly used for evaluating graph alignment, e.g., the edge score[43,44].

However, we did not use it as we found that it can be deceived in certain situations. For example, Extended Data Fig. 1c shows two graphs with known ground truth node pairing information. "Alignment 1" is the correct matching (grey lines) and "Alignment 2" has four mismatched pairs (highlighted by red lines), but the two alignments get the same edge score.

Benchmark tasks

We benchmarked the alignment methods based on the following four tasks: (1) real world alignment, (2) split slice alignment, (3) 3D reconstruction, (4) scalability test:

In real world alignment, the slices to be aligned are different slices produced from the same position of the same tissue[5,8,42] (Supplementary Fig. 1). Each algorithm was run eight times with different model random seeds. Performance of the alignment methods is quantified by the CRI metric defined in Eq. (16) and the Euclidean distance metric defined in Eq. (20).

In split slice alignment, we split the first slice of each technology used in real world alignment into two pseudo-slices of equal size by randomly sampling the cells without replacement (Supplementary Fig. 3). Each algorithm was run eight times with different model random seeds. Performance quantification is the same as in real world alignment.

3D reconstruction is carried out with the same 10x Visium datasets (Supplementary Fig. 4) used in PASTE. Specifically, this dataset includes three groups of 10x Visium slices generated from human dorsolateral prefrontal cortex, each containing three consecutive slices from same tissue. We ran each algorithm eight times with different model random seeds. Performance is also quantified with the same metric used in PASTE, i.e., region accuracy defined in Eq. (18).

For scalability benchmark, we randomly subsampled the Stereo-seq dataset used in real world benchmark to a range of cell numbers (3200, 6400, 12800, 25600, 51200, 102400). The subsampling process was repeated eight times with different random seeds. In addition to CRI and Euclidean distance, we also report the run time of each method.

Benchmark workflow

We used Snakemake (v7.12.0) to manage the whole benchmark workflow. All benchmarked methods were allocated 16 cores of Intel Xeon Platinum 8358 CPU, 128 GB of RAM, and a NVIDIA A100 GPU with 80 GB VRAM by the Slurm workload management system.

Hyperparameter robustness

We tested SLAT's robustness to key hyperparameters including: (1) neighbor size K for the spatial graph, (2) input embedding methods, (3) number of training epochs, (4) number of LGCN layers L, (5) MLP hidden layer dimension, (6) dimension P of SLAT embedding, (7) weight α of the adversarial loss, (8) dynamic clipping ratio l, (9) learning rate of the SLAT model.

We run SLAT in two homogeneous Stereo-seq datasets from the same E15.5 mouse embryo, each containing more than 100,000 cells so we randomly subsampled 8,000 cells from each dataset to save time. Results are reported in Extended Data Fig.2a in the CRI metric (defined in below) and Euclidean distance metric (defined in below). Every experiment was run 8 times with different model random seeds.

Robustness to noise in the spatial graph

In practice, the spatial graph may be imperfect due to technical limitations. Therefore, we tested SLAT's robustness to graph corruption in the same dataset used in hyperparameter robustness evaluation. Specifically, we randomly masked the edges in the graph by increasing ratios (from 0.1

and 0.9). Results are reported in Fig. 2c in the CRI metric (defined in below) and Euclidean distance metric (defined in below). Every experiment was run 8 times with different masking random seeds.

Heterogeneous alignment across distinct technologies and modalities

Visium and Xenium data alignment

The Visium and Xenium datasets were generated from consecutive slices of the human breast cancer tissue sample[4]. In order to maintain the comparability with the original paper, we chose the exact same slices used in their analysis (see Fig. 4c of ref[4]). We further selected the shared region between the two slices as the original authors reported[4] for follow-up analysis.

Considering that the Xenium slice contains more than 100,642 cells while Visium only contains less than 3,841 spots in same physical region, we used different neighbor sizes proportional to cell density ($K = 5$ for Visium, and $K = 130$ for Xenium) when constructing the spatial graphs, in order to ensure that the GCNs have similar spatial receptive fields. SLAT is then run with otherwise default parameters. We selected Visium triple positive spots based on the number of aligned Xenium triple positive cells (Supplementary Fig. 5b). 7 spots with more than 2 aligned cells were chosen. Following the original paper, we did differential gene expression analysis of the SLAT identified and manually curated triple positive spots against all other spots, respectively, using the Scanpy[45] function "scanpy.tl.rank_genes_groups" with parameter "methods='wilcoxon'". For comparison, all methods included in the benchmarks (PASTE, Harmony, Seurat, and Harmony) were applied to the same Visium and Xenium datasets with their default parameters.

For comparison, all methods benchmarked (PASTE, Harmony, Seurat, and Harmony) were applied to the same datasets with their default parameters (Extended Data Fig. 4).

SeqFISH data and Stereo-seq data alignment

For Stereo-seq and seqFISH, we chose the E9.5 slice with the most complete cell type annotation by

768 the original authors (i.e., the slices with the least proportion of unannotated cells), respectively. To
769 align the chosen seqFISH and Stereo-seq slices, we run SLAT using the default parameters and
770 filtered out alignment pairs with similarity scores lower than 0.8. We next refined cell type
771 annotations in the Stereo-seq dataset based on the higher resolution annotation in seqFISH through
772 label transfer for "Neural crest" cells (Fig. 4b). We also manually annotated "Neural crest" cells for
773 independent validation. Scanpy[45] was used for the analysis following its official tutorial: the data
774 were log-normalized using the functions "sc.pp.normalize_total" and "sc.pp.log1p". Highly variable
775 genes were identified with "sc.pp.highly_variable_genes". The first 50 principal components after
776 PCA ("sc.tl.pca") were used to generate neighborhood graphs ("sc.pp.neighbors") for computing
777 UMAP embeddings ("sc.tl.umap") and Leiden clustering ("sc.tl.leiden"). The "Neural crest" cells
778 were annotated via marker genes of different germ layers (*Foxc2* and *Vcan* for mesoderm, *Msx1*, *Mif*,
779 and *Dik1* for ectoderm, see Extended Data Fig. 3d).
780
781 For comparison, all methods benchmarked (PASTE, STAGATE, Seurat and Harmony) were applied
782 to the same datasets with their default parameters (Extended Data Fig. 5).
783
784 **Spatial-ATAC-seq and Stereo-seq data alignment**
785 For cross-modality alignment, the slices to align were two 11.5-day mouse embryo datasets from
786 Stereo-seq (RNA) and spatial-ATAC-seq (ATAC), respectively. For Stereo-seq, we chose the E11.5
787 slice with the most complete cell type annotation by the original authors (i.e., the slice with the least
788 proportion of unannotated cells). For spatial-ATAC-seq, we chose the E11.5 slice with the highest
789 spatial resolution (20 μm). Given that the current spatial-ATAC-seq data did not cover the entire
790 embryo due to technical limitations, we extracted the anatomically corresponding regions from the
791 Stereo-seq dataset under expert guidance.
792
793 The SVD-based cross-dataset matrix decomposition strategy cannot handle different modalities with
794 disjoint feature spaces. Thanks to SLAT's modular design, we replaced SVD with GLUE[23], a unified
795 multi-modality embedding framework, to project cells from different modalities into a shared latent

space. We then ran SLAT with the default hyperparameters. Based on SLAT alignment, we transferred cell type labels from Stereo-seq to spatial-ATAC-seq which was not annotated, and then applied SCENIC+[46] for joint regulatory inference.

To compare with benchmarked methods (PASTE, Harmony, Seurat and Harmony), we also used GLUE embedding as their input followed by their default pipeline. Exceptions are Seurat and STAGATE, which do not support low dimensional embeddings as input, thus cannot be compared. In addition, we also compared with the GLUE embeddings directly (Supplementary Fig. 6).

Spatial-temporal alignment

We applied SLAT to align E11.5 and E12.5 heterogeneous mouse embryo Stereo-seq datasets with the default hyperparameters. In order to maintain the comparability with the original paper, we chose the exact same slices used in their analysis (see Fig. 3a of ref[8]). Regions with lower SLAT similarity scores were marked in Fig. 5b. To focus on kidney development, we extracted cells labeled as "Kidney" in E12.5 and their aligned cells in E11.5, then clustered these cells by using the standard Scanpy clustering pipeline mentioned above and annotated them via well-defined kidney markers[32,33] (*Osr1*, *Foxc1* and *Podxl* for nephron progenitors; *Uncx*, *Nr2f2*, *Dach*, *Wt1*, *Nphs1*, and *Cd44* for kidney; see Fig. 5f). To further demonstrate the robustness of SLAT, we rerun the analysis against two additional slices randomly chosen from E11.5 and E12.5, and obtained similar results (Supplementary Fig. 7).

For comparison, all methods benchmarked (PASTE, STAGATE, Seurat and Harmony) were applied to the same datasets with their default parameters (Extended Data Fig. 8).

Data availability

All datasets used in this study were already published and were obtained from public data

repositories. See Supplementary Table 1 for detailed information on the entire datasets used in this study, including publication and downloading URLs.

Code availability

The SLAT framework was implemented in the "scSLAT" Python package, which is available at https://github.com/gao-lab/SLAT. For reproducibility, the scripts for all benchmarks were assembled using Snakemake (v7.12.0), which is also available in the above repository.