# Spatial Transcriptomic Alignment, Integration, and de novo 3D Reconstruction by STAIR

**Yuanyuan Yu**
Sun Yat-sen University
**Zhi Xie** ( xiezhi@gmail.com )
Sun Yat-sen University

**Method Article**

**Keywords:** Spatial transcriptome, de novo 3D reconstruction, Alignment, Integration, Heterogeneous graph attention network

**Posted Date:** February 9th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-3939678/v1



**Additional Declarations:** The authors declare no competing interests.

# Spatial Transcriptomic Alignment, Integration, and *de novo* 3D Reconstruction by STAIR

Yuanyuan Yu[1], Zhi Xie[1*]

[1] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

[*] Email: xiezhi@gmail.com

## Abstract

Spatial transcriptomics (ST) has emerged as a transformative approach for comprehending tissue architecture with molecular profiles. However, amalgamating discrete two-dimensional (2D) ST snapshots into a unified 3D atlas remains an outstanding challenge. To this end, we introduce STAIR, an end-to-end solution for ST alignment, integration, and *de novo* 3D reconstruction. STAIR uses a heterogeneous graph attention network with spot-level and slice-level attention mechanisms to obtain a unified embedding space and guide 3D space reconstruction in an unsupervised manner. We demonstrate STAIR's marked improvements in slice alignment and integration across samples and platforms over the previous methods. Furthermore, STAIR shows first-of-its-kind performance in *de novo* 3D reconstruction, with demonstrations in mouse hypothalamus preoptic area, mouse brain, and breast tumor tissue, which provides precise delineation of brain regions and reveals tumor progression in 3D space. Additionally, STAIR integrates additional slices into the existing 3D atlas incorporating both molecular features and physical coordinates. STAIR is the first to address the core obstacles limiting 3D positioning and harnessing alignments for atlas construction and assimilation. It lays a computational foundation to construct unified tissue maps and provides novel biological insights from a 3D perspective.

## Keywords

Spatial transcriptome, *de novo* 3D reconstruction, Alignment, Integration, Heterogeneous graph attention network

# Introduction

Recently remarkable advancements have enabled the generation of spatial transcriptomic (ST) data that captures gene expression with preserved spatial context. Various sequencing techniques[1-7] facilitate exploring molecular features within tissues from a spatial perspective. To study an organ or tissue in three-dimensional (3D) space, researchers sampled parallel slices of specific samples at certain distance intervals, where individual slices capture spatially resolved molecular features in a single plane[8-10]. At present, construction of ST-based molecular atlases is ongoing, such as mouse brain[8], macaque brain[10], and drosophila embryo[9]. By revealing topological structures organ-wide, these studies advance our comprehension of molecular drivers of tissue organization.

Several computational methods including PASTE[11], PRECAST[12], STAligner[13] and STitch3D[14] have enabled important advancements in handling multi-slice ST data. Specifically, PRECAST[12] introduced a novel probabilistic approach mainly focusing on spatial embedding integration. PASTE[11] proposed an innovative method for 2D coordinate alignment by balancing expression and physical proximity to obtain optimal transport between slices. STAligner[13] made pioneering work in the unified alignment of both embeddings and coordinates. It first employed graph attention networks with triplet loss for integrating spatial embeddings, followed by aligning 2D coordinates based on the selection of landmark domains that are uniformly shaped across slices. While most methods aligning physical space perform only 2D coordinate registration, more recently, STitch3D[14] pioneered intriguing 3D analytical capability by jointly modeling multiple slices. It reconstructed 3D views of tissue structures, cell distributions, and developmental trajectories with the requirement of additional single-cell data.

The previous methods such as STaligner and STitch3D utilized known inter-slice distance or 3D coordinates for 3D reconstruction. However, this information can be unavailable or unreliable when integrating across-sample, across-study sections with non-standardizable slice distances, severely limiting multi-source 3D assembly. Therefore, computed inference of inter-relationship positioning is imperative to enable coherent fusion. Regrettably, current solutions do not furnish this capability. In addition, existing reconstructed atlases also have minimal capacity for quantitative annotation transfer or assimilation of emerging data, impeding continued knowledge accumulation

within fixed spatial references.

To address these challenges, we developed STAIR, an integrated solution for end-to-end alignment, integration, and *de novo* 3D construction. STAIR utilizes heterogeneous graph attention network[15] to learn spatial embedding, and completes the reconstruction of physical location based on this. Unlike previous methods relying on known slice distance or known 3D coordinates, STAIR requires only ST data as input and infers the relative positioning of slices along z-axis in a completely unsupervised manner. In addition, STAIR seamlessly integrates new slices into the existing 3D atlas, expanding and updating the reference 3D atlas. We demonstrated STAIR's superior performance over existing methods for spatial features and 2D coordinates alignment across various datasets. Furthermore, STAIR was the first-of-its-kind method for *de novo* 3D reconstruction, with demonstrations of mouse hypothalamic preoptic area, whole mouse brain, and tumor tissue. Finally, STAIR achieved seamless integration of a new slice from a different ST platform into a constructed 3D atlas, accurately transferring the annotation information in the atlas to the new slice. STAIR is publicly available as an open-source Python package at https://github.com/yuyuanyuana/STAIR.

# Results

### Overview of the STAIR framework

STAIR achieves integration and alignment of molecular features and physical coordinates for ST data, enabling 3D reconstruction and assimilation of new slices into the reference atlases (Fig. 1). It takes only ST data as input and outputs aligned spatial embeddings and coordinates. It comprises two central modules (Fig. 1A), STAIR-Emb performs embedding alignment, while STAIR-Loc handles coordinate registration. Spatial domains and developmental trajectories are derived solely from the aligned embeddings via clustering and trajectory inference, respectively. Thus, relying on multi-slice data alone, STAIR is capable of *de novo* 3D reconstruction and further obtaining discrete organizational structures and continuous dynamic relationships in 3D space.

For embedding alignment (Fig. 1A, left), an autoencoder with batch factor[16,17] performs

nonlinear dimensionality reduction to account for platform-specific effects in gene expression. STAIR-Emb then constructs a heterogeneous graph with spots across all slices as nodes, and the node attribution is decided by their original slices. Connectivity for spots in the same slice is defined by spatial neighborhoods while inter-slice edges are weighted by gene expression affinities from the encoder output since the relative positions between slices are unknown. Subsequently, STAIR-Emb employs an attention mechanism[15] consisting of spot-level and slice-level attention for adaptive information aggregation within and between slices, obtaining low-dimensional spatial features of spots that encapsulate rich biological signals. In parallel, the attention score produces inter-slice relationship matrices reflecting higher-order tissue correspondences.

For coordinate registration (Fig. 1A, right), STAIR-Loc implements a two-step procedure between slice pairs. In the initial alignment, rotation, scaling, and translation guided by anchor correspondences in the embedding produce an approximate overlay. In the fine alignment, STAIR-Loc employs the Iterative Closest Point (ICP)[18] algorithm for precise registration based on boundary points of the slices and the most aggregated domain. By sequential application across ordered slices, this procedure reconstructs a stacked 3D physical map.

For a fully unsupervised 3D reconstruction from an arbitrary set of parallel slices (Fig. 1B), z-axis coordinates are first reconstructed from inter-slice attention scores derived from STAIR-Emb using a minimum spanning tree (MST)[19]. These predicted z-coordinates are then used to guide x-axis and y-axis alignment using STAIR-Loc. Post-reconstruction, seamlessly assimilating additional acquisitions is also enabled (Fig. 1C). The z-coordinate of a new slice is initially predicted by leveraging attention-based inter-slice proximity scoring to identify the nearest neighboring atlas slices for weighted positional estimation. The most proximal atlas slice serves as a spatial template to scale and align the x- and y-coordinates of the new data. At this stage, annotated atlas information like standard anatomical region labels can be accurately propagated to newly integrated slices. This continual expansion over time enhances the utility and applicability of an established 3D reference.

**STAIR effectively integrates heterogeneous ST data in spatial embeddings**

We first quantitatively evaluated STAIR's efficacy in integrating spatial embeddings and identifying unified spatial domains across diverse tissue sections, a prerequisite for downstream analysis.

Effective integration should identify shared signals across specimens while retaining biologically unique variations within each individual sample.

Our initial testing utilized 10X Visium dataset derived from the human dorsolateral prefrontal cortex (DLPFC)[20], spanning three samples with four sequential sections per sample. The original study performed precise manual annotation, delineating white matter (WM) and six gray matter layers ranging from Layer 1 to Layer 6 to provide ground truth labels (Fig. 2A). First, we conducted separate spatial domain identification from slices of each sample, which had close locations with some differences. STAIR achieved the most accurate domain division results on all three samples (Fig. 2B, Supplementary Fig. S1), with respective median Adjusted Rand Index (ARI) values of 0.60, 0.53, 0.62 (Fig. 2C).

Then, we processed twelve slices from three samples simultaneously (Fig. 2D; Supplementary Fig. S2A). STitch3D was excluded from this test due to its requirement for 3D coordinates to handle multiple slices, which were not available. Despite the challenge, STAIR maintained the highest consistency with annotations, achieving a median ARI value of 0.65, far exceeding the second-ranked STAligner with a median value of 0.46 (Fig. 2C, right). Notably, simultaneous consideration of three samples proved superior to evaluating each sample individually, resulting in higher ARI and a clearer demarcation between Layer 4, Layer 5, and Layer 6 (Fig. 2B to D). Subsequently, we performed low-dimensional visualization using uniform manifold approximation and projection (UMAP)[21] based on the spatial embeddings derived by these methods (Supplementary Fig. S2B). In the STAIR-based visualization, all spots exhibited an arrangement according to the layers, with thorough mixing between the different samples. In contrast, both STAligner and PRECAST lacked clarity in arranging and distinguishing these known layers. Specifically, STAligner mixed Layer 2 and Layer 3, as well as Layer 4 to Layer 6, while PRECAST only distinguished WM. We also quantified the effects of spatial embedding learning and integration using Average Silhouette width (ASW). We calculated ASW for spatial embeddings with respect to spatial domains ($ASW_{domain}$) and samples ($ASW_{batch}$), as well as $ASW_{F1}$ to evaluate overall capability (Methods). Higher $ASW_{F1}$ and $ASW_{domain}$ coupled with lower $ASW_{batch}$ indicates better performance. STAIR achieved best spatial embedding learning and integrating with the highest $ASW_{F1}$ (Supplementary

Fig. S2C). STAligner also adequately integrated samples, as indicated by an $ASW_{batch}$ value similar to STAIR. However, it had weaker feature learning capabilities with a much lower $ASW_{domain}$, which was consistent with the unclear UMAP pattern we observed previously.

Moreover, to evaluate cross-platform integration, we applied STAIR to integrate mouse olfactory bulb data from Stereo-seq[22] and Slide-seqV2[23], which differed in resolution and area. Stereo-seq covered the main olfactory bulb (MOB) at sub-single-cell resolution, whereas Slide-seqV2 encompassed both MOB and the accessory olfactory bulb (AOB) with a resolution of 10 μm, approximating single-cell. STAIR clearly delineated the MOB region common to both datasets and the AOB region unique to Slide-seqV2 (Fig. 2E, top), highly consistent with standard Allen Brain Atlas (ABA) partitioning (Supplementary Fig. S2D). The AOB region, situated in the middle and upper part of the olfactory bulb, comprised two sub-regions: AOBmi and AOBgr. The MOB region encompassed seven sub-regions arranged in a concentric ring: Rostral migratory stream (RMS), Granule cell layer (GCL), two Mitral cell layers (MCLs), External plexiform layer (EPL), Glomerular layer (GL), and Olfactory nerve layer (ONL). In contrast, STAligner struggled to identify the RMS layer and confused the EPL and GL layers of Stereo-seq (Fig. 2E, middle). PRECAST failed to identify coherent spatial patterns (Fig. 2E, bottom). UMAP visualization further showed STAIR effectively distinguished between AOB and MOB in low-dimensional space, preserving the AOBmi and AOBgr sub-regions in Slide-seqV2 data (Fig. 2F, G). Simultaneously, STAIR seamlessly integrated the shared MOB region, arranging sub-layers consistent by physical locations. Conversely, while STAligner achieved integration to a certain extent, it failed to discern the distinct difference between AOB and MOB. PRECAST completely mixed the datasets, losing Slide-seqV2 specificity.

In summary, STAIR successfully integrated the spatial embeddings across slices within samples, across samples, and across platforms, ensuring a consistent spatial region division and preserving unique biological variations.

**Precise alignment of 2D coordinates across slices by STAIR**

A single ST sequencing only acquires data from one slice, resulting in the loss of the unified physical space across multiple slices. Aligning spatial embeddings across diverse ST slices enables unified

molecular feature space that encapsulates spatial information. In this section, we evaluate STAIR's ability in aligning and integrating ST slices in 2D coordinates (x- and y-axis).

We utilized 12 MERFISH slices of the preoptic area of the mouse hypothalamus[2] with known 3D positions. To test the 2D position alignment of slices, we introduced random rotations and translations to 11 slices while keeping the first slice fixed (Fig. 3A). Subsequently, we employed STAIR, PASTE, and STitch3D on the rotated slices and assessed their effectiveness (Fig. 3B). STAIR stood out as the most precise alignment, closely recovering actual coordinates. In contrast, STitch3D and PASTE struggled with accurate alignment. Quantitatively, the median rotation and translation errors (Methods) for STAIR were 0.03 and 0.04 mm (Fig. 3C). Similar results were obtained across 500 random simulations (Supplementary Fig. S3A). In contrast, PASTE exhibited rotation and translation errors of 1.53 and 1.81, while STitch3D had errors of 2.0 and 2.1, respectively (Fig. 3C).

We further assessed STAIR's robustness on the alignment of 2D coordinates. First, we examined the impact of resolution by aggregating neighboring cells into virtual lower resolution spots (Supplementary Fig. S3B). With 2-5 aggregated cells per spot, STAIR maintained low median rotation errors of 0.02-0.07 and translation errors of 0.04-0.10 (Supplementary Fig. S3C). In contrast, PASTE had errors between 1.54-1.55 for rotation and 1.78-1.80 for translation, while STitch3D ranged from 1.23-1.93 and 1.78-2.25, respectively. Additionally, given that fine alignment relied on spatial region information, we examined performance across 8-15 spatial regions (Supplementary Fig. S3D). STAIR demonstrated stable alignment, with median rotation errors of 0.02-0.08 and translation errors of 0.03-0.09.

In short, ground truth hypothalamus data confirmed STAIR's capabilities for precise 2D slice alignment and robustness across varied resolution and spatial domains.

**Construction of *de novo* 3D atlas by STAIR**

Although we aligned spatial embeddings and 2D spatial coordinates, computational construction of *de novo* 3D atlas with only ST data posed a significant challenge. One of the key difficulties was computationally determining inter-slice position because the slices used for constructing atlas might come from varied samples, lacking standardized inter-slice distances. Hence, we aimed to build a

196 3D atlas without prior position knowledge or paired images.

197 First, we tested z-axis coordinate reconstruction on 12 mouse hypothalamic preoptic area slices
198 generated by MERFISH, where the z-axis location of each slice was known which could serve as a
199 ground truth[2]. We employed STAIR-Emb (Supplementary Fig. S4A, B) and the inter-slice attention
200 score derived from STAIR-Emb showed a strong negative correlation with physical distance
201 (Spearman's $\rho$ = -0.88, Fig. 4A, B), indicating its potential to reconstructing z-axis. By employing
202 MST on attention matrix, we achieved accurate z-axis reconstruction, evidenced by the Pearson
203 correlation coefficient $(PCC)$ of 1 and coefficient of determination $(R^2)$ of 1 with the ground
204 truth of z-axis coordinates (Fig. 4C). We further sequentially aligned the x-axis and y-axis of slices
205 followed by the reconstructed z-axis. This process led to precise *de novo* 3D coordinates (Fig. 4A,
206 right), showing high consistency with the known 3D coordinates (Fig. 4D). Moreover, we added an
207 additional 12 slices from another sample of a different gender, and the results showed high accuracy
208 for the prediction of the total 24 slices by STAIR, with $PCC$ and $R^2$ of 0.99 and 0.94, respectively
209 (Supplementary Fig. S4C to E).

210 Next, we took a more challenging task using 40 mouse brain slices across three different
211 samples generated from the ST platform[8]. This dataset covered 40 coronal hemi-brain sections from
212 the olfactory bulb to the hindbrain. The original study provided anteroposterior (AP)-axis
213 coordinates based on paired images and Bregma coordinates of the brain[24] (Fig. 4E, left). We started
214 by using STAIR-Emb to achieve efficient spatial feature integration and spatial region division (Fig.
215 4F). Despite complexity, a robust correlation between attention scores and physical positions
216 persisted, with a $\rho$ of -0.83 (Fig. 4E middle, Supplementary Fig. S5A, left). Further, $PCC$ and $R^2$
217 between reconstructed AP-axis and the ground truth were as high as 0.98 and 0.96, respectively
218 (Supplementary Fig. S5A, right). Subsequently, we employed STAIR-Loc to align dorsoventral
219 (DV)-axis and mediolateral (ML)-axis coordinates based on the slice order of reconstructed AP axis
220 coordinates, completing the 3D *de novo* construction (Fig. 4E, right; Supplementary Fig. S5B). The
221 accuracy of spatial positioning of each domain was evident by high expression of their
222 corresponding marker genes (Fig. 4G). For instance, there was high expression of marker genes *Dsp*
223 [25], *Gpr88* [26,27], *Ramp3* [28,29], and *Camk2n1* [30] in the hippocampus, striatum, thalamus, and superficial

cortex, respectively. To sum up, STAIR enabled accurate *de novo* 3D reconstruction based on ST slices only.

**3D modeling of breast tumor microenvironment**

To further illustrate biological insights provided by 3D atlas, we analyzed the HER2-positive breast cancer[31] ST data comprising three consecutive slices (H1 to H3) (Supplementary Fig. S6A). The pathologists[31] annotated one slice (H1) with six tissue types: invasive cancer, adipose tissue, connective tissue, breast glands, *in situ* cancer, and immune infiltrates (Fig. 5A).

First, STAIR integrated the spatial embeddings of all the three slices, yielding spatial domains highly consistent with the pathological annotations (Supplementary Fig. S6B). On slice H1, STAIR achieved an ARI of 0.36, surpassing that of STAligner, STitch3D, and PRECAST, which ranged from 0.30 to 0.32 (Supplementary Fig. S6C). Recognizing the inherent distinctions between transcriptome and pathological phenotypes, we deconvoluted[32] each spot with scRNA-seq[33] data to facilitate spatial domain annotation (Supplementary Fig. S7). STAIR's spatial regions were annotated as connective tissue, immune cancer, breast glands, adipose tissue, fibrous tissue near the tumor, invasive cancer, and two *in situ* cancer regions (Fig. 5B). In contrast, STAligner failed to differentiate the *in situ* cancer area from the invasive cancer area (cluster 0) (Supplementary Fig. S6B). STitch3D could not distinguish between the two *in situ* cancer areas (cluster 4). PRECAST struggled to separate the two spatially distinct *in situ* cancer areas (cluster 2, 3, and 5). Additionally, none of them could detect a heterogeneous region near tumor with fewer cancer cells and more immune cells (cluster 7).

Next, we reconstructed the 3D coordinates. The inferred z-axis distance, at 0.52/0.43, closely corresponded the true distance ratio. We then aligned the x-axis and y-axis positions in the order of the z-axis, revealing a continuous structure for each spatial region in the 3D space (Fig. 5C). While STAligner and STitch3D cannot infer z-axis, we compared STAIR with them in 2D coordinates alignment (Supplementary Fig. S6D) and employed LISI metric to assess the spatial clustered pattern of domains in the stacked 2D space, which was a comprehensive measure of spatial domain identification and 2D coordinates alignment (Supplementary Fig. S6E). STAIR showed best performance with the lowest median LISI value of 1.53.

Furthermore, we examined developmental trajectories[34] and 3D heterogeneity of tumor-associated domains. To select the initial domain of the developmental trajectory, we conducted a differential expression analysis of two *in situ* cancer regions (Supplementary Fig. S8A). *In situ* cancer-1, marked by high expression of the ERBB2 [35] gene and S100 family genes [36], showcased stronger malignancy and invasive potential. In contrast, *in situ* cancer-2, chosen as the starting region for trajectory inference, showed a relatively lower malignancy and a higher immune level, confirmed by overexpressed HLA family[37] genes and enriched immune cells. The trajectory unveiled a progression from *in situ* cancer-2 to *in situ* cancer-1, invasive cancer, and fibrous tissue near the tumor (Fig. 5D, Supplementary Fig. S8B), with genes exhibiting expression changes along the developmental trajectory (Fig. 5E). Notably, within the same area, tumor invasion exhibited inter-slice heterogeneity on the z-axis. For instance, *in situ* cancer-1 displayed a trend from H3 to H1, while *in situ* cancer-2 exhibited a trend from H1 and H3 to H2. Invasive cancer also followed a progression from H3 to H1 (Fig. 5F, Supplementary Fig. S8C).

To conclude, STAIR discerned tumor heterogeneity that cannot be identified by the other methods by integrating three ST slices. The 3D reconstruction of tumor tissue enabled analyzing heterogeneity of invasion paths in a 3D view.

**Assimilating new sections into a reference atlas with STAIR**

Finally, we tested STAIR's ability to integrate new slices into 3D atlas with molecular feature and physical coordinates, which enables continuous expansion and update of the existing 3D atlas.

We first assessed the ability for predicting the z-axis of new slices by STAIR. By sequentially masking the known z-axis coordinates of 12 slices in the MERFISH mouse hypothalamic preoptic dataset, we predicted the masked values based on attention scores. Remarkably, the $PCC$ between predicted and actual coordinates was 0.98, with an $R^2$ of 0.95. Expanding this analysis to 24 slices from two hypothalamic preoptic samples, the $PCC$ and $R^2$ still reached 0.98 and 0.94, respectively. Further testing on 40 ST mouse brain slices resulted in even higher $PCC$ and $R^2$ values of 0.99 and 0.98, respectively (Supplementary Fig. S9). Our analysis showed that STAIR enabled accurate z-axis prediction, even for slices from multiple samples.

Next, we evaluated its capability to integrate new slices into the existing 3D atlas. We

integrated a new slice from the Visium platform into an existing 3D ST-platform brain atlas [8], which was aligned to Allen Mouse Common Coordinate Framework (CCF) [38] 3D space through experimental information and image registration. We input 40 ST and the Visium slice into STAIR-Emb for inter-slice attention scoring (Fig. 6A) and seamlessly mixed spatial embeddings (Fig. 6B). Leveraging attention scores to predict the position of the new slice (Methods), we identified the nearest neighbor slice, 20A, in the 3D brain atlas. By taking the 2D coordinates of 20A as a reference template, STAIR-Loc performed scaling, rotation, and translation on the new Visium slice to map CCF 3D coordinates (Fig. 6C). Additionally, leveraging detailed spatial anatomical region annotations from the ABA, we secured regional information for each spot in the Visium slice (Fig. 6D). The UMAP-based visualization revealed internal clustering of anatomical regions and separation between regions (Fig. 6E). We further demonstrated the accuracy of regional annotation of the new Visium slice based on the expression of specific markers, such as *Cabp7*, *Hpca1*, *Gpr88*, *Rora*, *Mbp* and *Pmch*, which exhibited high expression in hippocampus, hippocampal formation, striatum, thalamus, fiber tracts and hypothalamus, respectively.

To summarize, STAIR enabled integrating new slices into a 3D atlas which might be generated from a different ST platform.

# Discussion

The rapid advances in ST have ushered in new opportunities for exploring tissue architecture with gene expression patterns. However, connecting perspectives across discrete 2D slices to enable unambiguous 3D biological comprehension presented persistent computational challenges. Here, we provide an integrated solution for robust alignment, construction to 3D modelling, and assimilation of emerging data.

STAIR achieves accurate spatial embedding and coordinate alignment, enabling pioneering *de novo* 3D atlas construction from ST slices only. Its superior efficacy and broad functionality are attributed to the attention mechanism of the heterogeneous graph attention network. First, it enables adaptive feature aggregation by capturing spot-level and slice-level relationships, thereby obtaining spatial embeddings informed by broader context. These embeddings further enable slice anchor

pairs to initialize 2D alignments. Second, the attention mechanism is highly interpretable and versatile. High-order semantic modeling captures the physical relationships between slices and provides an innovative z-axis positioning mechanism, which facilitates multi-sample 3D reconstruction and seamless assimilation of new slices.

While most of the previous ST aligners focus on 2D alignment or integration, STAIR and STitch3D take multiple parallel ST slices as input and aim to create integrated 3D spatial atlases. However, STAIR is technically distinguished from STitch3D in several aspects. (1) Input flexibility: STAIR requires only ST data as input, making it more readily applicable to datasets where matched single-cell data is not available. (2) *De novo* 3D reconstruction: STAIR can infer relative positioning of slices along the z-axis in a completely unsupervised manner. This enables fully data-driven 3D reconstruction without relying on known slice spacing or order information. (3) New slice assimilation: STAIR can predict the positioning of new slices based on attention score proximities and situate them within existing 3D atlases by transferring coordinate templates. This unique capability paves the way for continually expanding spatial references. (4) Coordinate alignment: While STitch3D aligns spots by 3D coordinate registration, STAIR offers an alternative approach using sequential 2D alignment between slice pairs guided by learned spatial embeddings. This provides greater flexibility when 3D location inputs are unreliable or unavailable.

Despite progress, further innovation must meet escalating biological complexity and spatial dimension. First, one current limitation is the effective integration of datasets from platforms with highly divergent resolutions, such as attempting to jointly analyze Stereo-seq data that profiles transcriptomes at subcellular resolution with ST data that assays much larger tissue spots. Second, integration of ST datasets with other data types such as spatial proteome will be a significant challenge. At application horizon, advancing analytical tools based on 3D atlas are needed. For example, most current spatially variable gene identification primarily models in 2D space, whereas capturing full organ system patterns in 3D could provide deeper biological insights.

In conclusion, STAIR provides a unified algorithm advancing analysis of multi-slice ST data, including alignment, integration, and *de novo* 3D reconstruction. This advancement facilitates the creation of spatial maps for diverse organs, enabling the exploration of molecular mechanisms

underlying tissue structure phenotypes in a true 3D dimension.

**STAIR algorithm**

**Data preprocessing.** Suppose that there are $S$ slices denoted as $A_1, A_2, \cdots, A_s$, each containing respective numbers of spots $N_1, N_2, \cdots, N_s$. Among these slices, $G$ genes are common. Consequently, the gene expression data can be represented as a matrix $X$ of dimensions $N \times G$, where $N = N_1 + N_2 + \cdots + N_s$ is the total number of spots across all slices. The spatial coordinate data is stored in a matrix $Y$ of dimensions $N \times 2$ or $N \times 3$ depending on if the z-dimension exists, and it is sufficient to input an $N \times 2$ dimensional coordinate matrix for STAIR. The raw gene expression counts in $X$ are normalized based on library size and then log-transformed to get the normalized expression matrix $\tilde{X}$.

To remove batch effects in expression features across various slices, we employed an autoencoder alongside batch annotation [16] $b_{ns}$. If the $n$th spot is from the $s$th slice, we set $b_{ns} = 1$; otherwise, we set $b_{ns} = 0$. Given the assumptions of negative binomial (NB) or zero-inflated negative binomial (ZINB) distributions in gene expression, the preprocessing framework is:

$$Z = f_1(\tilde{X}||B)$$

$$Z' = f_2(Z||B)$$

$$R = f_3(Z')$$

$$\Pi = f_4(Z')$$

Encoder $f_1$ contains two layers compressing the expression matrix into a 32-dimensional low-dimensional feature matrix $Z$. The decoder first uses $f_2$ to decode $Z$ and batch information into a 128-dimensional feature matrix $Z'$, the parameters of NB distribution $(R,\ L)$ and zero-inflated probability $\Pi$ are then learned based on $Z'$. The objective is to minimize reconstruction loss between $X$ and the NB/ZINB distribution parameterized by $(\Pi, \mathrm{R}, \mathrm{P})$. The loss function is the NB/ZINB negative log-likelihood.

**Spatial embedding alignment.** We use a heterogeneous graph attention network [15] to learn integrated spatial features across slices.

*Construction of heterogeneous graph.* We attribute node characteristics to each spot based on the slice it belongs to. Distinct approaches are employed to establish edges for spots located within the same slice and those positioned across different slices.

For spots in the same slice, we build homogeneous edges based on the physical location of the spot. For a spot with index $j$ in slice $A_J$, the set of its intra-slice neighbors $\mathcal{N}_j$ is identified by K-Nearest Neighbor (KNN) based on spatial coordinates.

For spots not within the same slice, since their relative physical coordinates are unknown, we construct heterogeneous edges based on the expression similarity obtained in the preprocessing step. Only nodes with highly consistent expression features are connected. For spots $j$ from slice $A_J$ and $k$ from slice $A_K$, the edge exists if $z_k \cdot z_j > \mathrm{t}$. We usually set $\mathrm{t} = 0.9$ and denote the inter-slice neighbors of spot $j$ from slice $A_K$ as $\mathcal{N}_j^{JK}$.

*Intra-slice information aggregation.* To learn the spatial features of each spot, we first aggregate the information from intra-slice neighbors based on homogeneous edges separately for each slice.

We adopted the same model architecture and aggregation method as the graph attention network (GAT) module in SECE[39], and obtained the homogeneous spatial feature $U^{hom}$ based on homogeneous edges,

$$U^{hom} = \mathrm{GAT}(Z, Y)$$

where $Z$ is the expression feature of the spot obtained by the preprocessing module.

*Inter-slice information aggregation.* Further, we employ spot and slice level attention to aggregate inter-slice information adaptively.

First, we learn spot-level attention. For two spots $i$ and $j$ from slices $A_I$ and $A_J$ connected by heterogeneous edge, the spot-level attention $e_{ij}^{IJ}$ between them represents the importance of spot $j$ to $i$, namely,

$$e_{ij}^{IJ} = LeakyReLU(q_{IJ}^T[W_{IJ}^1 u_i^{hom} || W_{IJ}^2 u_j^{hom}])$$

where $W_{IJ}^1$ and $W_{IJ}^2$ are weight matrices specific to slice pairs, $q_{IJ}$ is a learnable vector, and $e_{ij}^{IJ}$

is an asymmetric attention matrix that contains pairwise attention between slices $A_I$ and $A_J$. The final inter-slice spot level attention matrix $\alpha_{ij}^{IJ}$ is obtained by normalizing $e_{ij}^{IJ}$:

$$\alpha_{ij}^{IJ} = softmax_j(e_{ij}^{IJ}) = \frac{\exp(e_{ij}^{IJ})}{\sum_{k \in \mathcal{N}_i^{IJ}} \exp(e_{ik}^{IJ})}.$$

For a node $i$ from slice $A_I$, its spot-level heterogeneity based on neighbors in slice $A_J$ is:

$$u_i^{IJ} = \sigma\left(\sum_{j \in \mathcal{N}_i^{IJ}} \alpha_{ij}^{IJ} \cdot u_j^{hom}\right).$$

Then, we learn slice-level attention. Different slices contribute differently to target slice and we adaptively learn how much each slice contributes to others. For the slice $A_I$ to be learned, the slice-level importance of slice $A_J$ to $A_I$ is:

$$w_{IJ} = \frac{1}{N_I} \sum_{i \in A_I} q^T \cdot \tanh\left(W \cdot u_i^{IJ} + b\right),$$

where $W$ and $b$ are weight matrix and bias vector, $q$ is a learnable vector. Then normalize $w_{IJ}$ to get the final slice level attention coefficient:

$$\beta_{IJ} = \frac{\exp(w_{IJ})}{\sum_{K \neq I} \exp(w_{IK})}.$$

Therefore, the heterogeneity representation of slice $A_I$ based on other slices is:

$$U_I^{het} = \sum_{J \neq I} \beta_{IJ} \, U^{IJ}.$$

*Model learning and training.* For each slice, the final spatial feature matrix $U$ is obtained by combining the intra-slice representation $U^{hom}$ and inter-slice heterogeneous representation $U^{het}$:

$$U = \lambda \cdot U^{hom} + (1 - \lambda) \cdot U^{het}$$

where $\lambda$ weights the homogeneous and heterogeneous components with default value 0.8. The model is trained to ensure $U$ reliably represents the original spatial gene expression features $Z$. The objective function is defined as the mean squared error between $U$ and $Z$:

$$Loss = MSE(U, Z)$$

In the training process, $U$ is set to 32 dimensions. The Adam optimizer is used for training with learning rate of 0.001. The default number of iterations is 150.

**Slice-level attention-based reconstruction and prediction in z-axis.** The attention weight at the

slice level, denoted as $\beta_{IJ}$, characterizes the collective impact of slice $A_J$ on $A_I$ within the spatial embedding learning process. This encapsulates intricate higher-order spatial semantic insights. Notably, the strong correlation between these weights and the actual distance between slices signifies the potential for deducing the physical placement of slices.

*De novo z-axis reconstruction for multiple slices.* When we lack information about the third-dimensional coordinates $(y_1^{(3)}, \cdots, y_I^{(3)}, \cdots y_S^{(3)})$ of the slice, we can infer the positional relationships between these parallel slices based on the inter-slice attention scores. Specifically, we start by defining the distances between slices $A_J$ and $A_I$ as $d_{IJ} = 1 - \frac{\beta_{IJ}+\beta_{JI}}{2}$ and proceed to construct a Kruskal algorithm [19] based MST using *network* package implemented in Python. Subsequently, we select the given root slice $A_I$ and set its position as $l_I = 0$. The positions of the remaining nodes are then determined by adding the distance between each node and its parent node's position. Lastly, we normalize the distances to obtain position predictions $(l_1, l_2 \cdots, l_S)$ within the range of 0 to 1. Furthermore, for comparison with actual coordinates in the article, we additionally scale the normalized distances $(l_1, l_2 \cdots, l_S)$ to match the real coordinate range.

*Z-axis prediction for new slice.* When the z-axis coordinates $(y_1^{(3)}, y_2^{(3)}, \cdots y_S^{(3)})$ are known, in order to align the new slice $A_{new}$ into the unified 3D space, we predict $y_{new}^{(3)}$ based on the attention vectors between $A_{new}$ and $A_1, A_2, \cdots, A_s$, denoted as $\beta_1 = (\beta_{new,1}, \beta_{new,2}, \cdots, \beta_{new,S})$ and $\beta_2 = (\beta_{1,new}, \beta_{2,new}, \cdots, \beta_{S,new})$.

We first sort the elements of $\beta_1$ and $\beta_2$ in descending order, and choose the top $N$ elements $\hat{\beta}^1 = (\hat{\beta}_{new,R_1^1}, \cdots, \hat{\beta}_{new,R_N^1})$ and $\hat{\beta}^2 = (\hat{\beta}_{new,R_1^2}, \cdots, \hat{\beta}_{new,R_N^2})$. Following this, we identify the intersection of these element indices $\{R_1, \cdots, R_{N'}\}$, where $N' \leq N$. Finally, we perform weighted average on the z-axis corresponding to$(R_1, \cdots, R_{N'})$ to get the predicted value of $y_{new}^{(3)}$:

$$\hat{y}_{new}^{(3)} = \frac{\sum_{K\in\{R_1,\cdots,R_{N'}\}} \frac{\beta_{new,K} + \beta_{K,new}}{2} \cdot y_K^{(3)}}{\sum_{K\in\{R_1,\cdots,R_{N'}\}} \frac{\beta_{new,K} + \beta_{K,new}}{2}} \qquad (1)$$

**Spatial location alignment.** We employ a two-stage approach for 2D alignment in x-axis and y-

axis. In the initial stage, spatial features are utilized to identify spot pairs with precise matches, and initial transformation matrices are determined based on these pairs. In fine alignment stage, we first identify spots that effectively capture both global and local information within slices. The coordinate positions of these spots are then leveraged to execute ICP[18] fine registration.

*Initial alignment.* Consider slice $A_I$ with $n_I$ spots and slice $A_J$ with $n_J$ spots, along with their spatial feature sets $\{u_1^I, u_2^I, \cdots, u_{n_I}^I\}$ and $\{u_1^J, u_2^J, \cdots, u_{n_J}^J\}$, respectively. First, we measure the similarity between spots through the cosine distance based on spatial embeddings. The mutual nearest neighbors (MNN) of $k = 1$ located in different slices form pairs for the initial alignment, resulting in curated pairs $\{(i_1, j_1), (i_2, j_2), \cdots, (i_n, j_n)\}$ along with their coordinates $\{(y_{i_1}^I, y_{j_1}^J), (y_{i_2}^I, y_{j_2}^J), \cdots, (y_{i_n}^I, y_{j_n}^J)\}$. subsequently, we seek the optimal rotation matrix $R$ and translation vector $t$, to align the physical positions of the corresponding point sets. Consequently, the loss function for the initial alignment is formulated as:

$$Loss_{init} = \sum_{k=1}^{n} \left| y_{i_k}^I - (R y_{j_k}^J + t) \right|^2$$

Singular value decomposition[40] solves for $R$ and $t$. All spots are aligned to their targets accordingly.

*Fine alignment.* The fine alignment stage involves two main steps: identifying informative anchor spots for capturing both regional details and global outlines specific to the slice, followed by the application of the classic ICP [18] algorithm. First, we select highly concentrated regions that are common to both slices, according to median of LISI value which implemented in R package *lisi*. Subsequently, we identify the concave hulls[41] corresponding to the chosen regions and the slices. These concave hulls serve to depict both the local attributes of the region and the broader shape characteristics of the entire slice. Finally, ICP [18] algorithm is employed on these sets of informative points to achieve precise fine registration, resulting in our desired alignment outcome.

**Aligning new brain slices into existing ABA atlases.** We employed a dataset consisting of 40 ST

coronal brain slices with CCF 3D coordinate information, derived from a previous study [8], to serve as a foundational reference of whole-brain framework. To obtain the CCF 3D coordinates and anatomical region annotations of each spot in new slice, we incorporated it into the reference dataset through the following steps.

First, we integrate the spatial embeddings of the new slice $A_{new}$ with the reference set $\{A_1, A_2, \cdots, A_{40}\}$, obtaining the attention score associated with the $A_{new}$ and each reference slice. By applying formula (1), we predict the AP coordinate of the new slice, denoted as $y_{AP}^{new}$.

Then, we adjusted the ML and DV dimensions of $A_{new}$ to align with the CCF scale. We select the reference slice $A_J$ closest to $A_{new}$ according to $y_{AP}^{new}$, followed by filtering MNN spot pairs based on cosine distances between spatial features. This filtering process generated a set of spot pairs $\{(i_1, j_1), (i_2, j_2), \cdots, (i_n, j_n)\}$, each possessing corresponding ML and DV coordinates $\left\{\left(y_{i_1}^{new}, y_{j_1}^{J}\right), \left(y_{i_2}^{new}, y_{j_2}^{J}\right), \cdots, \left(y_{i_n}^{new}, y_{j_n}^{J}\right)\right\}$. Consequently, the scaling factor was calculated as:

$$scale = Median_{k_1 \neq k_2}\left\{\frac{\left|y_{j_{k_1}}^{J} - y_{j_{k_2}}^{J}\right|}{\left|y_{i_{k_1}}^{new} - y_{i_{k_2}}^{new}\right|}\right\}.$$

Next, the ML and DV orientations of slice $A_{new}$ are aligned into the CCF. We perform a two-stage alignment of the scaled coordinates (see **Spatial location alignment**), thus effectively integrating the 3D coordinates of $A_{new}$ into the CCF.

Finally, anatomical regions are assigned to each new spot based on location-specific information within the ABA annotation file *annotation_25.nrrd* from https://portal.brain-map.org/, establishing a mapping of the anatomical context for slice $A_{new}$.

## Toolkit for ST analysis

**Clustering for spatial domains.** Clustering is conducted on integrated spatial embeddings to obtain unified spatial domains across ST slices. We employ mClust [42] clustering method implemented in R package *rmclust*.

**Spatial trajectory inference.** Trajectory inference based on spatial embeddings is used to track the

development in spatial dimension. We utilize Monocle3 [34] to perform the pseudo time inference for each spot based on UMAP derived from spatial embedding of STAIR by applying *learn_graph* and *order_cells* in Monocle3 package. Function *graph_test* is employed to find genes that change with pseudo time.

**Deconvolution.** For HER2+ breast cancer dataset, we perform deconvolution to analyze the cell type composition using software Cell2location [32] and annotated scRNA-seq dataset [33]. Cell2location is run according to the tutorial and default parameters.

**Differential expression analysis.** We employ Seurat V4[43] to perform differential expression analysis. Differential expression analysis is used to identify cluster-specific marker genes where all the clusters are pairwise compared using the Wilcoxon method. Each identified marker gene was expressed in a minimum of 25% of cells and at a minimum log fold change threshold of 0.25.

## Evaluation

**Evaluation of spatial embedding alignment**. *Adjusted Rand index (ARI).* ARI measures the consistency between spatial domains identified by different algorithms and the known anatomical region labels. Given the contingency table of intersections between the algorithm-generated domains and annotation-based labels, it is calculated as

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}},$$

where $n_{ij}, a_i$ and $b_i$ are values from the contingency table. ARI values range from -1 to 1. Higher ARI indicates greater agreement with the ground truth annotations.

*Average Silhouette width (ASW).* ASW evaluates how well the features match true clusters in the data. For every sample $i$, Silhouette width $S(i)$ is calculated as

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where $a(i)$ is the average distance between i and points in its own cluster, and $b(i)$ is that to adjacent cluster points. ASW values range from -1 to 1, with greater ASW indicates better match.

In this paper, we employ $ASW_{domain}$ and $ASW_{batch}$ to measure the fitness of spatial features to known spatial regions and batches of slices, respectively. A larger $ASW_{domain}$ and a smaller $ASW_{batch}$ value represent stronger feature learning and data integration capabilities. To evaluate the comprehensive performance of spatial features, we calculated the harmonic mean $ASW_{F1}$ of $ASW_{domain}$ and $1 - ASW_{batch}$:

$$ASW_{F1} = \frac{2(1 - ASW_{batch})ASW_{domain}}{1 - ASW_{batch} + ASW_{domain}}$$

Higher $ASW_{F1}$ indicates stronger comprehensive ability of feature integration and biological specificity retention.

**Evaluation of spatial alignment in x-axis and y-axis.** To assess STAIR's capability for spatial alignment, we conducted simulations involving multiple slices, each with known 3D coordinates. Specifically, we maintained the spatial location of the first slice unchanged, while applying random rotations and translations to the remaining slices. Rotation angle $\theta \in (0, \pi)$, translation distance $t_{(1)} \in (-a_{(1)}, a_{(1)})$, $t_{(2)} \in (-a_{(2)}, a_{(2)})$, where $a_{(1)} = \max(y_{(1)}) - \min(y_{(1)})$, $a_{(2)} = \max(y_{(2)}) - \min(y_{(2)})$. Note that the ground truth rotation angle and translation distance of $S - 1$ slices are $\{\theta^2, \theta^3, \cdots, \theta^S\}$ and $\{(t^2_{(1)}, t^2_{(2)}), (t^3_{(1)}, t^3_{(2)}), \cdots, (t^S_{(1)}, t^S_{(2)})\}$, respectively. The error of rotation and translation are:

$$\Delta_\theta = \frac{1}{S-1}\sum_{I=2}^{S}\left|\theta^I - \hat{\theta}^I\right|$$

$$\Delta_t = \frac{1}{S-1}\sum_{I=2}^{S}\sqrt{\left(t^I_{(1)} - \hat{t}^I_{(1)}\right)^2 + \left(t^I_{(2)} - \hat{t}^I_{(2)}\right)^2}$$

where $\hat{\theta}^I$ and $(\hat{t}^I_{(1)}, \hat{t}^I_{(2)})$ are rotation angle and translation distance of slice $A_I$ obtained by the algorithm to be evaluated, respectively.

**Evaluation of spatial domain identification and 2D coordinate alignment in HER2+ breast cancer.** Local inverse Simpson's index (LISI) measures the degree of local mixing, and we use it to

evaluate the spatial aggregation pattern of domains in stacked 2D space. For each spot $i$, LISI is formulated as:

$$LISI(i) = \frac{1}{\sum_{l \in L} p_i(l)},$$

where $p_i(l)$ is the probability that the spatial domain label l exists in the local neighborhood of sample i, and L is the set of spatial domains. Local neighborhoods are selected by stacked 2D coordinates. The value of LISI is in the range of $[1, \infty)$, and smaller LISI indicates better aggregation pattern.

**Evaluation of z-axis reconstruction and prediction.** We utilize the coefficient of determination ($R^2$) and the Pearson correlation coefficient ($PCC$) to quantify the effectiveness of z-axis reconstruction and prediction. Denoting the ground truth z-axis coordinate of the $S$ slices as $\{y_{(3)}^1, y_{(3)}^2, \cdots, y_{(3)}^S\}$, and the reconstructed or predicted outcomes are $\{\hat{y}_{(3)}^1, \hat{y}_{(3)}^2, \cdots, \hat{y}_{(3)}^S\}$. The corresponding $R^2$ and $PCC$ values are calculated as:

$$R^2 = 1 - \frac{\sum_{I=1}^{S} (y_{(3)}^I - \hat{y}_{(3)}^I)^2}{\sum_{I=1}^{S} (y_{(3)}^I - \bar{y}_{(3)})^2}$$

$$PCC = \frac{\sum_{I=1}^{S} (y_{(3)}^I - \bar{y}_{(3)})(\hat{y}_{(3)}^I - \bar{\hat{y}}_{(3)})}{\sqrt{\sum_{I=1}^{S} (y_{(3)}^I - \bar{y}_{(3)})^2} \sqrt{\sum_{I=1}^{S} (\hat{y}_{(3)}^I - \bar{\hat{y}}_{(3)})^2}}$$

where $\bar{y}_{(3)} = \sum_{I=1}^{S} y_{(3)}^I$ and $\bar{\hat{y}}_{(3)} = \sum_{I=1}^{S} \hat{y}_{(3)}^I$.

**Assessment of alternative methods.** We conducted a comparative evaluation of STAIR against other alignment methods, including STAligner, Stitch3D, PRECAST, GraphST, and PASTE. In our assessment, we employed the default parameters for all methods unless specific parameters were outlined in the original text or tutorial.

*STAligner.* STAligner integrates ST data across different conditions, technologies, and devel opmental stages. It employs STAGATE and triplet loss to integrate the ST datasets until b atch-corrected embeddings are generated. It further considers shared spatial domain and M NNs identified by STAligner as corresponding pairs to guide the 2D alignment. We downl

oaded the package from https://github.com/zhoux85/STAligner, and ran STAligner following its tutorial https://staligner.readthedocs.io/en/latest/index.html.

*STitch3D.* STitch3D first unified 3D spatial coordinates for spots using ICP or PASTE, followed by graph construction based on 3D coordinates. It performed spatial embedding learning and integration by graph attention network and slice- and gene-specific parameters. We downloaded the package from https://github.com/YangLabHKUST/STitch3D, and ran STitch3D following https://stitch3d-tutorial.readthedocs.io/en/latest/tutorials/index.html. Given the requirement for scRNA-seq datasets from the same tissue, we utilized single cell DLPFC and HER2+ breast cancer data accessed at the Gene Expression Omnibus (GEO) under the accession code GSE144136 and GSE176078, respectively.

*PRECAST.* PRECAST is a probabilistic method for spatial embedding learning, clustering, and alignment. We downloaded the R package from https://github.com/feiyoung/PRECAST/, and ran PRECAST following its tutorial https://feiyoung.github.io/PRECAST/index.html

*PASTE.* PASTE provides the flexibility to align two slices either through pairwise alignment or to simultaneously align multiple slices using center alignment. We opted for pairwise alignment in our testing, as other approaches also employ pairwise processes. In pairwise slice alignment, it aims to find the best possible way to connect spots in one slice with spots in another slice, followed by constructing a stacked 3D alignment of a tissue. The connection, denoted as $\Pi$, is chosen to reduce both the differences in gene expression patterns between connected spots from different slices and the differences in physical distances between connected spots within the same slice. Parameter $\alpha$ was used to balance these two differences, and we set its default value of $\alpha = 0.1$ in our test. We download the package from https://github.com/raphael-group/paste/tree/main, and ran PASTE following https://github.com/raphael-group/paste/tree/main/docs/source/notebooks.

# Data availability

Visium DLPFC dataset can be accessed from the spatialLIBD package http:://spatial.libd.org/spatialLIBD. Olfactory bulb dataset sequenced by Stereo-seq is available at https://github.com/JinmiaoChenLab/SEDR_analyses. Olfactory bulb dataset sequenced by Slide-seqV2 is the Puck_200127_15 data in https://singlecell.broadinstitute.org/. MERFISH hypothalamic preoptic region data is downloaded from https://github.com/ZhuangLab/MERFISH_analysis. Dataset of ST brain slices is available at http://molecularatlas.org/, and the Visium brain data aligned to them is downloaded from https://www.10xgenomics.com/resources/datasets/mouse-brain-coronal-section-1-ffpe-2-standard. HER2+ breast cancer ST dataset is available at https://doi.org/10.5281/zenodo.4751624, and single cell data is accessed at GEO under the accession code GSE176078.

# Code availability

The STAIR algorithm is implemented and provided as a pip installable Python package which is available on GitHub https://github.com/yuyuanyuana/STAIR. All scripts used to reproduce all the analyses are also available at the same website.

# Declarations

**Ethics approval and consent to participate**

No ethnical approval was required for this study.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This project was supported by National Key R&D Program of China (2019YFA0904400, Z.X.), Guangzhou Science and Technology Project (202201020336, Z.X.).

**Authors' contributions**

Z.X. conceived and supervised the study. YY.Y. and Z.X. designed the study. YY.Y. analyzed the data. YY.Y. and Z.X. wrote the manuscript. All authors read and approved the manuscript.

**Acknowledgements**

We would like to thank all the data and software contributors who make this research possible. We also thank Zhongshan Ophthalmic Center and the Center for Precision Medicine at Sun Yat-sen University for the long-term support.

**Authors' information**

State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

Yuanyuan Yu & Zhi Xie

# Figures

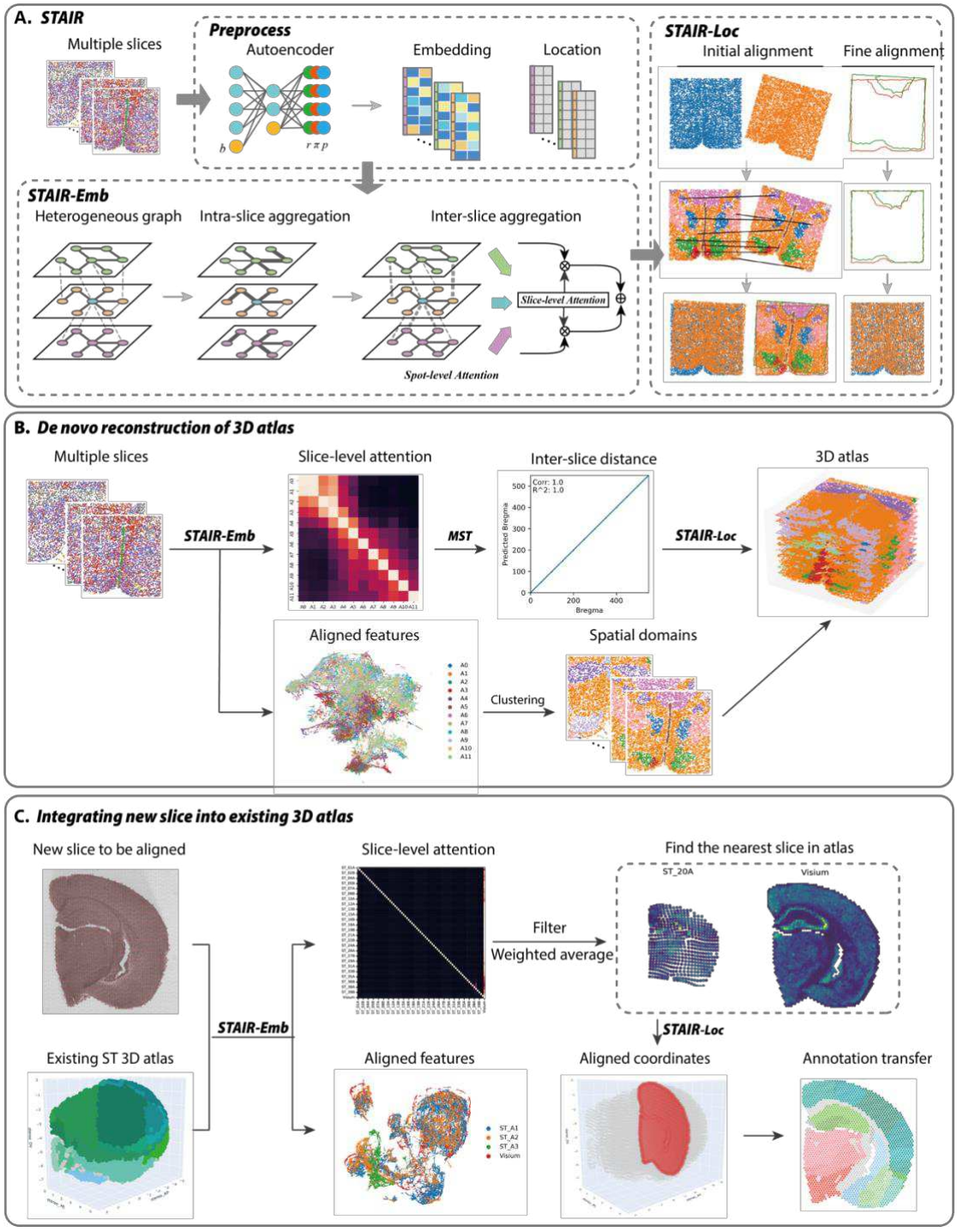


**Fig. 1. Overview of STAIR framework. (A)** STAIR processes multiple ST slices by utilizing an autoencoder to compress the expression matrices of each slice, resulting in integrated expression features. Subsequently, STAIR-Emb takes the expression features and the 2D coordinates of each slice as input, employing a heterogeneous graph attention network to learn integrated spatial features. Finally, STAIR-Loc utilizes these spatial features to establish the initial alignment of spatial

coordinates, followed by refining the alignment further by incorporating boundary points of slices and their respective domains. **(B)** *De novo* reconstruction of 3D atlas without prior knowledge about physical location. STAIR-Emb learns integrated spatial features and establishes high-order semantic relationships between slices, transforming them into distance matrices. Next, minimum spanning trees (MST) reconstruct relative positional relationships, guiding sequential spatial alignment. **(C)** Seamlessly integrating new slices into an existing 3D atlas. STAIR-Emb integrates spatial features of the new slice with the 3D atlas, followed by predicting the new slice's z-axis location and aligning its 2D coordinates (x- and y-axis) with the 3D atlas using STAIR-Loc.

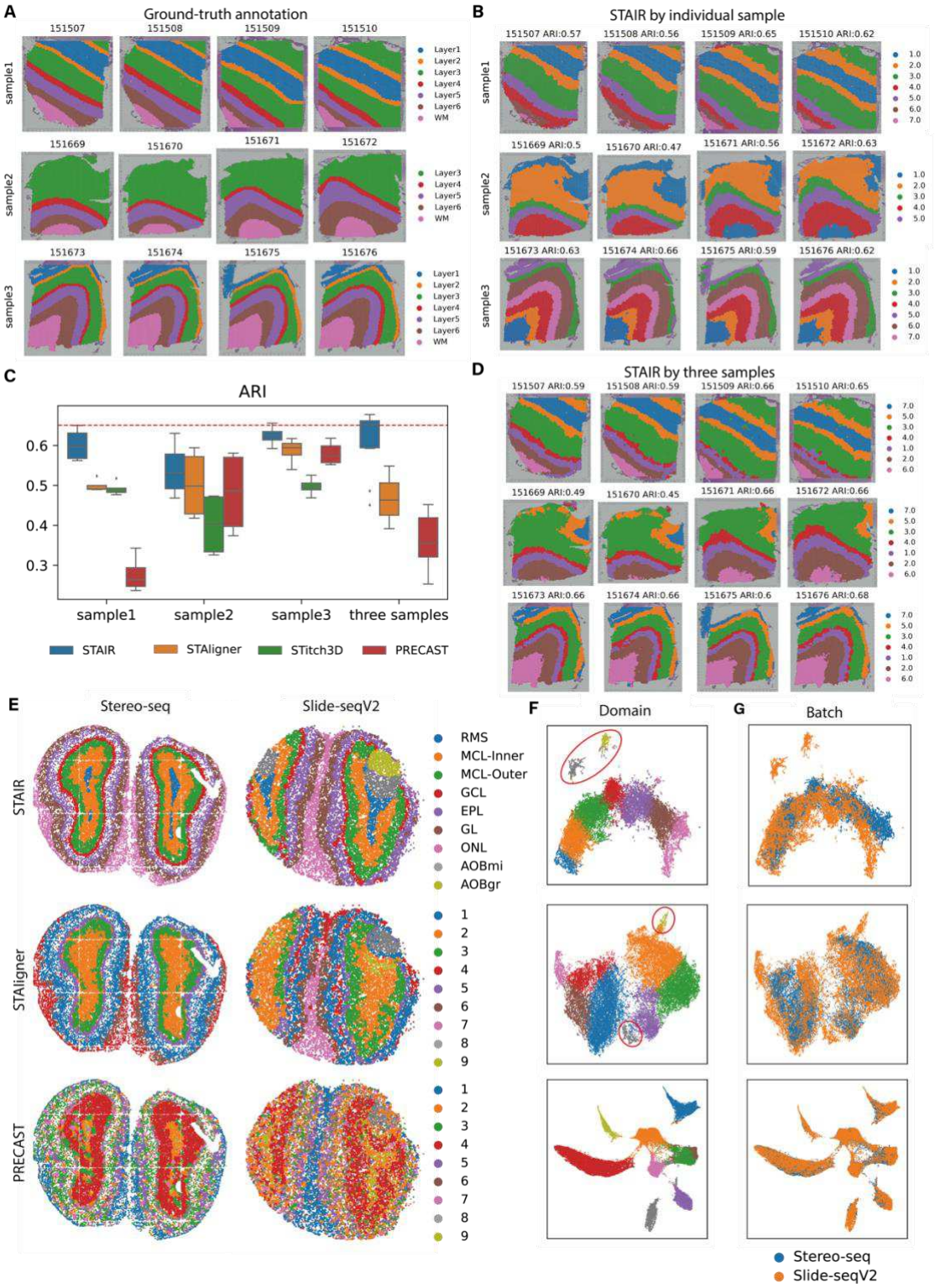


**Fig. 2. STAIR effectively integrates heterogeneous spatial transcriptomics data.** **(A)** Ground-truth segmentation of manually annotated regions in 12 DLPFC sections. **(B)** STAIR's spatial domain identification based on the 4 DLPFC slices for each sample. **(C)** Boxplots of adjusted rand index (ARI) scores of the four methods applied to the 4 DLPFC slices of each sample and to the

total 12 slices of the three samples. In the boxplot, the center line denotes the median, box limits denote the upper and lower quartiles, and whiskers denote the 1.5 × interquartile range. **(D)** STAIR's spatial domain identification based on the 12 DLPFC slices. **(E)** Integrative spatial regions identification of Stereo-seq (left) and Slide-seqV2 (right) mouse olfactory bulb data using STAIR, STAligner and PRECAST. **(F)** UMAPs derived from spatial embedding of STAIR, STAligner and PRECAST, with colors determined by the spatial regions they identify. **(G)** UMAPs derived from spatial embedding of STAIR, STAligner and PRECAST, with colors determined by the dataset to which the spot belongs.

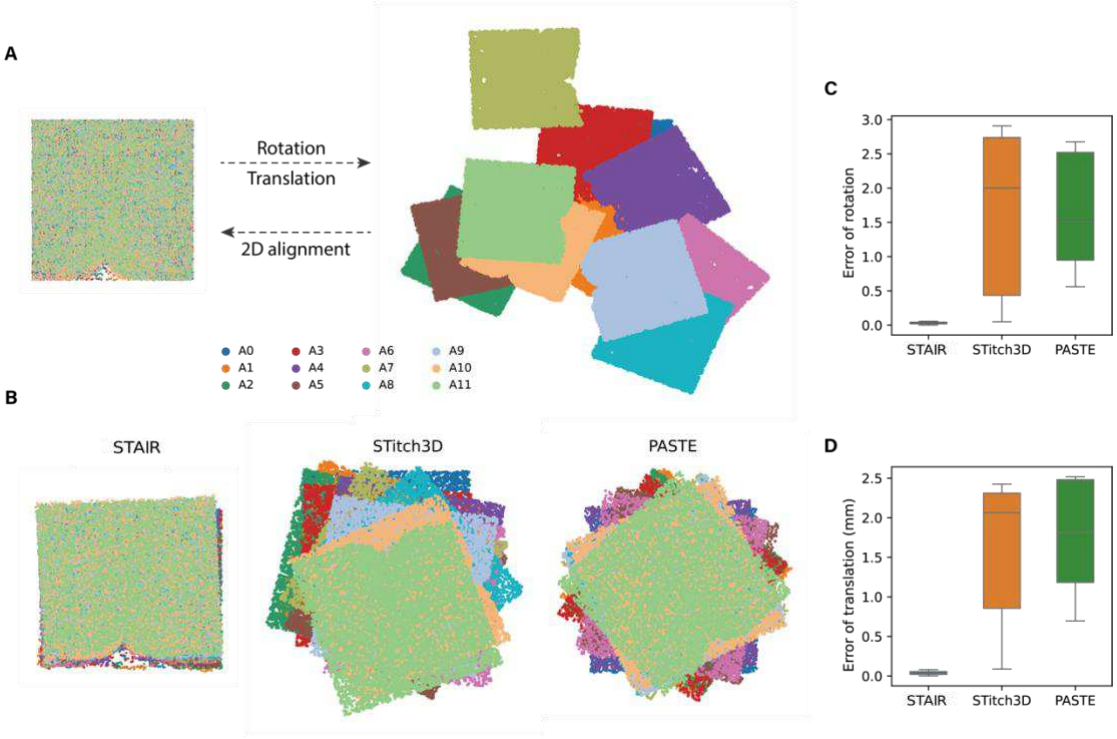


**Fig. 3. Precise alignment of 2D coordinates by STAIR.** **(A)** Schematic diagram of spatial position alignment. 2D coordinates in the first slice were fixed, and the remaining 11 slices were randomly rotated and translated. STAIR, PASTE, and STitch3D were employed to align the spatial coordinates of the rotated data. **(B)** Results of 2D spatial alignment using STAIR, STitch3D and PASTE. **(C)** Boxplots show the rotation errors of each method. In the boxplot, the center line denotes the median, box limits denote the upper and lower quartiles, and whiskers denote the 1.5 × interquartile range. **(D)** Boxplots show the translation errors of each method. In the boxplot, the center line denotes the median, box limits denote the upper and lower quartiles, and whiskers denote the 1.5 × interquartile range.

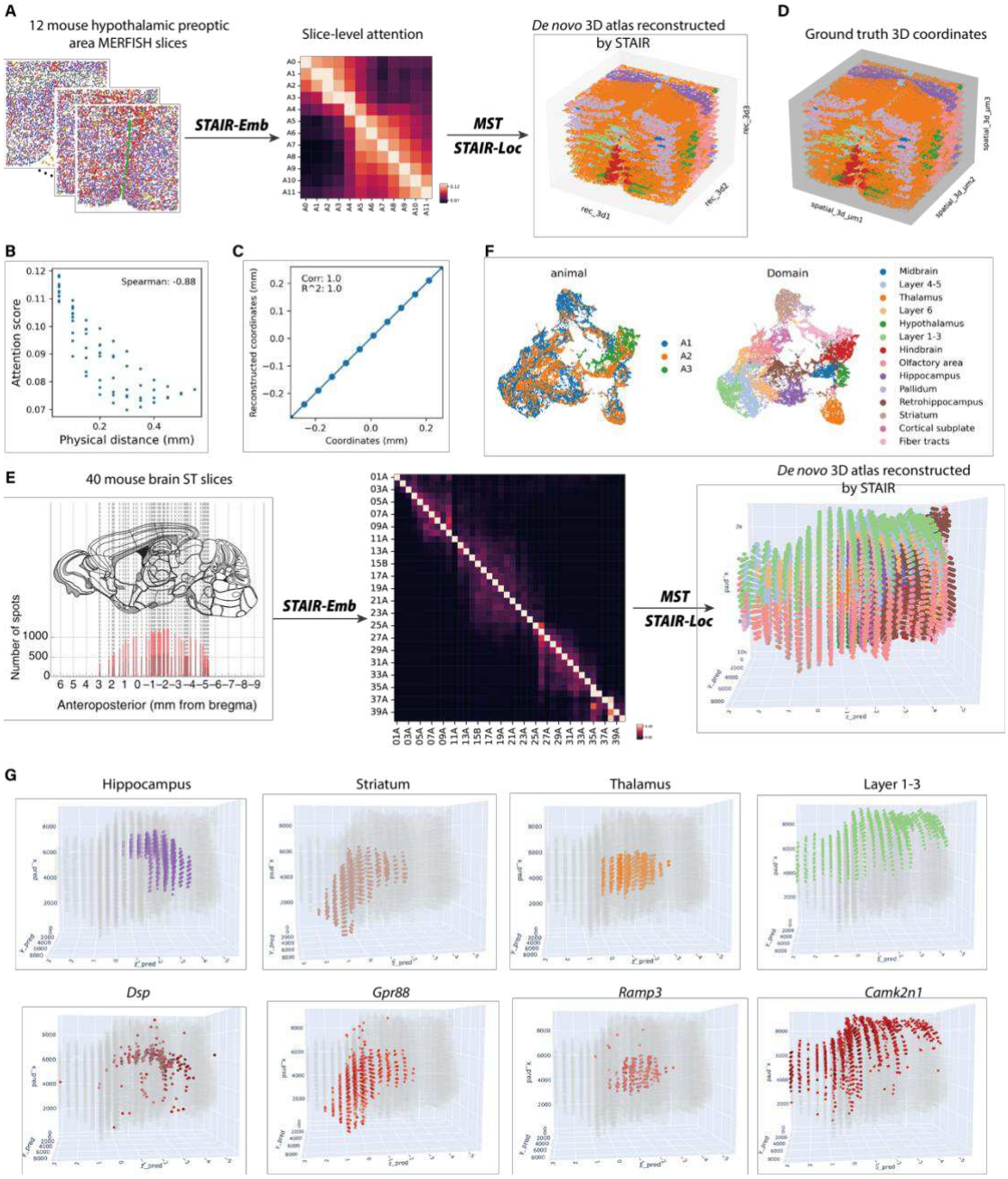


**Fig. 4. STAIR constructs *de novo* 3D atlas for the hypothalamic preoptic region and the mouse brain. (A)** Diagram depicting the *de novo* 3D atlas reconstruction based on 12 MERFISH slices in the hypothalamic preoptic region. We acquired slice-level attention scores by STAIR-Emb. Subsequently, these attention scores were used for reconstructing distances along the parallel direction of the slices. Finally, STAIR-Loc was employed to align the 2D coordinates guided by inter-slice distance. **(B)** Scatterplot showing the correlation between pairwise attention scores and physical distance between slices, with a Spearman correlation coefficient of -0.88. **(C)** Comparing reconstructed and actual physical coordinates in z-axis reveals a perfect correlation, with both the

Pearson correlation coefficient and determination coefficient of 1. The reconstructed coordinates are proportionally scaled to match the dimensions of the real physical coordinates. **(D)** Visualization of ground truth 3D coordinates, colored by spatial domains. **(E)** Left: Distribution of 40 coronal sections generated by the ST platform used to generate the atlas, adapted from the Ortiz's work [8]. Middle: Heatmap of attention scores across 40 slices in the of ST mouse brain data. Right: Visualization of *de novo* reconstructed 3D coordinates, colored by spatial domains. **(F)** UMAP visualization generated by spatial embeddings of STAIR, colored by original samples (left) and domains (right). **(G)** Visualization of *de novo* reconstructed 3D coordinates, colored by spatial domains (top) and their corresponding marker genes (bottom).

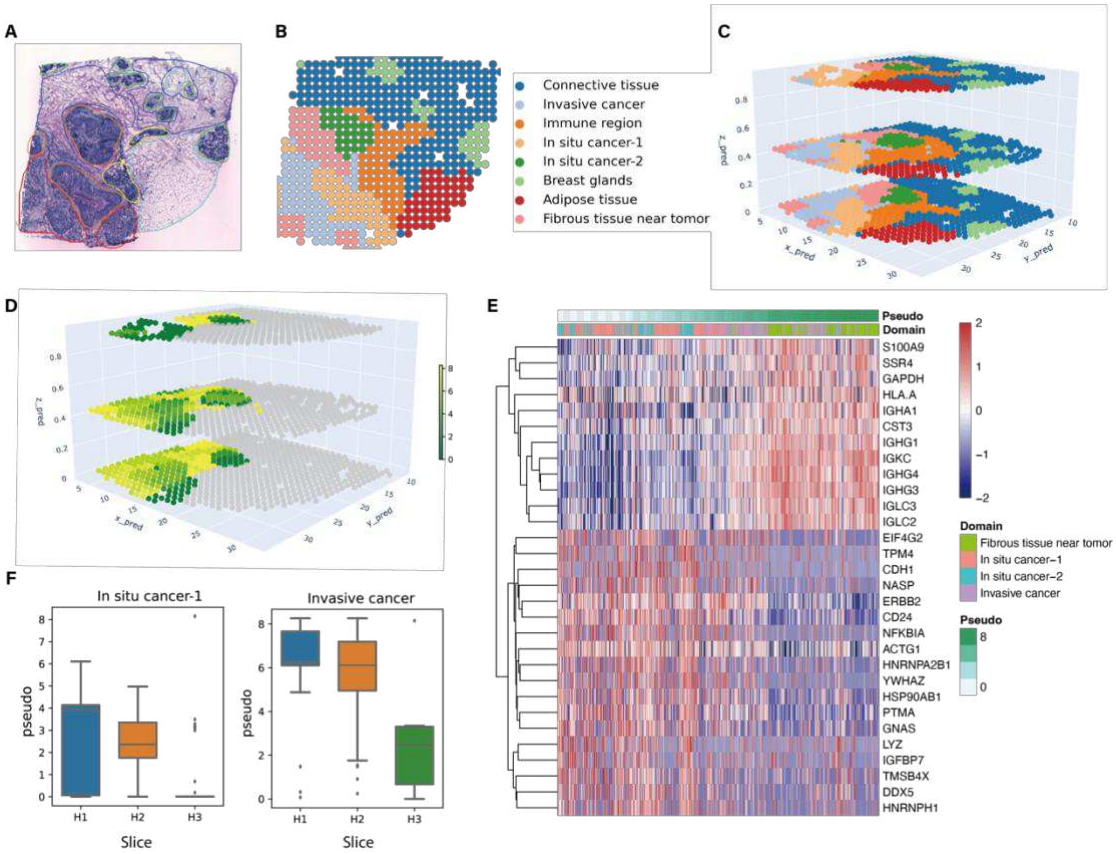


**Fig. 5. *De novo* 3D reconstruction and analysis of HER2+ breast cancer slices. (A)** Annotations of slice H1 in the original study[31] into six distinct categories: invasive cancer (red), adipose tissue (cyan), connective tissue (blue), breast glands (green), *in situ* cancer (orange) and immune infiltrates (yellow). **(B)** 2D spatial visualization shows the domains identified by STAIR in slice H1. That of the other two slices are displayed in Fig. S6B. **(C)** Visualization of *de novo* reconstructed 3D coordinates, colored by spatial domains. **(D)** Pseudo-time of each spot inferred by Monocle3 based on spatial embedding from STAIR. **(E)** Heatmap displaying genes with expression changes along the Monocle-derived pseudo-time, with spots ordered by pseudo-time. **(F)** Boxplot shows the pseudo-time of the spots in in situ cancer-1 (left) and invasive cancer (right) for each slice. In the boxplot, the center line denotes the median, box limits denote the upper and lower quartiles, and whiskers denote the 1.5 × interquartile range.

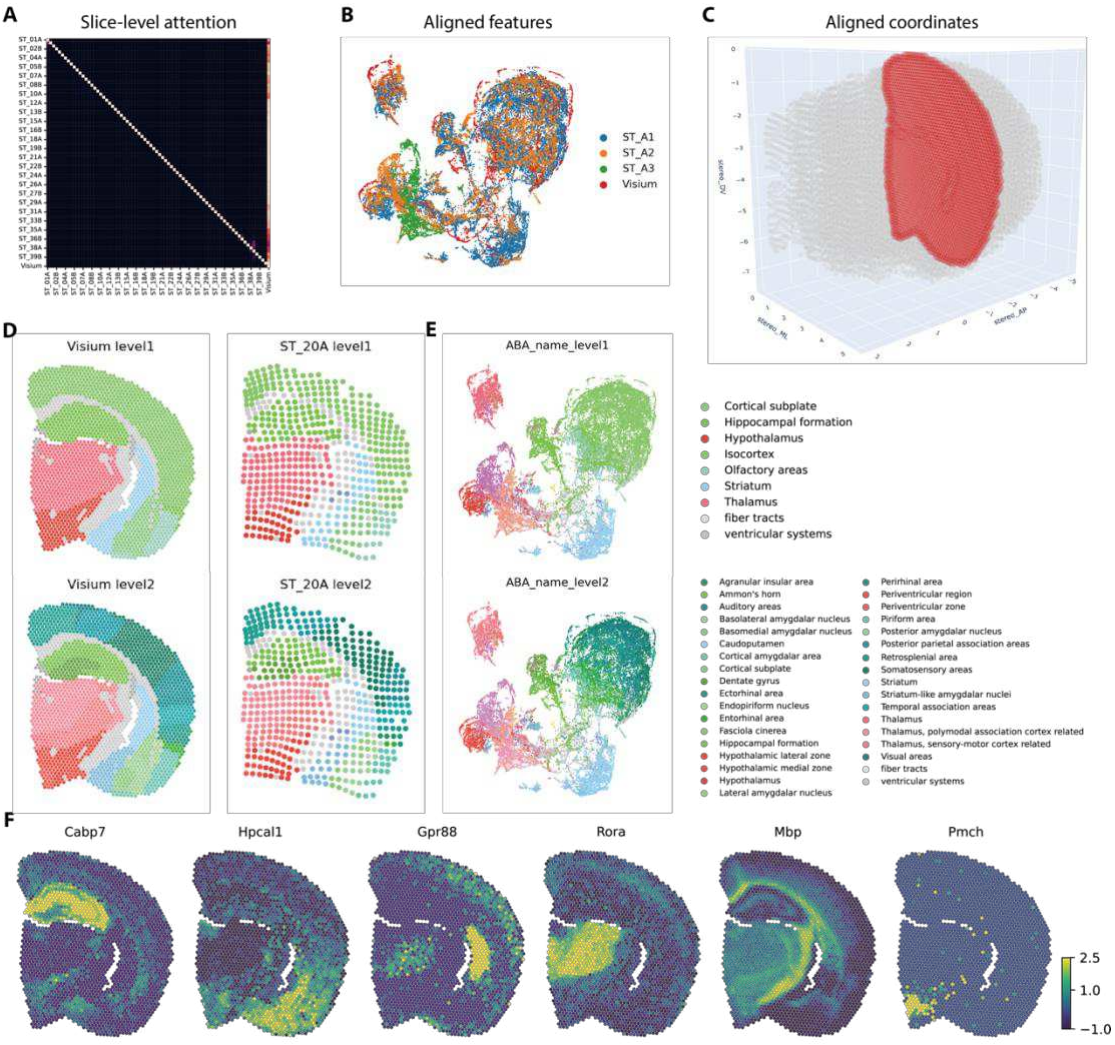


**Fig. 6. Assimilating new sections into a reference atlas. (A)** Heatmap depicting attention scores among 41 mouse brain slices, including 40 slices of the reference atlas from the ST platform, along with an additional slice generated from the Visium platform. **(B)** UMAP visualization of spatial embedding generated by STAIR, with colors indicating the respective sample of origin. **(C)** Visualization of the unified three-dimensional space after aligning the coordinates of the Visium slice with the 3D atlas. Spots from the Visium slice and ST slices are shown in red and gray respectively. **(D)** Spatial visualization of the Visium slice (left) and the ST slice (right) closest to Visium slice, with colors indicating the anatomical regions of the first (top) and the second (bottom) levels. **(E)** UMAP visualization of spatial embedding generated by STAIR, with colors indicating the anatomical regions of the first (top) and second (bottom) levels. **(F)** Spatial visualization of known regional marker genes.

# References


1 Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).

2 Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, doi:10.1126/science.aau5324 (2018).

3 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).

4 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology* **39**, 313-319, doi:10.1038/s41587-020-0739-1 (2021).

5 Eng, C. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239, doi:10.1038/s41586-019-1049-y (2019).

6 Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, doi:10.1126/science.aat5691 (2018).

7 Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792 e1721, doi:10.1016/j.cell.2022.04.003 (2022).

8 Ortiz, C. *et al.* Molecular atlas of the adult mouse brain. *Sci Adv* **6**, eabb3446, doi:10.1126/sciadv.abb3446 (2020).

9 Wang, M. *et al.* High-resolution 3D spatiotemporal transcriptomic maps of developing Drosophila embryos and larvae. *Dev Cell* **57**, 1271-1283 e1274, doi:10.1016/j.devcel.2022.04.006 (2022).

10 Chen, A. *et al.* Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. *Cell* **186**, 3726-3743 e3724, doi:10.1016/j.cell.2023.06.009 (2023).

11 Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nat Methods* **19**, 567-575, doi:10.1038/s41592-022-01459-6 (2022).

12 Liu, W. *et al.* Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with PRECAST. *Nat Commun* **14**, 296, doi:10.1038/s41467-023-35947-w (2023).

13 Zhou, X., Dong, K. & Zhang, S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science* **3**, 894-906, doi:10.1038/s43588-023-00528-w (2023).

14 Wang, G. *et al.* Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nature Machine Intelligence* **5**, 1200-1213, doi:10.1038/s42256-023-00734-1 (2023).

15 Wang, X. *et al.* in *The World Wide Web Conference* 2022–2032 (Association for Computing Machinery, San Francisco, CA, USA, 2019).

16 Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative

modeling for single-cell transcriptomics. *Nat Methods* **15**, 1053-1058, doi:10.1038/s41592-018-0229-2 (2018).

17 Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* **10**, 390, doi:10.1038/s41467-018-07931-2 (2019).

18 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, doi:10.1109/34.121791 (1992).

19 Kruskal, J. B. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society* **7**, 48-50, doi:10.2307/2033241 (1956).

20 Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425-436, doi:10.1038/s41593-020-00787-0 (2021).

21 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 (2018). <https://ui.adsabs.harvard.edu/abs/2018arXiv180203426M>.

22 Fu, H. *et al.* Unsupervised Spatially Embedded Deep Representation of Spatial Transcriptomics. *bioRxiv*, 2021.2006.2015.448542, doi:10.1101/2021.06.15.448542 (2021).

23 Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* **39**, 313-319, doi:10.1038/s41587-020-0739-1 (2021).

24 Cecyn, M. N. & Abrahao, K. P. Where do you measure the Bregma for rodent stereotaxic surgery? *IBRO Neuroscience Reports* **15**, 143-148, doi:https://doi.org/10.1016/j.ibneur.2023.07.003 (2023).

25 Hagihara, H., Fujita, M., Umemori, J., Hashimoto, M. & Miyakawa, T. Immature-like molecular expression patterns in the hippocampus of a mouse model of dementia with Lewy body-linked mutant beta-synuclein. *Mol Brain* **11**, 38, doi:10.1186/s13041-018-0378-3 (2018).

26 Spark, D. L. *et al.* Gpr88 Deletion Impacts Motivational Control Without Overt Disruptions to Striatal Dopamine. *Biol Psychiatry Glob Open Sci* **3**, 1053-1061, doi:10.1016/j.bpsgos.2022.10.008 (2023).

27 Laboute, T. *et al.* The orphan receptor GPR88 blunts the signaling of opioid receptors and multiple striatal GPCRs. *Elife* **9**, doi:10.7554/eLife.50519 (2020).

28 Wende, B. *et al.* Expression of the Calcitonin Receptor-like Receptor (CALCRL) in Normal and Neoplastic Tissues. *Int J Mol Sci* **24**, doi:10.3390/ijms24043960 (2023).

29 Oliver, K. R. *et al.* Cloning, characterization and central nervous system distribution of receptor activity modifying proteins in the rat. *Eur J Neurosci* **14**, 618-628, doi:10.1046/j.0953-816x.2001.01688.x (2001).

30 Wei, J. R. *et al.* Identification of visual cortex cell types and species differences

using single-cell RNA sequencing. *Nat Commun* **13**, 6902, doi:10.1038/s41467-022-34590-1 (2022).

31 Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun* **12**, 6012, doi:10.1038/s41467-021-26271-2 (2021).

32 Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661-671, doi:10.1038/s41587-021-01139-4 (2022).

33 Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334-1347, doi:10.1038/s41588-021-00911-1 (2021).

34 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

35 Tan, M. & Yu, D. Molecular mechanisms of erbB2-mediated breast cancer chemoresistance. *Adv Exp Med Biol* **608**, 119-129, doi:10.1007/978-0-387-74039-3_9 (2007).

36 Chen, H., Xu, C., Jin, Q. & Liu, Z. S100 protein family in human cancer. *Am J Cancer Res* **4**, 89-115 (2014).

37 Kaneko, K. *et al.* Clinical implication of HLA class I expression in breast cancer. *BMC Cancer* **11**, 454, doi:10.1186/1471-2407-11-454 (2011).

38 Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953 e920, doi:10.1016/j.cell.2020.04.007 (2020).

39 Yu, Y., He, Y. & Xie, Z. SECE: accurate identification of spatial domain by incorporating global spatial proximity and local expression proximity. *bioRxiv*, 2023.2012.2026.573377, doi:10.1101/2023.12.26.573377 (2023).

40 Arun, K. S., Huang, T. S. & Blostein, S. D. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, 698-700, doi:10.1109/TPAMI.1987.4767965 (1987).

41 Edelsbrunner, H. & Mücke, E. P. Three-dimensional alpha shapes. *ACM Trans. Graph.* **13**, 43–72, doi:10.1145/174462.156635 (1994).

42 Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289-317 (2016).

43 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

## Supplementary Files

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

- STAIRsupp.pdf