# DeepST: A versatile graph contrastive learning framework for spatially informed clustering, integration, and deconvolution of spatial transcriptomics

Yahui Long[1], Kok Siong Ang[1], Mengwei Li[1], Kian Long Kelvin Chong[1], Raman Sethi[1], Chengwei Zhong[1], Hang Xu[1], Zhiwei Ong[2], Karishma Sachaphibulkij[3], Ao Chen[4], Zeng Li[2], Huazhu Fu[5], Min Wu[6], Hsiu Kim Lina Lim[3], Longqi Liu[4], Jinmiao Chen[1,3]*

[1] Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos Building, Level 3, Singapore 138648

[2] National Neuroscience Institute (NNI), 11 Jalan Tan Tock Seng, Singapore 308433

[3] Immunology Translational Research Program, Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore (NUS), 5 Science Drive 2, Blk MD4, Level 3, Singapore 117545

[4] BGI-ShenZhen, Shenzhen, China 518103

[5] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632

[6] Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632

*Corresponding author. Email: chen_jinmiao@immunol.a-star.edu.sg

**Keywords:** Spatial transcriptomics; Single-cell RNA-seq; Graph neural networks; Self-supervised contrastive learning; Spatial clustering; Data integration

## Abstract

Advances in spatial transcriptomics technologies has enabled gene expression profiling of tissues while retaining the spatial context. To effectively exploit the data, spatially informed analysis tools are required. Here, we present DeepST, a versatile graph self-supervised contrastive learning framework that incorporates spatial location information and gene expression profiles to accomplish three key tasks, spatial clustering, spatial transcriptomics (ST) data integration, and single-cell RNA-seq (scRNA-seq) data transfer onto ST. DeepST combines graph neural networks (GNNs) with self-supervised contrastive learning to learn spot representations in the ST data, and an auto-encoder to extract informative features in the scRNA-seq data. Spatial self-supervised contrastive learning enables the learned spatial spot representation to be more informative and discriminative by minimizing the embedding distance between spatially adjacent spots and vice versa. With DeepST, we found biologically consistent clusters with higher accuracy than competing methods. We next demonstrated DeepST's ability to jointly analyze multiple tissue slices in both vertical and horizontal integration while correcting for batch effects. Lastly, we used DeepST to deconvolute cell types present in ST with scRNA-seq data, showing better performance than cell2location. We also demonstrated DeepST's accurate cell type mapping to recover immune cell distribution in the different regions of breast tumor tissue. DeepST is a user-friendly and computationally efficient tool for capturing and dissecting the heterogeneity within ST data, enabling biologists to gain insights into the cellular states within tissues.

## Introduction

Within the tissue of multicellular organisms, cells are organized with groups of similar cells physically clustered together. Linking gene expression of cells with their spatial distribution is crucial for understanding the tissue's emergent properties and pathology [1]. Current spatial transcriptomics (ST) combine gene expression profiling and spatial information to yield greater insights into both healthy and diseased tissues [2–4]. Spatial information is also useful for inferring cell-cell communications, especially juxtacrine signalling [5]. Within the spatial transcriptomics analysis workflow, assigning capture spots to spatial domains with unsupervised clustering is an essential task. Among the existing clustering methods employed in spatial domain identification, K-means, Louvain's method [6], Seurat [7] utilize only gene expression data to cluster spots into different domains. The domains identified by these methods are often discontinuous as they do not employ spatial information to identify colocalized cells that are likely to belong to the same domain. Recently, several methods have been proposed to improve spatial domain identification by exploiting spatial information. For example, Giotto [8] uses a Hidden Markov Random Field (HMRF) model to detect spatial domains with coherent gene expression patterns by fully exploiting the spatial dependency between spots. SpaGCN [9] uses a graph convolutional network-based model to identify spatial domain by integrating gene expression, spatial location, and the histology image. stLearn [10] combines morphological features obtained from the histology image with gene expression of adjacent spots to cluster similar spots in the tissue. BayesSpace [11] adopts a Bayesian statistical method that employs the information from spatial neighborhoods to improve clustering analysis. More recently, STAGATE [12] uses a graph attention auto-encoder framework to identify spatial domains through integrating spatial information and gene expression profiles. Another method, Cell Clustering for Spatial Transcriptomics data (CCST), uses graph convolutional network for unsupervised cell clustering [13]. However, these methods employ unsupervised learning and they often show sub-optimal clustering performance as the boundaries of identified domains are often fragmented and poorly match

pathological annotations. As ST ground-truth segmentation is usually not available, supervised learning cannot be employed to improve performance. Alternatively, advances in self-supervised learning have made the approach superior to unsupervised learning for representation learning, which can improve downstream clustering [14].

Spatial transcriptomics technologies have restrictions on the area captured during data acquisition. To perform spatial transcriptomics on a tissue slice encompassing a whole organ of interest, the sample is dissected into multiple sections. These adjacent multiple sections will have to be inferred jointly to accurately identify tissue compartments within the whole organ. Multiple tissue sections can also be obtained via serial sectioning of the organ of interest, yielding serial tissue slices that capture 3D information in each spatial transcriptomics experiment. Therefore, there is a need for methods to integrate and learn the joint representation of adjacent tissue sections (horizontal integration) and serial tissue sections (vertical integration). Most current analysis methods are suitable for only single tissue slice and cannot jointly identify spatial domains from multiple tissue slices. Moreover, batch correction methods developed for scRNA-seq are commonly employed. These methods are not suitable as they only consider gene expression and ignore related spatial information. Although STAGATE can be used for analyzing multiple tissue slices, its performance is limited by the lack of batch effect removal capabilities. As such, spatially informed batch correction is needed for ST data.

Current technological limitations also prevent ST from achieving single cell resolution with gene coverage comparable to single cell RNA sequencing (scRNA-seq). The popular 10x Visium platform can capture scRNA-seq scale transcriptomes but uses 55 µm capture spots that are larger than typical cells (5-10 µm) [15,16]. More recent sequencing-based technologies including Slide-seq [17], DBiT-seq [18], Stereo-seq [19], PIXEL-seq [20], and Seq-Scope [21], offer sub-cellular spatial resolution, but suffer from high dropout events that give rise to a very sparse gene-spot matrix. Meanwhile, fluorescent in situ hybridization (FISH) based methods are capable of subcellular resolution but lack genome scale gene coverage with the latest seqFish having a 10,000 gene limit [22]. To analyze data from low resolution capture methods, computational methods such as RCTD [23], stereoscope [24], SPOTlight [25], cell2location [26], CARD [27], and spatialDWLS [28], have been developed. These methods perform cell type deconvolution of spots measured in spatial transcriptomes by leveraging on cell type specific gene expression from RNA-seq. However, all current deconvolution methods except CARD ignore spatial location information. Moreover, they only return a matrix of cell type composition, except RCTD and cell2location which also calculate the cell-type specific gene expressions for each location. Tangram [29] and other similar packages adopt a batch integration approach to correct for technical variations in the scRNA-seq and ST data to accomplish integration, but do not employ spatial information. Currently cell-level deconvolution methods for achieving single-cell resolution spatial transcriptomes are still lacking, and performance is dependent on the cell type annotation quality and resolution. For FISH-based methods, spatially informed integration to project scRNA-seq data onto ST data are useful in creating single-cell resolution spatial transcriptomic maps with genome-wide gene coverage. Therefore, spatially informed integration of scRNA-seq data with ST is needed to achieve more accurate cell type deconvolution of ST and spatial registration of scRNA-seq.

To address the aforementioned challenges, we developed DeepST, a novel graph self-supervised contrastive learning framework that makes full use of spatial information and gene expression profiles for spatially informed clustering, batch integration, and cell type deconvolution. To our knowledge, few existing methods use self-supervised contrastive learning on spatial transcriptomics. Using self-supervised contrastive learning improves

performance in learning relevant latent features and has the additional benefit of removing batch effects. DeepST first combines graph neural networks with self-supervised contrastive learning to learn latent representations of spots for spatial clustering by encoding gene expression and spatial information. Alongside, we train an auto-encoder to learn the informative cell features from scRNA-seq data in an unsupervised way. Thereafter, DeepST learns a mapping matrix to project the scRNA-seq data into the ST space based on their learned features via an augmentation-free contrastive learning mechanism where the similarities of spatially neighboring spots are maximized and those of spatially non-neighboring spots are minimized. We extensively tested DeepST on different 10x Visium, Stereo-seq, and Slide-seqV2 datasets of human and mice tissues, including human brain, human breast cancer, human lymph node, mouse breast cancer, mouse olfactory bulb, mouse brain, and mouse embryo. The tests demonstrated DeepST's superiority over five existing methods in identifying spatial domains. Joint analyses on the mouse breast cancer and mouse brain datasets showed that DeepST was able to accurately identify spatial domains from multiple tissue slices while removing batch effects effectively without the need to explicitly detect batch factors. We also tested DeepST's projection of scRNA-seq data onto ST to predict cell states (cell types and sample types) in spatial spots. The computed cell-to-spot mapping matrix gave a more accurate estimate of cell type compositions than the best performing deconvolution method. Moreover, DeepST can transfer scRNA-seq derived sample phenotypes onto ST. We demonstrated this capability by delineating tumor and normal adjacent regions in a tumor derived tissue slice.

## Results

### Overview of DeepST

DeepST consists of three major modules: 1) graph self-supervised contrastive learning for spatially informed clustering (Figure 1A), 2) vertical and horizontal batch integration of multiple tissue sections (Figure 1B), and 3) spatially informed contrastive learning for ST and scRNA-seq data integration (Figure 1C). In the spatial clustering module, we leverage the spatial information of the spatial transcriptomics dataset to construct a neighbor graph where spots spatially close to each other are connected. Next, a graph self-supervised contrastive learning framework is used to embed the gene expression profiles into a latent representation space, with data augmentation and contrastive learning employed to help capture spatial context. The learned representation $Z_S$ is then reversed back into the original space to obtain the reconstructed gene expression matrix $H_S$ . The learning process minimizes both self-reconstruction loss and contrastive loss. Self-reconstruction loss enforces $H_S$ to capture the informative features from the gene expression and spatial location. Contrastive loss encourages the gene expression matrix $H_S$ to capture the spatial context information and be more informative and discriminative. This relies on the neighbor graph and its associated corrupted version generated from data augmentation that are used to construct positive and negative pairs. During model training, the similarities of positive pairs are maximized while those of negative pairs are minimized.

Unlike most existing approaches that remove batch effects by explicitly detecting batch factors, such as harmony [30], mutual nearest neighbors (MNN) approach [31], and Seurat 3.0 [8], our DeepST framework is able to implicitly removes batch effects during model training. Before training the model in the DeepST framework, multiple sections are aligned or stitched together, and a shared graph is constructed. Thereafter, graph self-supervised contrastive learning is performed using the shared graph, which allows for DeepST model to automatically smooth features of adjacent spots from both intra- and inter-section. Therefore,

after model training, not only neighbor spots from the same section have similar representations, but also neighbor spots from different sections have similar representations.

In the third module for ST and scRNA-seq data integration, we use an auto-encoder to learn cell representation $H_c$ from scRNA-seq gene expression. With the cell representation $H_c$ and spot representation $H_s$ learned from the first module, we aim to project scRNA-seq data into spatial transcriptomes via a trainable mapping matrix $M$, denoting the probability that cells are projected into each spot of spatial data. This process is implemented by aligning the predicted spatial gene expression $H_s' = M^T \cdot H_c$ to the reconstructed spatial gene expression $H_s$ via an augmentation-free contrastive learning mechanism, where the similarities of positive pairs (i.e., spot $i$ and its spatial neighbors) are maximized while those of negative pairs (i.e., spot $i$ and its spatial non-neighbors) are minimized. During model training, the first and third modules are independently executed. Subsequently, after model training, $H_s$ and $M$ can be applied for various tasks, such as spatial segmentation, multiple ST data integration, and scRNA-seq and ST data integration.

**DeepST significantly improves spatial clustering over existing methods**

*Application to human dorsolateral prefrontal cortex 10x Visium data.* We first assessed DeepST's spatial clustering performance on the LIBD human dorsolateral prefrontal cortex (DLPFC) dataset [32]. This dataset contains spatially resolved transcriptomic profiles of 12 DLPFC slices, each depicting the four or six layers of the human dorsolateral prefrontal cortex and white matter (WM). Here we tested DeepST and competing methods for their ability to recover the annotated anatomical cortex layers in an unsupervised manner. Across all 12 slices, DeepST achieved the highest median score of 0.6 among all methods with only STAGATE's performance as a close second (Figure 2A). The remaining methods had median ARI scores of less than 0.5, and for some methods like SpaGCN and STAGATE, there was a wide variance in performance across different slices. DeepST also showed low performance variance across slices than other methods. Seurat showed similarly low performance variance but its median performance was also the worst.

Next, we illustrate the results with one slice (#151673) in Figure 2B-C. Visual comparison clearly showed that the Seurat clustering had the poorest performance. It was only able to recover the Layer 1 clustering with the remaining clusters mixed among the other layers and clustered the WM with a portion of Layer 6. The boundaries between clusters were also ragged with no clean separation. Giotto accomplished better separation between WM and Layer 6, but Layers 3-5 were not correctly recovered. SpaGCN, BayesSpace, and STAGATE produced layers that were closer in shape to the annotation but with different layer thicknesses. The methods also did not accurately capture the boundaries between Layer 6 and WM, nor capture Layer 4 as a thin layer. The former was accomplished by DeepST, the only method that could do so. The latter is particularly challenging as no method was able to do so. For quantitative assessment, we employed the widely used Adjusted Rand Index (ARI). DeepST achieved the highest score of 0.62 while Seurat was the poorest at 0.29 followed by Giotto at 0.34. The remaining methods (SpaGCN, BayesSpace, STAGATE) obtained similar score between 0.51 and 0.58. The results with all other slices are in the Supplementary Figure S1.

Among the methods tested in this example, only Seurat, STAGATE, and DeepST can be applied to the subsequent Stereo-seq and Slide-seq examples due to data format incompatibility. Consequently, we also tested all six methods on a 10x Visium mouse brain dataset [16] and showed that DeepST was the best method. The full analysis results are available in Supplementary file 1.

*Application to mouse olfactory bulb Stereo-seq data.* In this second example, we used a coronal mouse olfactory bulb tissue dataset acquired with Stereo-seq [19]. We first annotated the coronal mouse olfactory bulb's laminar structure using the DAPI-stained image, identifying the olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), granule cell layer (GCL), and rostral migratory stream (RMS) (Figure 2D). Overall, all three tested methods were able to separate the outer layers of the organ, namely ONL, GL, and EPL (Figure 2E). The results of Seurat are slightly poorer with greater mixing between clusters, which we attribute to its lack of consideration of spatial information. For the inner structure, only DeepST was able to demarcate the GCL and RMS regions. Seurat was able to find the RMS but merged the GCL with the outer IPL region. STAGATE was unable to clearly capture the laminar structure for effective identification of the relevant inner layers, particularly the RMS, GCL and IPL regions. We next used the respective marker genes of each anatomical region to validate DeepST's results (Figure 2F). We see good correspondence between DeepST's clusters and the known marker genes. For some marker genes such as Mbp and Pcp4, their high expression levels overlap with neighboring regions. This is expected as cell types are often shared among the different inner structures of organs and their differences are small. Nevertheless, DeepST was able to leverage the whole transcriptome and spatial information to discern the relevant anatomical regions.

*Application to mouse hippocampus Slide-seqV2 data.* In this example, we again compared Seurat, STAGATE, and DeepST with a mouse hippocampus dataset acquired with Slide-seqV2. For this comparison, we use the annotated Allen Brain Atlas as the ground truth (Figure 2G). Although Seurat was able to outline the major anatomical regions, there were many clusters that were intermixed (Figure 2H). We hypothesize that this is a result of Seurat capturing different cell types that reside across different anatomical regions; this highlights the importance of spatial information in accomplishing anatomically relevant clustering. STAGATE and DeepST produced more spatially consistent clustering and captured the major anatomical regions such as the dentate gyrus (DG) and the pyramidal layers within Ammon's horn, which can be further separated into fields CA1, CA2, and CA3. Here, DeepST was better than STAGATE in delineating the CA3 and DG regions with sharper boundaries and higher concordance with the anatomical annotation (Figure 2G) and marker gene expressions (Figure 2I). Unlike the other two methods, DeepST was able to differentiate between the third ventricle (V3), medial habenula (MH), and lateral habenula (LH). Comparatively, Seurat merged the LH with the rest of the thalamus while STAGATE merged the MH and LH into one region.

We further examined the regions' marker gene expressions to verify DeepST's clustering (Figure 2I). For most regions and their corresponding markers, they show good alignment. In particular, only DeepST was able to demarcate the MH and LH regions in high concordance to their respective markers, Nwd2 and Gpr151. For the third ventricle, the Enpp2 expression did not align well with DeepST's detected V3 region, but the latter better resembled the V3 region in the annotated brain reference. For comparison, Seurat and STAGATE's V3 regions were closer to the shape of Enpp2 expression region, but did not match the anatomical shape well.

*Application to mouse embryo Stereo-seq data*. In this final clustering comparison, we used Stereo-seq acquired datasets of mouse embryos at E9.5 and E14.5, with 5,913 bins and 25,568 genes, and 92,928 bins and 18,566 genes respectively [19]. Tissue domain annotations were obtained from the original Stereo-seq study wherein Leiden clustering from SCANPY [33] was performed on the union of neighborhood graphs constructed using spatial proximity and transcriptomic similarity, and the computed clusters were annotated based on

differentially expressed genes. We first examine the clustering results of the E9.5 embryo. Although the original annotation had 12 reference clusters (Figure 3A), we set the number of clusters in our testing to 22 to acquire a higher resolution of tissue segmentation. STAGATE output was notable with many identified clusters collectively forming a thick layer around the embryo (Figure 3B). The identified clusters also did not match the annotation well. In contrast, DeepST's clusters better matched the annotated regions (Figure 3B and Supplementary Figure S3A). More importantly, they showed high concordance with known marker genes of major organs. Particularly, the liver region marked by Afp, Fgb, Alb, Itih2, mesenchyme by Meox1, Meox2, Pcp4, dermomyotome by Myog, head mesenchyme by Crym, Six2, Alx4, Sclerotome by Pax1, Pax9, Meox1, Meox2, heart by Myl2, Myl7, Nppa, Myh6, Myh7, Tnni3, Ttn, and connective tissue by Postn were all well matched by clusters of DeepST (Figure 3C and Supplementary Figure S3B). DeepST also demarcated two clusters for the embryo heart's atrium and ventricle chambers, with compartment specific markers such as Myl7 and Nppa marking the atrium and ventricle chambers, respectively. Nppa is unique with a spatiotemporal expression pattern during heart development where it is first expressed in the developing mouse heart at E8.0–8.5 [34], specifically at the ventral portion of linear heart tube which develops into the ventricles. At E9.0–9.5, Nppa is expressed mostly in the left and right ventricles within the looped embryonic heart, but its expression becomes restricted to the atria upon birth. We also examined the transcription factors essential for heart development and cardiac septation, namely Gata4, Tbx5, and Gata6, and their regulon activities showed spatial distribution that matched DeepST's clusters (Figure 3C and Supplementary Figure S3B). Although STAGATE was also able to identify the two halves of the heart but with a lower concordance with cardiac gene expression, it was unable to identify the other major organs.

We next compared the clustering results with the E14.5 mouse embryo (Figure 3D). Here we set the number of clusters to 16, matching the original annotation. STAGATE produced overly smoothened clusters and failed to reveal any fine-grained tissue complexity (Figure 3E). In contrast, DeepST's clusters were able to capture much of the fine-grained structures in embryo and they were also highly concordant with the original annotation, accurately identifying major areas such as the heart, liver, and muscle regions (Figure 3F and Supplementary Figure S3C). Furthermore, DeepST was able to accurately outline the epidermis as confirmed by the expressions of Krt5, Krt14, and Krt15 (Figure 3G), while the original annotation assigned parts of the epidermis to the cavity instead. DeepST was also better at delineating the meninges region as a continuous closed loop. In contrast, the annotation's meninges cluster appeared more fragmented and discontinuous. Notably, the original annotation placed the choroid plexus as a separate region while DeepST clustered it as part of the meninges. As the choroid plexus resides in the meninges' innermost layer (the pia mater) of three, it is difficult to be detected as a separate region. Within the cranial region, DeepST demarcated several sub areas that were not in the original annotation. One key identified area is the cerebral cortex, which we verified with the expression of cortical glutamatergic neuroblast markers Neurod6, Tbr1, and Eomes (Figure 3G). Though DeepST did not identify the olfactory epithelium found in the original annotation, it was able to delineate a separate olfactory epithelium cluster when we increased the number of clusters to 20 (Supplementary Figure S3D). Finally, DeepST also delineated a cluster of osteoblasts marked by Nov and Mfap5. To identify it, we employed gene set enrichment analysis with Enrichr, which suggested that this cluster is associated with mouse osteoblast day 14 and day 21 (Supplementary Figure S3E). Overall, the mouse embryo clustering results affirm DeepST's strength over STAGATE in identifying anatomically distinct regions in complex tissue samples.

## DeepST corrects batch effects for vertical and horizontal integration of multiple tissue sections

*Vertical integration*. Serial tissue slices are experimentally probed to increase accuracy and enable reconstruction on the vertical axis. An obstacle to this data integration is the presence of batch effects. To date, there are few existing methods developed to integrate multiple tissue sections for joint analysis. Here we evaluated DeepST and STAGATE's ability to vertically integrate serial tissue sections using two mouse breast cancer serial sections acquired with 10x Visium. We first assessed the integration results with UMAP. In the uncorrected data, strong batch effects were clearly present (Figure 4A). DeepST was able to evenly mix both slices while STAGATE was only able to pull the slices closer (Figure 4B). In the subsequent clustering, DeepST's clusters in both slices were also highly overlapping, unlike the clusters from STAGATE. This shows that DeepST was able to correct the batch effects and accurately align the regions across different serial slices. STAGATE's clusters were also more fragmented than DeepST's and achieved a lower ARI score (0.27 vs 0.38).

We next tested both methods on another mouse breast cancer dataset with batch effect present (Figure 4C). DeepST again merged the tissue slices such that the same cell types from different slices are evenly mixed in the UMAP plot, while the batch difference are still visible in STAGATE's output (Figure 4D). DeepST again found clusters that aligned well across the two slices. STAGATE's clusters showed better overlap than in the previous example but still poorer than DeepST's with significant differences visible in clusters such as 2, 3, 4, and 8.

*Horizontal integration*. As tissue samples can be significantly larger than the capture slides used for transcriptomics, horizontal integration enables the data from multiple capture slides to be stitched together. Here we tested DeepST and SpaGCN's horizontal integration capabilities using the mouse brain 10x Visium data with two sections, both divided into anterior and posterior slices (Figure 4E). We first aligned the tissue slices of each sections and performed clustering. For comparison purposes, we set the number of clusters to 26 for both algorithms, the same number used in the demonstration of SpaGCN [9]. Using the mouse brain atlas annotation [35] (Figure 4F), we compared the results for both sections. We find that most of the SpaGCN and DeepST clusters showed good agreement with known anatomy. However, many of the SpaGCN clusters were fragmented and were not aligned along the shared edge of the two tissue sections. Unlike SpaGCN, DeepST's clusters again showed better alignment along the edge of the anterior and posterior sections. In addition, DeepST's clusters were also more accurate at capturing the structure of the cerebral cortex layers, cerebellum, and corpus callosum. This example again highlights DeepST's consistent capability in aligning shared domains across tissue sections and meanwhile removing batch effects.

## DeepST projects scRNA-seq to ST for cell-type deconvolution of ST and identification of spatial niches for cell subsets and associated gene regulation

*Accurate spatial deconvolution of simulation data.* With both scRNA-seq and ST data, DeepST can perform cell type deconvolution of ST data by mapping single-cell gene expression profiles onto spots. Upon mapping, DeepST produces a cell-to-spot mapping probability matrix that can be combined with cell type annotation to derive the spatial probability distribution of each cell type, which is equivalent to cell type deconvolution of spatial spots. We compared DeepST's deconvolution with cell2location which, to our knowledge, is the top performing method for ST deconvolution [36]. We first used the simulated datasets derived from seqFISH+ and STARmap acquired ST data used in the benchmarking study. The original high-resolution data was overlaid with a coarse grid and

cellular gene expression was summed within each grid square to simulate spots capturing multiple cells. We calculated four indices, Pearson correlation coefficient (PCC), structural similarity index measure (SSIM), root mean squared error (RMSE), and Jensen-Shannon divergence (JSD) to compare performance (Figure 5A). With the seqFISH+ derived simulated dataset, DeepST slightly outperformed cell2location in the first three indices PCC and SSIM (higher is better), and JSD (lower is better), while being slightly poorer in terms of RSME (lower is better). For STARmap derived dataset, DeepST also produced better PCC, SSIM, and JSD scores, while having similar average performance in terms of RSME.

*Resolving the cellular compartments of the human lymph node*. To demonstrate the ability of DeepST in integrating scRNA-seq with experimentally acquired ST data wherein each spot contains more than one cell, we first tested DeepST and competing methods with a publicly available 10x Visium dataset acquired from human lymph node tissue containing germinal centers (GC) [37]. The reference scRNA-seq data used for deconvolution was created from three studies of human secondary lymphoid organs, and is composed of 34 cell types and 73,260 cells [38–40]. The human lymph node is a dynamic environment with different cell types spatially intermixed, thus creating a challenge for cell type deconvolution. We first visually compared the results of DeepST and cell2location by examining the mapping of cell types onto GC and non-GC locations. The ground truth annotations of germinal centers used here were obtained from the original cell2location study. Comparatively, DeepST mapped more GC preplasmablast (prePB) and cycling, dark zone (DZ), light zone (LZ) B cells to the annotated germinal center locations (Figure 5B and C). Furthermore, the cycling and DZ B cells correctly colocalized while LZ B cells and prePB cells also spatially coincided as expected. The naïve B and preGC B cells were also correctly mapped to regions surrounding the GCs. We also quantified performance using ROC curves for the light zone, dark zone, and pre-plasmablast cells (Figure 5D). In all three cases, DeepST achieved the best performance with cell2location as a close second, while the remaining methods were significantly poorer. Finally, we quantified the GC cell type mapping by computing the odds ratio between the sum of deconvolution scores of GC spots and non-GC spots. For all computed cell types (LZ, DZ, and prePB), DeepST's results again showed higher scores than cell2location, indicating a higher concordance between ground truth annotation and DeepST deconvolution (Figure 5E). The mapping results of all 34 cell types are in Supplementary Figures S4 and S5 for DeepST and cell2location, respectively.

*Accurate spatial mapping of cells to human dorsolateral prefrontal cortex 10x Visium data.* We also tested DeepST and cell2location on the human dorsolateral prefrontal cortex (DLPFC) dataset. The DLPFC has a clear layered tissue structure (Figure 5F) which can help better distinguish performance differences between the methods. We first selected slice #151673 (3,639 spots and 33,538 genes) and paired it with the scRNA-seq data used in the CellDART [41] study, with 78,886 cells and 30,062 genes, annotated with 33 cell types. The cells spatially mapped by DeepST exhibited cortical layer-specific distribution patterns (Figure 5F). Here we see Ex_10_L2_4, Ex_7_L4_6, Ex_1_L5_6, Ex_8_L5_6, and Ex_4_L_6 form discernable ordered layers from the outermost to the innermost, demonstrating DeepST's capability to distinguish the progressive layering of the cortical regions that are representative of the actual physiological anatomy. DeepST also mapped the oligos cells onto the WM layer, which is known to be enriched with oligodendrocytes. In cell2location's output, the oligos cell mappings were similar to DeepST's, but without clear edges. However, the other cortex cell mappings were much more scattered and did not correspond to their expected positions in the layers. The results of all 33 cell types are in Supplementary Figures S6 and S7 for DeepST and cell2location, respectively.

We further tested DeepST and cell2location for cell type deconvolution with the mouse brain anterior dataset. DeepST similarly showed mappings with higher density and sharper edges while cell2location's mappings tend to be more diffuse. The full analysis results are presented in Supplementary File1.

*Accurate spatial mapping of cells to human breast cancer 10x Visium data revealed T cell suppression in IDC regions.* In this comparison, we first performed clustering on a human breast cancer 10x Visium dataset with 3,798 spots and 36,601 genes. The data was manually annotated with 20 regions and we also adjusted the algorithms to obtain 20 clusters (Supplementary Figure S11). For Seurat, Giotto, and STAGATE, many of the computed clusters are fragmented and discontinuous, while BayesSpace, SpaGCN, and DeepST produced less disjointed clusters. This result is also reflected in the ARI scores where Seurat, Giotto, and STAGATE had lower scores in the range of 0.41 to 0.47, while BayesSpace, SpaGCN, and DeepST had higher scores between 0.54 and 0.57. We next applied DeepST to deconvolute the cell types present in each spot. We mapped cells from the DISCO scRNA-Seq human breast tissue atlas [42] onto the ST dataset (Figure 6A). The fibroblasts, muscle, lymphatic endothelial, and vascular endothelial cells were expectedly mapped onto the Healthy and Tumor edge regions, while the myoepithelial cells were mostly mapped to the Tumor edge regions. The luminal and luminal progenitor cells were primarily mapped onto the IDC and DCIS/LCIS regions.

Focusing on the immune cells, we found a significant presence of diverse immune subsets, namely T cells and macrophage/DC/monocytes (myeloid cells) in the IDC regions (IDC 5, 6, 7, and 8). The correspondence of the single cells with their mapped regions are also visualized with UMAP (Figure 6B, C). Of the DCIS/LCIS regions, 4 and 5 only had macrophage/DC/monocytes present, while 1 and 2 had no immune cells mapped onto them. We only found small amounts of B and plasma cells to be mapped. B cells were mainly mapped to IDC 3 and a portion of Tumor edge 3, while small concentrations of plasma cells found in various Tumor edge regions. Macrophages being present in tumor tissues is of clinical concern as they are linked to tumor progression and hence poor patient survival [43]. The lack of B cells is also significant as it is also associated with poor patient prognosis [44].

As the T cells were found across different region types (DCIS/LCIS, Healthy, IDC, and Tumor edge), we further investigated their differences (Figure 6C). We first identified the differentially expressed genes (DEGs), followed by the associated differentially regulated pathways. We found the T cells in IDC regions to have the largest number of DEGs, including genes associated with T cell exhaustion (LAG3, CTLA4, TIGIT, HAVCR2, PDCD1), and CXCL13 (Figure 6D). In the pathway analysis, the most significant differentially regulated pathways were associated with leukocyte activation and response (Figure 6E). These results highlight T cell dysregulation within the tumor environment. Such T cells were also observed by Zhang et al. [45] and Bassez et al. [46] in other human breast cancer samples and both studies demonstrated their reactivation with anti-PD-L1 therapy to harness their cytotoxicity against the tumor cells.

When comparing the spatial mapping of cells from adjacent normal and solid tumor sites, adjacent normal cells were mainly mapped onto the Healthy (1, 2) and Tumor edge (1, 2, 4, 5) regions (Figure 6F, Supplementary Figure S11C). The DCIS/LCIS regions (1, 2) also show high levels of adjacent normal cell mapping. Conversely, the solid tumor derived cells mostly mapped to the IDC regions. Unlike other DCIS/LCIS regions, DCIS/LCIS 4 also had more solid tumor derived cells mapped to it. Other regions, notably Tumor edge 3 and DCIS/LCIS 5, show a mix of cell mappings from both adjacent normal and solid tumor cells.

These mappings suggest that DCIS/LCIS 4 and 5 were more advanced in tumor development than other DCIS/LCIS regions. For Tumor edge 3, its lower region was mapped with tumor cells while the upper region was primarily mapped with adjacent normal cells. Interestingly, DeepST merged the upper portion of Tumor edge 3 with DCIS/ICIS 1, which was also mapped with adjacent normal cells (Supplementary Figure S11B). Another notable feature of Tumor edge 3's lower region is that it had one of the two rare concentrations of B cells in this tissue sample. Further investigation is needed to reveal the reasons behind B cell infiltration into this region but not others, as B cell infiltration is considered positive for patient prognosis [44]. This cell type and sample type deconvolution step is valuable in exploring the heterogeneity present within tissue that may otherwise not be apparent from histological analysis. Moreover, such analysis can also help to determine the appropriate number of regions for clustering purposes.

## Discussion

Spatial transcriptomics enables the measurement of gene expression while retaining associated spatial context. By combining gene expression and spatial information, spatially informed clustering can be employed to identify biologically relevant domains. ST capture areas are currently limited in size, necessitating multiple adjacent tissue slides to capture larger tissue samples. Thus, horizontal data integration is needed for integrated studies. Extending to three-dimensional characterization of a sample, vertical data integration is then required for serial tissue slides. Current ST technologies are also limited to either low spatial resolution with each spot containing more than one cell or lower gene coverage at a higher subcellular resolution. Meanwhile, scRNA-seq offers single-cell resolution genome-wide gene coverage but lacks spatial information. As such, it is beneficial to integrate scRNA-seq with ST to enhance the spatial resolution and gene coverage of ST, as well as infer the spatial locations of scRNA-seq. Here we present DeepST, a versatile graph self-supervised contrastive learning framework that can effectively and efficiently perform spatial clustering, spatial transcriptomics data integration, and scRNA-seq data transfer onto ST data. DeepST employs graph neural networks to learn informative representations of the gene expression profiles with spatial locations. The learned representations are further refined with graph self-supervised contrastive learning with the spatial context information to be more informative and discriminative. For scRNA-seq data transfer, we use an auto-encoder to extract features from the scRNA-seq data while reducing noise.

Clustering on diverse ST datasets, DeepST was better than competing methods at identifying heterogeneous tissue structures. We also demonstrated DeepST's ability in vertical and horizontal data integration with multiple tissue samples. DeepST detected domains that are more continuous across samples when integrating ST data horizontally. When integrating ST data vertically, it identified spatial domains that aligned across samples while removing batch effect present. Finally, we validated DeepST's ability in mapping scRNA-seq onto ST acquired spots to quantify the spatial distribution of cell types. Overall DeepST is a user-friendly and computationally efficient tool.

The key features of DeepST responsible for its superior performance are the use of graph neural network and contrastive learning in capturing gene expression together with spatial information. In comparison, existing methods such as SpaGCN and STAGATE also use graph neural networks to learn gene expression and spatial information. These methods show poorer clustering performance and lack the ability to accomplish batch integration. The main differences between DeepST and these methods lie in the use of graph self-supervised contrastive learning to strengthen the latent representation learning and additional

reservation of spatial context information of spots. This makes the learned representation more informative and discriminative, thus improving clustering performance. Moreover, while there are similarities between DeepST and STAGATE in the integration of multiple samples, the use of contrastive learning enhances DeepST's ability in removing batch effects. In this scenario, STAGATE's limited capability in batch effect removal greatly limited its performance.

Finally, we showed DeepST can accurately assign scRNA-seq data onto ST spots. Most existing deconvolution methods, such as SPOTLight [25] and cell2location [26], require accompanied cell type annotation. Moreover, they do not consider spatial information for deconvolution. In contrast, DeepST employs augmentation-free contrastive learning and spatial information to learn the cell-to-spot mapping matrix. The learning process also does not require prior cell type information and thus the mapping matrix is able to flexibly project arbitrary cell attributes (e.g., cell type and sample type) onto spatial space. Thus, DeepST can directly project scRNA-seq data onto the spatial spots to achieve cell-level deconvolution.

While the process of mapping matrix learning is similar between DeepST and Tangram, there are also differences. DeepST employs augmentation-free contrastive learning to learn the mapping matrix, exploiting the spatial location information. DeepST also takes informative and noise-reduced features from the graph deep learning modules as input instead of the raw gene expressions by Tangram, which helps to improve projection accuracy.

Currently, we have tested DeepST on 10x Visium, Slide-seqV2, and Stereo-seq. We plan to extend it to handle data from other ST platforms such as MERFISH and Nanostring CosMx SMI. We also anticipate that future developments in ST technologies will bring about subcellular resolution with full gene expression profiling, as well as growth in dataset sizes. Here, the largest dataset we tested contained about 100,000 spots (E14.5 mouse embryo), which took 30 minutes of wall-clock time with Intel Core i7-8665U CPU and NVIDIA RTX A6000 GPU. We believe that our algorithm can handle existing datasets and as well as newer ones in the near future. We also plan to further improve performance through distributed computing and batch training implementations.

## References


1. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).

2. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol.* **39**, 43–58 (2021).

3. Hunter, M. V, Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).

4. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991.e19 (2020).

5. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).

6. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).

7. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

8. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).

9. Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).

10. Pham, D. *et al.* stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *bioRxiv* 2020.05.31.125658 (2020) doi:10.1101/2020.05.31.125658.

11. Zhao, E. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).

12. Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).

13. Li, J., Chen, S., Pan, X., Yuan, Y. & Shen, H.-B. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat. Comput. Sci.* **2**, 399–408 (2022).

14. You, Y. *et al.* Graph Contrastive Learning with Augmentations. in *Advances in Neural Information Processing Systems* (eds. Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H.) vol. 33 5812–5823 (Curran Associates, Inc., 2020).

15. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* **42**, e1900221 (2020).

16. 10x Genomics. https://www.10xgenomics.com/resources/datasets/.

17. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

18. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e18 (2020).

19. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).

20. Fu, X. *et al.* Continuous Polony Gels for Tissue Mapping with High Resolution and RNA Capture Efficiency. *bioRxiv* 2021.03.17.435795 (2021) doi:10.1101/2021.03.17.435795.

21. Cho, C.-S. *et al.* Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559-3572.e22 (2021).

22. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).

23. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).

24. Andersson, A. *et al.* Spatial mapping of cell types by integration of transcriptomics data. *bioRxiv* 2019.12.13.874495 (2019) doi:10.1101/2019.12.13.874495.

25. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).

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

27. Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01273-7.

28. Dong, R. & Yuan, G.-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* **22**, 145 (2021).

29. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).

30. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

31. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

32. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).

33. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

34. Christoffels, V. M. *et al.* Chamber formation and morphogenesis in the developing mammalian heart. *Dev. Biol.* **223**, 266–278 (2000).

35. Allen Brain Atlas: Mouse Brain. https://mouse.brain-map.org/static/atlas.

36. Li, B. *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).

37. V1_Human_Lymph_Node - Datasets - Spatial Gene Expression - Official 10x Genomics Support. https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Human_Lymph_Node.

38. James, K. R. *et al.* Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020).

39. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, (2020).

40. King, H. W. *et al.* Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci. Immunol.* **6**, (2021).

41. Bae, S. *et al.* CellDART: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. *Nucleic Acids Res.* **50**, e57 (2022).

42. Li, M. *et al.* DISCO: a database of Deeply Integrated human Single-Cell Omics data. *Nucleic Acids Res.* **50**, D596–D602 (2022).

43. Carron, E. C. *et al.* Macrophages promote the progression of premalignant mammary lesions to invasive cancer. *Oncotarget* **8**, 50731–50746 (2017).

44. Hu, Q. *et al.* Atlas of breast cancer infiltrated B-lymphocytes revealed by paired single-cell RNA-sequencing and antigen receptor profiling. *Nat. Commun.* **12**, 2186

(2021).

45. Zhang, Y. *et al.* Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* **39**, 1578-1593.e8 (2021).

46. Bassez, A. *et al.* A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).

47. Kipf, T. N. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. (2016) doi:10.48550/ARXIV.1609.02907.

48. Veličković, P. *et al.* Deep Graph Infomax. (2018) doi:10.48550/ARXIV.1809.10341.

49. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014) doi:10.48550/ARXIV.1412.6980.

50. Fraley, C., Raftery, A. E., Murphy, T. B. & Scrucca, L. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. (2012).

51. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).

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

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

## Methods

**DeepST algorithm.** *Data pre-processing.* DeepST takes in gene expression counts and spatial position information. Raw gene expression counts are first log-transformed and normalized to library size via the SCANPY package [33]. The normalized gene expression counts are then scaled to unit variance and zero mean. Finally, the top 3,000 highly variable genes (HVGs) are selected to be input into the DeepST model. Similarly, for scRNA-seq data, raw gene expression counts are first log-transformed and normalized by library size and then scaled to unit variance and zero mean. Subsequently, the top 3,000 highly variable genes are used as input to the model for cell type representation learning. For ST cell composition deconvolution with scRNA-seq data, the common pre-processed HVGs of the scRNA-seq and ST data are used as the input to DeepST.

*Graph construction for spatial transcriptomics data.* Spatial location information in tissue structure can be exploited to measure similarities between spots, which can help identify spatially coherent spatial domains. To make full use of this information, we convert the spatial information into an undirected neighbor graph $G = (V, E)$ with a pre-defined neighbor number $k$. In the graph $G$, $V$ represents the set of spots while $E$ is the set of connected edges between spots. Assuming that $A \in \mathbb{R}^{N_{spot} \times N_{spot}}$ is the adjacency matrix of graph $G$ with $N_{spot}$ denoting the number of spots. If spot $j \in V$ is the neighbor of spot $i \in V$ , $a_{ij} = 1$, otherwise 0. Thus, for a given spot, its neighbors are defined by its proximity to other spots, evaluated by Euclidean distance computed from the spatial location information. Finally, we select the top $k$-nearest spots as its neighbors. From our testing, DeepST achieved its best performance in most of the tested datasets with $k$ set to 3.

*Graph self-supervised* contrastive *learning.* For ST data, we propose a graph self-supervised contrastive learning framework to learn representations for the spots. Figure 1 (A) illustrates the overview of this framework, divided into three major steps: 1) data augmentation, 2) GNN-based encoder for latent representation learning, and 3) contrastive learning for representation refinement. DeepST takes as inputs the pre-processed gene expressions and the neighbor graph constructed above to learn the spot representations for spatial clustering and ST data integration. Details on each step are described next.

*Data augmentation*. Data augmentation is crucial in contrastive learning. Here we first generate a corrupted neighbor graph via data augmentation. Specifically, given a neighbor graph $G$ and the normalized gene expression matrix $X$, we randomly shuffle the gene expression vectors among the spots while keeping the original graph structure unchanged. Let $G' = (V', E')$ and $X'$ denote the corrupted neighbor graph and gene expression, respectively.

*GNN-based encoder for latent representation learning*. The encoder takes the neighbor graph $G$ and gene expression matrix $X$ as inputs, and outputs the reconstructed spot gene expression (i.e., final spot representation) matrix $H_s$. Specifically, we utilize a graph convolutional network (GCN) [47] as encoder to learn a latent representation $z_i$ for spot $i$ by iteratively aggregating the representations of neighbors. The encoder enables the representation $z_i$ to simultaneously capture gene expression profile and spatial similarity. Formally, the $l$-th layer representations in the encoder can be formulated as follows,

$$Z_s^l = \sigma(\tilde{A} Z_s^{l-1} W_e^{l-1} + b_e^{l-1}), \quad (1)$$

where $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ represents the normalized adjacent matrix where $D$ is a diagonal matrix with diagonal elements being $D_{ii} = \sum_{j=1}^{N_{spot}} A_{ij}$. $W_e$, and $b_e$ denote a trainable weight matrix and a bias vector, respectively. $\sigma(\cdot)$ is a nonlinear activation function such as ReLU (Rectified Linear Unit). $Z_s^l$ denotes the $l$-th layer output representation and $Z_s^0$ is set as the original input gene expression matrix $X$. We denote $Z_s$ as the final output of the encoder, where the $i$-th row $z_i$ is the representation of spot $i$.

After that, the latent representations $Z_s$ are fed into a decoder to reverse them back into the raw feature space. Different from the encoder, the decoder adopts a symmetric architecture to reconstruct the gene expression. Specifically, the decoder is defined as follows,

$$H_s^t = \sigma(\tilde{A} H_s^{t-1} W_d^{t-1} + b_d^{t-1}), \quad (2)$$

where $H_s^t$ denotes the reconstructed representations at the $t$-th layer and $H_s^0$ is set as the final output representation $Z_s$ of the encoder. $W_d$ and $b_d$ represent the trainable weight matrix and bias vector, respectively, which are shared by all nodes in the graph. To make full use of the gene expression profile and spatial similarity, we train the model by minimizing self-reconstruction loss of spot gene expression as follows:

$$\mathcal{L}_{recon} = \sum_{i=1}^{N_{spot}} \|x_i - h_i\|. \quad (3)$$

The final output of the decoder $H_s$ is the reconstructed spot gene expression. $x_i$ and $h_i$ are the original normalized gene expression and reconstructed gene expression for spot $i$, respectively.

*Self-supervised* contrastive *learning for representation refinement*. To make the representation $H_s$ more informative and discriminative, we further adopt a self-supervised contrastive learning (SCL) mechanism. Specifically, with the original and corrupted graphs $G$

and $G'$ as inputs to the GNN-based encoder, we can derive two representation matrices $Z_s$ and $Z_s'$ for the spots. Following Velickovic et al. [48], we utilize a readout function $\mathcal{R}$: $Z_s$ $(Z_s') \in \mathbb{R}^{N_{spot} \times d} \rightarrow g$ $(g') \in \mathbb{R}^d$ to aggregate the neighbors' representations into the local context. Taking graph $G$ as an example, for spot $i$ in the graph, its representation $z_i$ and the local context $g$ form a positive pair while $z_i$ and the corrupted local context $g'$ form a negative pair. The key idea behind SCL is to maximize mutual information between the spot representation $z_i$ and its local context $g$ while minimizing the mutual information between $z_i$ and its corrupted local context $g'$. Next, we use binary cross-entropy (BCE) to model SCL. Formally, the contrastive loss can be defined as:

$$\mathcal{L}_{CSL} = -\frac{1}{2N_{spot}}\left(\sum_{i=1}^{N_{spot}} \mathbb{E}_{(X,A)}[log\Phi(z_i, g)] + \sum_{j=1}^{N_{spot}} \mathbb{E}_{(X',A')}\left[\log\left(1 - \Phi\left(z_j', g\right)\right)\right]\right), \quad (4)$$

where $\Phi(\cdot)$ is a discriminator $\mathcal{D}$: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, a dual neural network that distinguishes positive pairs from negative pairs. $\Phi(z_i, g)$ denotes the probability score that is assigned to the positive pair $(z_j, g)$. Since the corrupted graph $G'$ is another view of graph $G$, we define a similar contrastive loss $\mathcal{L}_{CSL\_corrupt}$ for the corrupted graph to make the model more stable and balanced,

$$\mathcal{L}_{CSL\_corrupt} = -\frac{1}{2N_{spot}}\left(\sum_{j=1}^{N_{spot}} \mathbb{E}_{(X',A')}\left[log\Phi\left(h_j', g'\right)\right] + \sum_{i=1}^{N_{spot}} \mathbb{E}_{(X,A)}[\log(1 - \Phi(h_i, g'))]\right), \quad (5)$$

*Overall loss function*. The representation learning module of ST data is trained by minimizing the self-reconstruction loss and contrastive loss. Briefly, the overall training loss of this module is defined as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{recon} + \lambda_2(\mathcal{L}_{CSL} + \mathcal{L}_{CSL\_corrupt}) + \lambda_3(\|W_e\|_F^2 + \|W_d\|_F^2), \quad (6)$$

where the third term is a penalty term to avoid over-fitting. $\lambda_1$, $\lambda_2$ and $\lambda_3$ are weight factors that trade-off the impact of reconstruction loss, contrastive loss, and penalty term. Empirically, we set $\lambda_1$, $\lambda_2$, and $\lambda_3$ as 10, 1, and 0.1. The training of this module is independent of the next integration module and we employ the Adam optimizer [49] for the optimization. The learning rate and training epoch are set to 0.001 and 600 for both spatial clustering and ST data integration tasks, while 0.001 and 1,200 were used for the scRNA-seq and ST data integration task.

*Spatial domain assignment via clustering and refinement.* After model training, we use the reconstructed spatial gene expression $H_s$ from the decoder (Figure 1A) with the non-spatial assignment algorithm, mclust [50], to cluster the spots into different spatial domains. Each cluster is regarded as a spatial domain, containing spots with similar gene expression profiles and spatially proximate. For tissue slices with manual annotation, we set the number of clusters to be the same as the ground truth. For tissue slices without manual annotations, we will test different cluster counts and select the count that gives the highest Silhouette score [51]. Although the reconstructed spatial gene expression $H_s$ was obtained using both gene expression and spatial information, some spots may be wrongly assigned to spatially disparate domains. We consider such occurrences to be noise and their presence may influence downstream biological analysis. To resolve this, we extended our model with an optional optimization step. In this step, for a given spot $i$, its surrounding spots within an $r$ radius circle are treated as its neighbors. DeepST re-assigns the spot $i$ to the same domain as the most common label of its surrounding spots. Setting $r$ to 50 provided the best clustering performance in our benchmarks. This step is not recommended for ST data with fine-grained domains (e.g., mouse brain anterior and posterior), Stereo-seq, and Slide-

seqV2. In this study, we only applied this refinement step to the human brain DLPFC and the human breast cancer dataset.

*Vertical and horizontal integration of multiple tissues via implicit batch effect correction.* The discussion so far assumes only a single tissue slice as input. For biological analysis of tissue samples, integrated analysis of multiple tissue slices can yield greater insights. Two types of multiple sample analysis are possible, vertically split tissue slices (such as the mouse breast cancer sections 1 and 2) and horizontally split tissue slices (such as the mouse brain anterior and posterior sections). For the former, one major challenge for integrated analysis is the presence of batch effects among different slices, hindering data integration. For the latter, the challenge is to assign spots to domains such that the domains straddling the joining edge are aligned.

To overcome these challenges, we extended our DeepST model to handle integrated analysis of multiple tissue slices. Here we consider an example with two slices, though the model can be extended to handle more slices. There are three major steps, as illustrated in Figure 1B. First, for two given tissue slices, we employ the algorithm PASTE [52] to align their H&E images to ensure that the two slices are adjacent in space. Next, with the aligned spatial coordinates, a joint neighbor graph for the two slices is constructed in the same way as with a single slice. The joint construction of the neighbor graph makes it possible to consider both intra- and inter-slice adjacent spots as neighbors for a specific spot, enabling feature smoothing between adjacent spots across slices during the representation learning. Finally, with the joint neighbor graph and gene expressions as inputs, DeepST learns the joint representation of the spots from the two slices for downstream spatial clustering (Figure 1A).

For the vertical integration, DeepST implicitly removes batch effects between slices without explicit batch factor detection. The batch effects originate mainly from the discrepancies of feature distributions between batches. In DeepST, two aspects contribute to batch effect elimination. Firstly, DeepST learns the representation by iteratively aggregating representations of neighbors, which smooths the feature distribution of batches and helps diminish the differences between batches. Secondly, by using the graph self-supervised contrastive learning mechanism, the learned representation captures local context information, which further makes spatially adjacent spots have similar representations.

*PASTE (probabilistic alignment of ST experiments) for multiple tissue slice alignment.* We employed PASTE to align and integrate multiple tissue slices into a single consensus slice. PASTE leverages both gene expression similarity and spatial distances between spots to align and integrate spatial transcriptomics data. In our analysis, we used center slice integration to overcome variability in individual slices due to varying sequencing coverage, tissue dissection, or tissue placement on the array. We first filtered genes with min_counts = 15 in each individual slice using Scanpy. We then filtered for common genes for each individual slice and used PASTE's 'center_align' algorithm or center slice integration mode. In this mode, PASTE infers a 'center' slice consisting of a low rank expression matrix and a collection of mappings from the spots of the center slice to the spots of each input slice, and integrates ST slices to the center slice by combining a fused Gromov–Wasserstein barycenter with non-negative matrix factorization (NMF).

*Spatially informed contrastive learning for scRNA-seq and ST data integration.* To integrate scRNA-seq and ST data, we aim to learn a trainable mapping matrix $M$ of dimensions $N_{cell} \times N_{spot}$ to project cells from scRNA-seq into the spatial space (Figure 1C)). Each

element $M_{ij}$ in $M$ represents the probability of cell $i$ being mapped in spot $j$, with the constraint that the total probability for all the cells is 1 for each spot, i.e., $\sum_{i}^{N_{cell}} M_{ij} = 1$.

The mapping matrix $M$ is learned using the gene expression profiles of scRNA-seq and ST data. Unlike Tangram [29] that learns the mapping matrix directly using raw gene expressions with high noise levels, DeepST retains and refines the informative and noise-reduced features from the raw gene expression profiles of both the scRNA-seq and ST data via deep learning modules before learning the mapping matrix. Specifically, for ST data, we take the output $H_s$ of Module 1 (Figure 1A) as the input for mapping matrix learning (Figure 1C). For the scRNA-seq data, we learn the cell representations via an auto-encoder. Specifically, with the normalized gene expression $e_i$ as input, a latent representation $q_i$ of cell $i$ is learned by an encoder:

$$q_i = f_{en}(e_i), \quad (7)$$

where $f_{en}(\cdot)$ is a multiple-layer neural network. After that, cell gene expression $y_i$ is reconstructed by a decoder:

$$y_i = f_{de}(q_i), \quad (8)$$

where $f_{de}(\cdot)$ is a multiple-layer neural network like the encoder. After model training, we can obtain the reconstructed cell gene expression matrix $H_c$, with which we can predict the spatial gene expression matrix $H_s'$ by combining with mapping matrix $M$:

$$H_s' = M^T \cdot H_c. \quad (9)$$

To learn the mapping matrix $M$, we designed an augmentation-free contrastive learning mechanism to align the predicted spatial gene expression $H_s'$ with the reconstructed spatial gene expression $H_s$ (Figure 1A). The overall loss $\mathcal{L}_{map}$ of the mapping matrix learning is formulated as:

$$\mathcal{L}_{map} = \alpha \sum_{i=1}^{N_{spot}} \sum_{j \in \mathcal{N}_i} -log \frac{\exp\left(sim\left(h_i, h_j'\right)/\tau\right)}{\sum_{p \notin i}^{N_{spot}} \exp\left(sim\left(h_i, h_p'\right)/\tau\right)} + \beta \| H_s - H_s' \|_F^2 , \qquad (10)$$

where $sim(i,j)$ denotes the cosine similarity of spot pair $i,j$ calculated by their representations, $\mathcal{N}_i$ is the set of neighbors of spot $i$, and $\tau$ represents temperature parameter (set as 1 by default). The first term computes the contrastive learning loss. In the contrastive learning architecture, for a given spot $i$, it forms positive pairs with itself and spatially adjacent spots while forming negative pairs with spatially non-adjacent spots. Contrastive learning aims to maximize the similarities of positive pairs and minimize those of negative pairs. The second term is to ensure that the predicted gene expression is proportional to the reconstructed gene expression. We use $\alpha$ and $\beta$ as weight factors to control the weightage consideration of the contrastive and reconstruction losses. $\alpha$ and $\beta$ are set to 1 and 10, respectively.

Unlike most existing deconvolution methods, such as SPOTLight and cell2location, our DeepST model is independent of scRNA-seq annotation information (e.g., cell types) during the mapping process.

*Spot-targeted annotation transfer.* As the output of Module 3 (Figure 1C), we can obtain the mapping matrix $M \in \mathbb{R}^{N_{cell} \times N_{spot}}$, which can be treated as a general transfer function. For certain annotation of scRNA-seq data, such as cell types, disease states, or disease grade, we can transfer it into the spatial space easily via $M$. Briefly, let $S_{cell} \in \mathbb{R}^{N_{cell} \times N_{annot\_cell}}$ denote a one-hot annotation matrix with rows representing cells and columns representing

annotation labels, and $N_{annot\_cell}$ is the number of annotation labels. The probability distribution $P_{spot} \in \mathbb{R}^{N_{spot} \times N_{annot\_cell}}$ of cell annotations in the spots is formulated as:

$$P_{spot} = M^T \cdot S_{cell}. \quad (11)$$

To circumvent the influences of spots with low scores, we empirically retain the scores of the top 10% cells for each spot in $P_{spot}$ and set the remaining values as 0.

*Domain-targeted annotation transfer.* Although we can determine the spatial distribution of annotations (taking cell type as an example) in the H&E images, some cell types may span different spatial domains such that it becomes difficult to distinguish which cell types are distributed in which spatial domains. Therefore, we further extend our model to transfer annotations of scRNA-seq data into the spatial domain. With function (11), we can derive the annotation-to-spot projection matrix $P_{spot} \in \mathbb{R}^{N_{spot} \times N_{annot\_cell}}$ . Assuming that $S_{spot} \in \mathbb{R}^{N_{spot} \times N_{annot\_spot}}$ represents a one-hot annotation matrix with $N_{annot\_spot}$ denoting the number of spatial domains. Similar to the spot-targeted annotation transfer, the probability distribution $P_{domain} \in \mathbb{R}^{N_{annot\_spot} \times N_{annot\_cell}}$ of cell annotations in spatial domains is formulated as:

$$P_{domain} = S_{spot}^T \cdot P_{spot}. \quad (12)$$

**Mouse breast cancer experiments.** All animal work was approved by the NUS Institutional Animal Care and Use Committee (IACUC) and was in accordance with the National Advisory Committee for laboratory Animal Research (NACLAR) Guidelines (Guidelines on the Care and Use of Animals for Scientific Purposes). 1 x $10^5$ metastatic murine breast cancer cells (4T1-12b) tagged with luciferase were implanted into the mammary fat pads of 16-20 weeks-old Balb/c mice. Upon reaching a tumor volume of approximately 100mm$^3$, mice were distributed randomly to experimental groups with and without electroporation with cisplatin and or an adjuvant CPG. Using the ESCOPE machine with a CUTIS probe attached to the machine, mice were intratumorally injected with compounds in various combinations 2-3 min prior to electroporation. 6-8 pulses were given to each tumor following injections with drugs. Tumor growth was monitored over 2 weeks using a digital caliper. Tumors were excised and extracted at 2 weeks for spatial transcriptomics.

*Histology and RNA quality assessment*. 4T1 implanted mammary pad tumors during post-treatment were excised, immediately embedded in optimal cutting temperature (OCT) compound and frozen on dry ice. Cryoblocks were kept at -80 °C and sent for histology to obtain 20 µm sections. RNA quality was assessed for all tissue blocks using the RNeasy Mini kit (Qiagen). Ten sections were obtained from each cryoblock to verify RNA quality for each respective block. 600 uL of buffer RLT was added to the ten sections and subsequently disrupted using a QIAshredder by centrifuging for 2 mins at maximum speed. RNA was extracted from the lysate with the RNeasy Mini kit with instructions from the “Purification of Total RNA from Animal Tissues” section and quality was assessed via RNA integrity number (RIN) value determined using Agilent 2100 Bioanalyzer (Agilent). Cryoblocks, where derived sections had RIN value ≥ 8, were used for subsequent spatial transcriptomics experiments.

*Spatial transcriptomics.* Sections obtained at 20 µm were placed within the grids on pre-chilled Visium Tissue Optimization and Gene Expression slides (10x Genomics) and stored at -80 °C. For H&E staining, slides with sections were thawed at 37 °C for 1 min before fixation in methanol at -20 °C for 30 mins. H&E staining was performed according to manufacturer’s protocol (10x Genomics; CG000160) where hematoxylin staining was reduced from 7 mins to 5.5 mins. Optimal tissue permeabilization time of individual

cryoblocks of each group was obtained from time-course assays performed on tissue optimization slides. This optimal timing was then used on gene expression slides with the respective blocks to capture mRNA post-permeabilization. Images were all obtained and stitched together using EVOS FL Auto 2 (Thermofisher Scientific) with 20x objective (Fluorite with correction collar; 20x objective, 0.7 N.A) and raw images were saved in the TIF format. Fluorescence images were acquired using a RFP filter cube (531/40 nm Ex; 593/40 nm Em) with 20x objective (Plan fluorite). Libraries were then sequenced with Novaseq.

**Data description.** *Spatial clustering.* We employed four spatial gene expression datasets for this task. The first dataset was the LIBD human dorsolateral prefrontal cortex (DLPFC) data acquired with 10X Visium [32] composed of spatial transcriptomics data acquired from twelve tissue slices (http://research.libd.org/spatialLIBD/). The numbers of spots in each slice range from 3,460 to 4,789, with 33,538 genes captured. Each slice was manually annotated to contain between five to seven regions, namely the DLPFC layers and white matter. The second dataset of mouse brain tissue was downloaded from the publicly available 10X Genomics Data Repository (https://www.10xgenomics.com/resources/datasets) . This dataset has two sections and we selected the anterior section. The selected section contains 2,695 spots with 32,285 genes captured and is manually annotated with 52 regions based on the Allen Brain Atlas reference (https://mouse.brain-map.org/static/atlas). The third dataset of the mouse olfactory bulb was acquired using Stereo-seq data [19], which was further processed and annotated [12,53]. The data used contains 19,109 spots and 14,376 genes. The fourth dataset of the mouse hippocampus was acquired with Slide-seqV2, and was downloaded from (https://portals.broadinstitute.org/single_cell/study/slide-seq-study). The section used was Puck_200115_08 with 52,869 spots. The last example consisted of Stereo-seq data acquired from two mouse embryos (E9.5 and E14.5) [19], which we downloaded from https://db.cngb.org/stomics/mosta/. The E9.5 embryo data consisted of 5,913 bins and 25,568 genes, while the E14.5 embryo data consisted of 92,928 bins and 18,566 genes.

In the second task of multiple sample integration, we performed vertical integration on two datasets and one for horizontal integration. For vertical batch integration, both datasets were acquired from mouse breast cancer tissue that is described above. We utilized two sets of samples derived from this tissue, of which each set is composed of two vertically adjacent sections. The numbers of spots range from 1,868 to 3,042 for each section, with 32,285 genes captured. For horizontal integration, we employed both sections (anterior & posterior) of the mouse brain tissue dataset that was also used for clustering.

In the third task of ST cell composition deconvolution with scRNA-seq reference data, we employed both simulated and experimentally acquired data. The first example tested simulated data obtained from the benchmarking study by Li et al. [36], downloaded from https://github.com/QuKunLab/SpatialBenchmarking/tree/main/FigureData/Figure4. We used datasets 4 and 10 from the study. Dataset 4 was generated from mouse cortex experimental data acquired with seqFISH+ and Smart-seq, with 72 spots created. Dataset 10 was created using mouse visual cortex tissue data captured with STARmap and Smart-seq technologies, with 189 spots created. For experimentally acquired datasets, we used four examples. The first example consisted for human lymph node tissue data for spatial transcriptomics and scRNA-seq, both obtained from an existing study by Kleshchevnikov et al. [26]. The ST data contained 4,035 spots while the scRNA-seq consisted of 73,260 cells with 10,217 genes captured. The second example employed the anterior section of the mouse brain sample data. The scRNA-seq reference used was acquired from mouse whole cortex and hippocampus tissues , encompassing more than 1.1 million cells with 22,764 genes captured (https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-

hippocampus-10x). The third example used the human brain sample DLPFC dataset, of which we used slice #151673. The corresponding snRNA-seq data of archived post-mortem dorsolateral prefrontal cortex (BA9) tissue was acquired with the 10X Genomics Chromium platform (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136). The snRNA-seq dataset is composed of 78,886 cells with 30,062 genes [41]. The final deconvolution example utilized the human breast cancer sample obtained from the publicly available 10X Genomics Data Repository (https://www.10xgenomics.com/resources/datasets/human-breast-cancer-block-a-section-1-1-standard-1-1-0). The dataset contains 3,798 spots with 36,601 genes and was manually annotated with 20 regions. The corresponding scRNA-seq data was downloaded from database DISCO [42], consisting of cells from two different samples, i.e., adjacent normal and solid tumor (https://www.immunesinglecell.org/). The data consisted of 476,571 cells with 5,000 genes, which we downsampled to 46,080 cells while maintaining the relative cell type composition.

**Comparison with baseline methods and evaluation**. To showcase the effectiveness of DeepST's spatial clustering, we compared DeepST with five state-of-the-art methods in spatial clustering, including the non-spatial method Seurat [7] and spatial methods Giotto [8], SpaGCN [9], BayesSpace [11], and STAGATE [12]. To evaluate the performance of multiple tissue slice integration, we compared DeepST with STAGATE and SpaGCN for vertical and horizontal integrations, respectively. For ST cell type deconvolution with scRNA-seq data, DeepST was compared with cell2location, which has been benchmarked to be the best existing method [36].

*Seurat*. Seurat is a popular single cell transcriptomics library that had been extended to handle spatial transcriptomics data. We used the default pipeline and default parameters in Seurat as described in the spatial clustering vignette. As we are unable to specify the number of clusters in Seurat, we ran the FindClusters function at different resolutions and chose the parameters that gave us the desired number of clusters.

*Giotto*. Giotto is a toolbox designed for spatial data analysis. We followed the tutorial on Giotto's Github repository with default settings to pre-process the data. For clustering, we used the hidden random Markov random field model based module designed for spatial pattern discovery. Specifically, we used the new Giotto functions, init_hmrf_v2 and do_hmrf_v2 with default parameter values. The HMRF workflow is given in: https://bitbucket.org/qzhudfci/smfishhmrf-r/src/master/TRANSITION.md

*SpaGCN*. SpaGCN is a graph convolutional network approach that integrates gene expression, spatial location information, and histology for ST data analysis. SpaGCN is the only other method that can perform horizontal ST data integration. Following the tutorial, we applied SpaGCN to spatial clustering and horizontal ST data integration with the default parameter settings. In particular, the parameter 'histology' was set to 'False'. The learning rate and max training epoch were set to 0.05 and 200, respectively.

*BayesSpace*. BayesSpace uses a Bayesian model with a Markov random field to model spatial transcriptomics data for clustering, utilizing both spatial and gene expression information. We followed the analysis tutorial for BayesSpace on its GitHub repository, and used the following parameters, nrep = 50,000 and gamma = 3, platform = "Visium" and mode = "normal".

*STAGTE*. STAGATE is another deep learning model that combines an auto-encoder with a graph attention mechanism to learn latent representation by modelling both gene expression profiles and spatial location information. We ran STAGATE for spatial clustering and vertical ST data integration. All experiments were implemented using the recommended parameters

in the package vignette. Specifically, with raw gene expressions, the top 3,000 highly variable genes were first selected, and then log-transformed and normalized according to library size. The parameter 'alpha' was set to 0. The learning rate and training epoch were left at the default 0.0001 and 500, respectively.

*cell2location* cell2location employs a Bayesian model to estimate the spatial distribution of cell types in the ST data of a given tissue using single-cell or nuclei RNA-seq data as reference. We first performed initial pre-processing of removing mitochondrial genes from the ST data. For the scRNA-seq reference dataset, we performed very permissive genes selection using the following parameters: 'cell_count_cutoff = 5'; 'cell_percentage_cutoff = 0.03' and 'nonz_mean_cutoff = 1.12'. We used this selection criteria instead of the standard highly-variable gene selection to retain rare markers genes while removing most of the uninformative genes. To estimate reference cell type signatures from single-cell RNA-seq profiles, we used cell2location with the default negative binomial regression. We used the following parameters: 'max_epochs = 250'; 'batch_size = 2500'; 'train_size=1'; 'lr=0.002'; 'use_gpu=True' to train the model and export the estimated cell abundance for each cell type. In the next step, cell2location performs spatial mapping by taking the spatial dataset and estimated cell abundance for each cell type from scRNA-seq reference dataset as input to output the estimated cell abundance at all spatial locations. Here, we set the hyperparameters: 'N_cells_per_location' (number of cells per location) and 'detection_alpha' (for within-experiment variation in RNA detection) and other parameters such as: 'max_epochs=30000'; 'batch_size=None'; 'train_size=1' and 'use_gpu=True'. The cell abundance in spatial coordinates can be visualized in scatter plots with Scanpy's 'scanpy.pl.spatial' command.

**Data availability**

The data used in this study has been uploaded to Zenodo and is freely available at: https://zenodo.org/record/6925603#.YuM5WXZBwuU.

**Code availability**

An open-source Python implementation of the DeepST toolkit is accessible at https://github.com/JinmiaoChenLab/DeepST.

**Acknowledgements**

The work was supported by A * STAR, AI, Analytics and Informatics (AI3) Horizontal Technology Programme Office (HTPO) seed grant C211118015, the Singapore National Medical Research Council Open Fund Individual Research Grant [NMRC-OFIRG18nov-2013], and the National Research Foundation (NRF-CRP19-2017-04).

**Author contributions**

Jinmiao Chen conceptualized and supervised the project. Yahui Long, Huazhu Fu and Min Wu designed the graph contrastive learning. Yahui Long developed DeepST. Yahui Long, Kok Siong Ang, Jinmiao Chen wrote the manuscript. Yahui Long, Mengwei Li, Kian Long Kelvin Chong, Sethi Raman, Chengwei Zhong, Hang Xu performed the analysis. Yahui Long generated the figures. Zhiwei Ong and Zeng Li annotated and interpreted the brain datasets. Kian Long Kelvin Chong performed the 10x Visium experiments to generate the mouse breast cancer spatial transcriptomic data. Karishma Sachaphibulkij and Hsiu Kim Lina Lim performed the mouse experiments, provided the breast cancer tissues, and annotate the ST data. Ao Chen and Longqi Liu provided the Stereo-seq data of mouse olfactory bulb.

**Competing interests**

The authors declare no competing interests.

## Figure Legends

**Figure 1. Overview of DeepST**. **A)** DeepST takes as inputs the preprocessed spatial gene expressions and neighbor graph constructed using spot coordinates $(x, y)$ of that fall within a distance threshold. Latent representation $Z_s$ is first learned by fully reserving the gene expression profiles, spatial location information, and local context information via the graph self-supervised contrastive learning mechanism. This is then reversed back into the original feature space to reconstruct the gene expression matrix $H_s$. **B)** The analysis workflow for spatial batch effect correction by DeepST. The first step is to align H&E images of two or more samples, followed by shared neighbor graph construction, where both intra- and inter-sample neighbors are considered. This provides the possibility for feature smoothing. Finally, sample batch effects are implicitly corrected by smoothing features across samples with DeepST. **C)** With the reconstructed spatial gene expression $H_s$ and the refined scRNA-seq feature matrix $H_c$ derived from an unsupervised auto-encoder, a cell-to-spot mapping matrix $M$ is trained via a spatially informed contrastive learning mechanism where the similarities of positive pairs (i.e., spatially adjacent spot pairs) are maximized and those of negative pairs (i.e., spatially non-adjacent spot pairs) are minimized. **D)** The outputs $H_s$ and $M$ of DeepST can be utilized for spatial segmentation, multiple ST data integration, and scRNA-seq data transfer onto ST data.

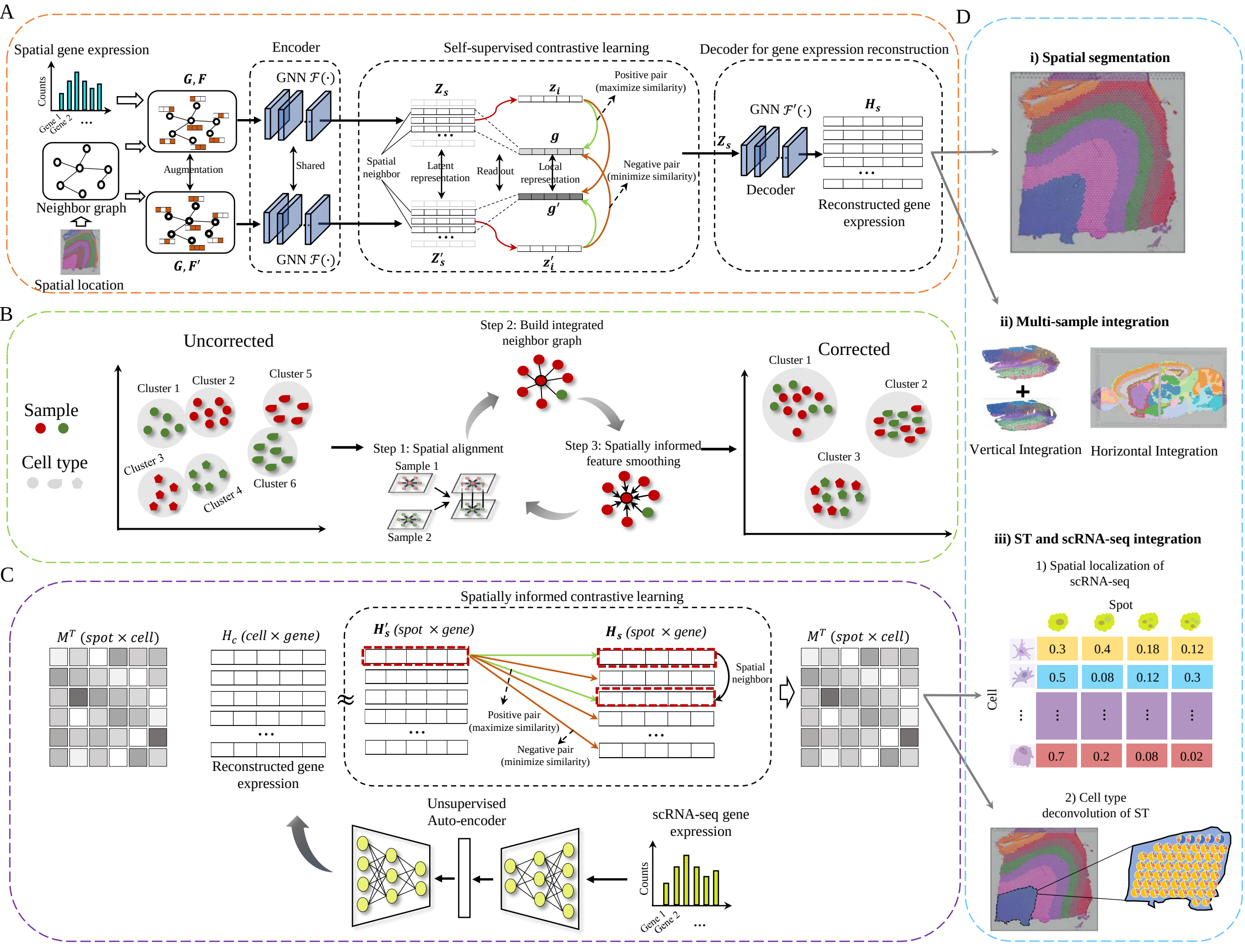
**Figure 2. DeepST clustering improves the identification of tissue structures in the human dorsolateral prefrontal cortex (DLPFC), mouse olfactory bulb, and mouse hippocampus tissue. A.** Boxplots of adjusted rand index (ARI) scores of the six methods applied to the 12 DLPFC slices. In the boxplot, the center line, box limits, and whiskers denote the median, upper and lower quartiles, and 1.5× interquartile range, respectively. **B.** H&E image and manual annotation from the original study. **C.** Clustering results by non-spatial and spatial methods, Seurat, Giotto, SpaGCN, BayesSpace, STAGATE, and DeepST on slice 151673 of the DLPFC dataset. Manual annotations and clustering results of the other DLPFC slices are shown in Supplementary Figure S1. **D**. Laminar organization of the mouse olfactory bulb annotated using the DAPI-stained image. **E**. Spatial domains identified by Seurat, STAGATE, and DeepST, in the mouse olfactory bulb Stereo-seq data. **F**. Visualization of the spatial domains identified by DeepST and the corresponding marker gene expressions. The identified domains are aligned with the annotated laminar organization of the mouse olfactory bulb. **G**. Allen Mouse Brain Atlas with the hippocampus region annotated. **H**. Spatial domains identified by Seurat, STAGATE, and DeepST, in mouse hippocampus tissue acquired with Slide-seqV2. **I**. Visualization of the spatial domains identified by DeepST and the corresponding marker gene expressions. The identified domains are aligned with the annotated hippocampus region of the Allen Mouse Brain Atlas.

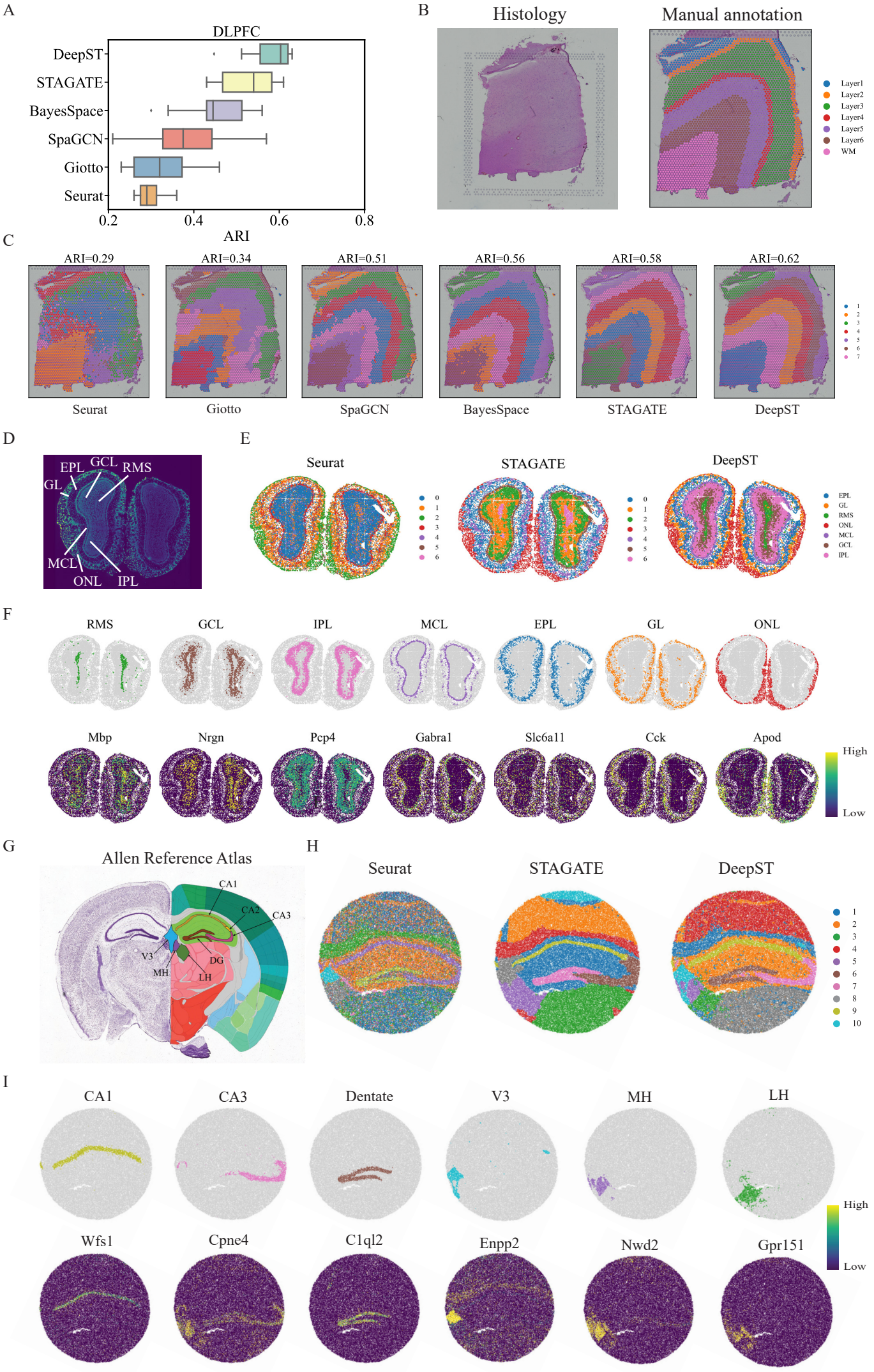
**Figure 3. DeepST enables accurate identification of different organs in the Stereo-seq mouse embryo. A.** Tissue domain annotations of the E9.5 mouse embryo data obtained from the original Stereo-seq study wherein the clusters are first identified using Leiden clustering from SCANPY and then annotated using differentially expressed genes. B. Clustering results of STAGATE and DeepST on the E9.5 mouse embryo data. **C.** Visualization of selected spatial domains identified by DeepST and the corresponding marker gene expressions. **D.** Tissue domain annotations of the E14.5 mouse embryo data obtained from the original Stereo-seq study. **E.** Clustering results by STAGATE and DeepST on the E14.5 mouse embryo. **F.** Visualization of selected spatial domains identified by the original Stereo-seq study and DeepST respectively. **G.** Visualization of marker gene expressions supporting the identified domains.

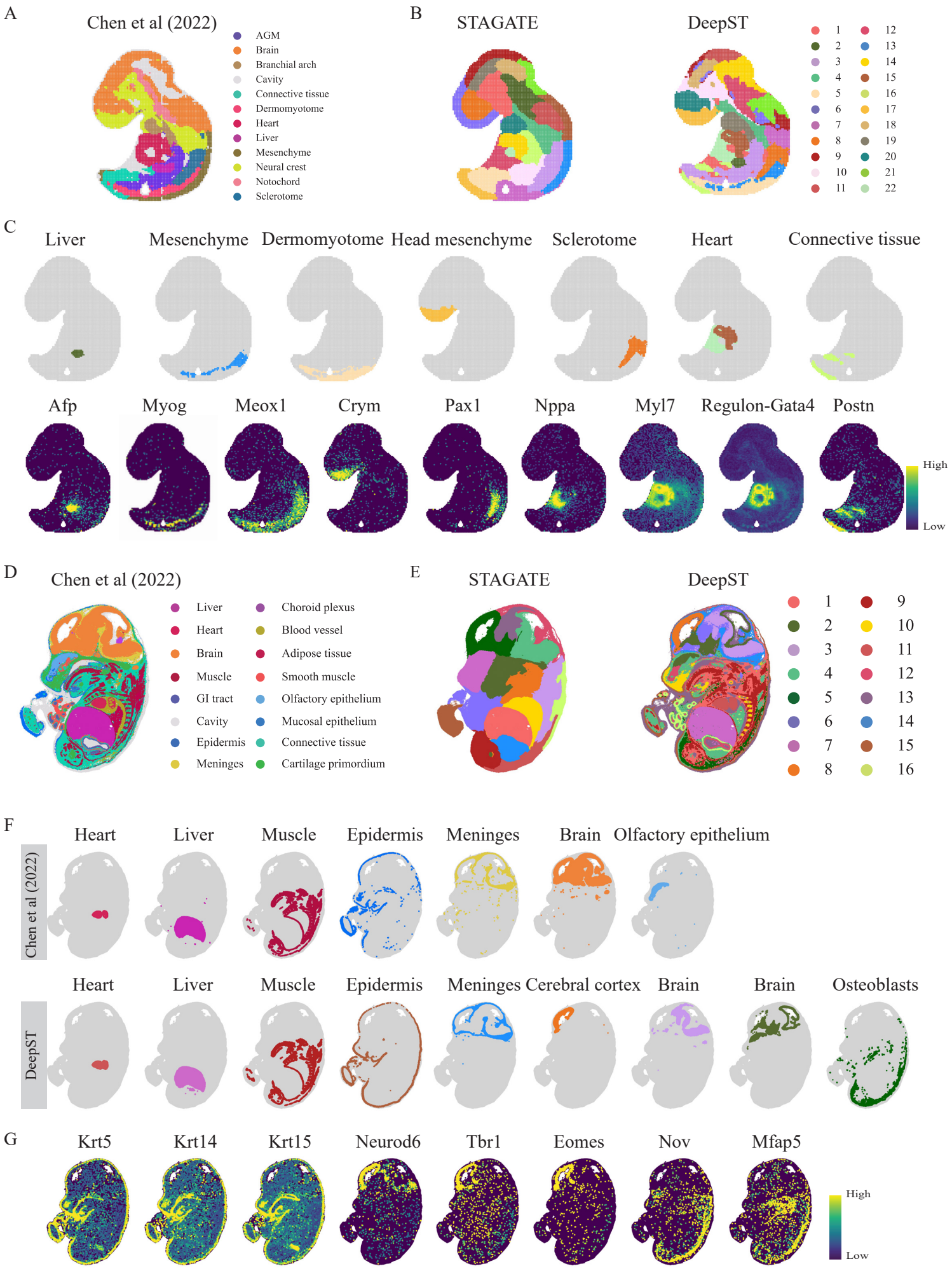
**Figure 4. DeepST enables accurate vertical and horizontal integrations of ST data on mouse breast cancer and mouse brain anterior, and posterior data, respectively. A.** First set of mouse breast cancer sample images aligned by the PASTE algorithm and plotted before batch effect correction. **B.** UMAP plots after batch effect correction and spatial clustering results from STAGATE and DeepST. Spots in the second column are colored according to the spatial domains identified by the respective clustering methods. Here we manually annotate section 2 and the clustering results are compared quantitatively between STAGATE and DeepST using ARI. **C.** Second set of mouse breast cancer sample images aligned and UMAP plot before batch effect correction. **D.** UMAP plots after batch effect correction and the spatial clusters detected by STAGATE and DeepST on sections 1 and 2, respectively. Similarly, spots in the second column are colored according to the spatial domains identified by the respective methods. **E.** Horizontal integration results with two mouse brain samples, of which each consists of an anterior and posterior brain sections. Top: spatial joint domains identified by DeepST on sections 1 and 2. Bottom: spatial joint domains identified by SpaGCN. **F.** Annotated Allen Mouse Brain Atlas for reference.

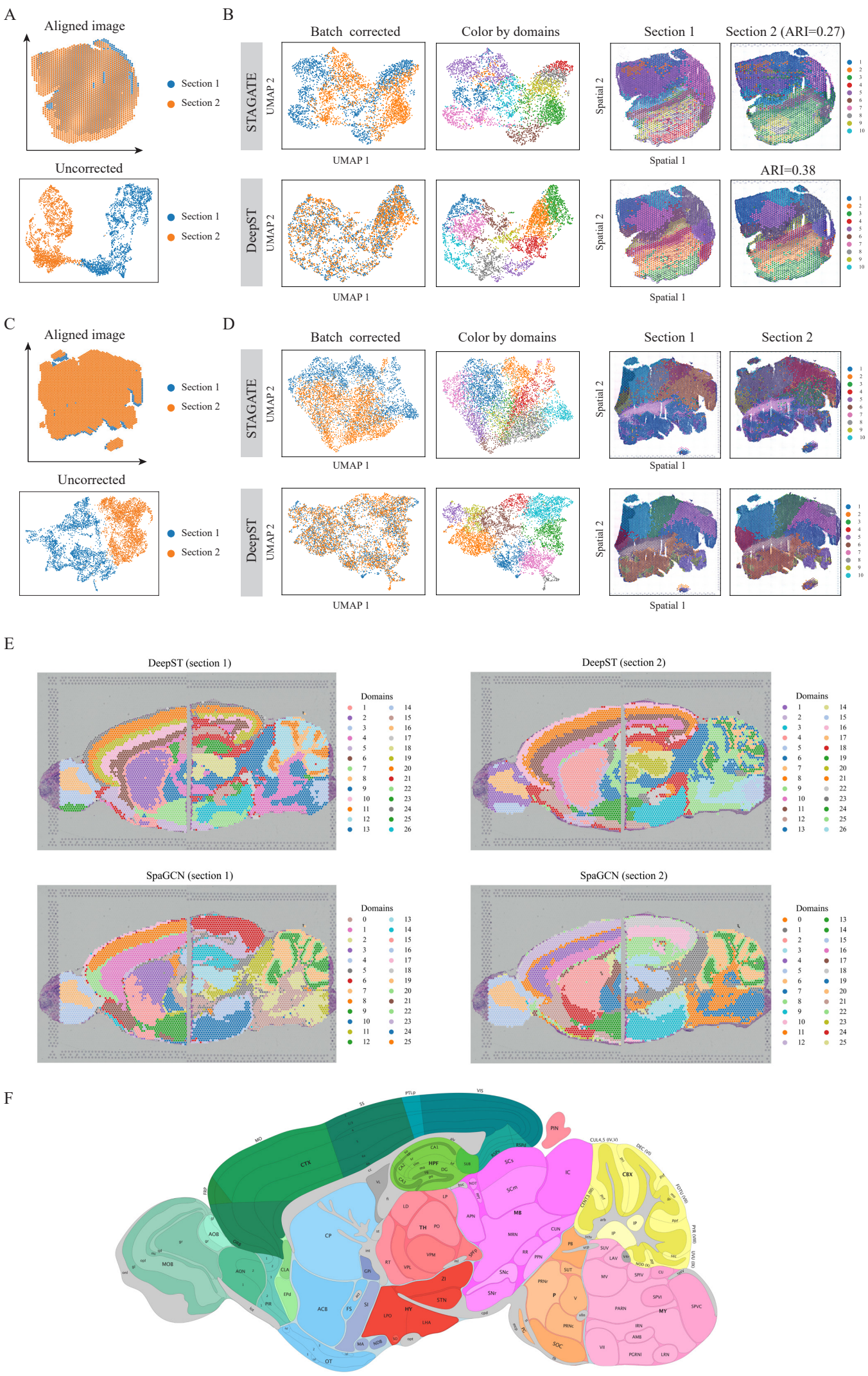
**Figure 5. Comparing the accuracy of DeepST with top deconvolution method cell2location in predicting spatial distributions of scRNA-seq data with simulated data, human lymph node, and the slice 151673 of DLPFC. A.** Boxplots of PCC, SSIM, RMSE, and JSD metrics for cell2location and DeepST results on simulated data created from seqFISH+ and STARmap experimental data. **B**. Left, annotations of germinal center (GC) locations from cell2location's study (GC locations annotated with yellow). Right, H&E image of human lymph node data. **C.** Comparison between cell2location and DeepST on the spatial distributions of selected cell types, namely B_Cycling, B_GC_DZ, B_GC_LZ, B_GC_prePB, B_naive, and B_preGC. **D.** Quantitative evaluation via AUC of three cell types (B_GC_DZ, B_GC_LZ, and B_GC_prePB) localized in the GCs using the annotated locations shown in **B**. **E**. Quantitative evaluation of GC cell type mapping of three cell types (B_GC_DZ, B_GC_LZ, and B_GC_prePB) between cell2location and DeepST using the odds ratio metric. **F**. Comparison between cell2location and DeepST on the spatial distribution of cell types Ex_10_L2_4, Ex_7_L4_6, Ex_1_L5_6, Ex_8_L5_6, Ex_4_L_6, and Oligos_1 with slice 151673 of the DLPFC dataset.

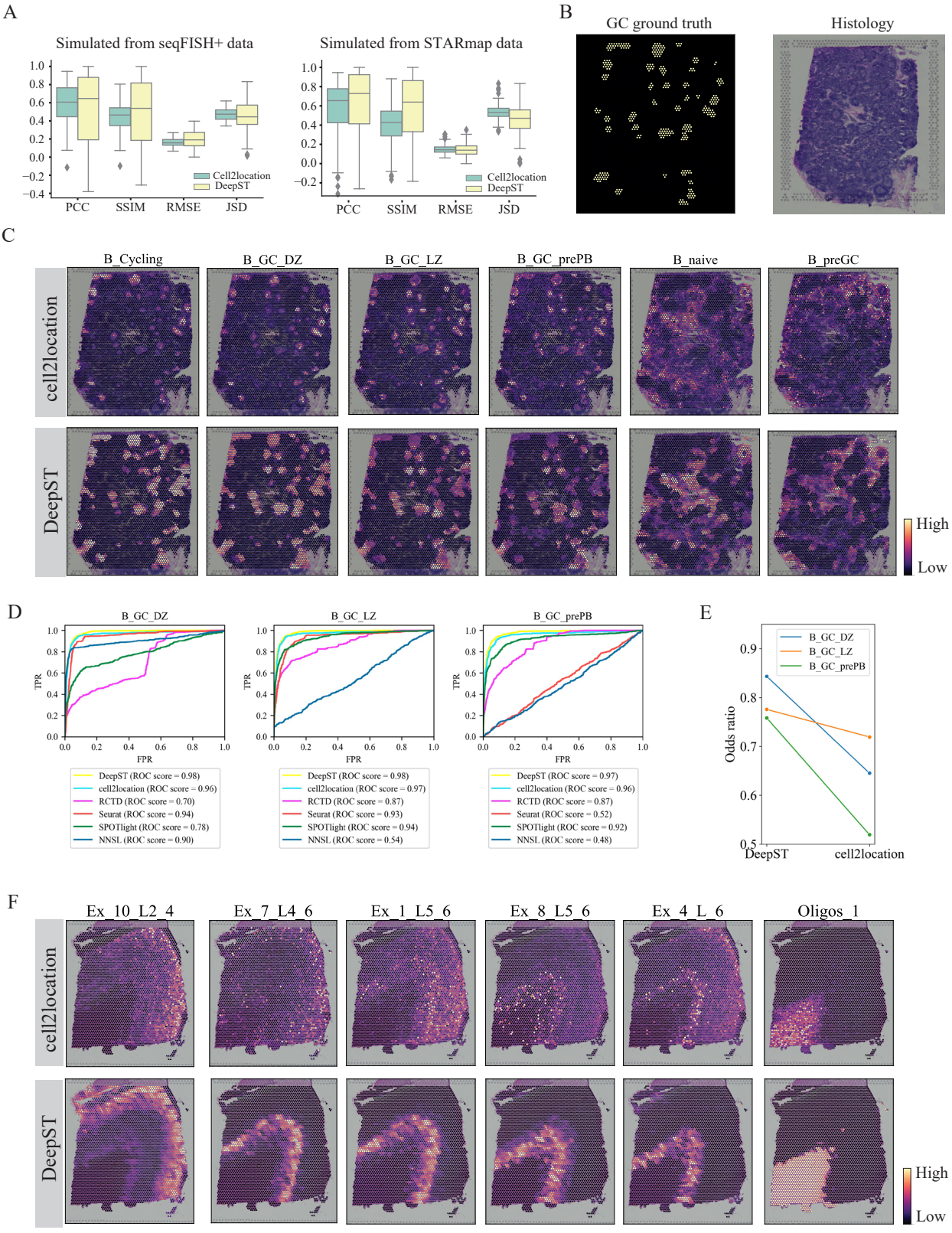
**Figure 6. DeepST enables comprehensive and accurate spatial mapping of scRNA-seq data in human breast cancer data. A.** Manual annotation and spatial distribution of major cell types mapped by DeepST, namely B cell, luminal cell, T cell, fibroblast, lymphatic endothelial cell, NK cell, plasma cell, myoepithelial cell, pDC, luminal progenitor, macrophage/DC/monocyte, muscle cell, and vascular endothelial cell. **B.** Visualization of scRNA-seq data and spatial localization of cell types with UMAP generated from the output cell representations of DeepST. **C.** Heatmap of spatial distribution of cell types. **D.** The gene expression of six T cell exhaustion related markers in different annotated domains. **E.** Functional enrichment results of the IDC domain specific differentially expressed genes. **F.** Predicted spatial distribution of cells from two sample types, adjacent normal and solid tumor.

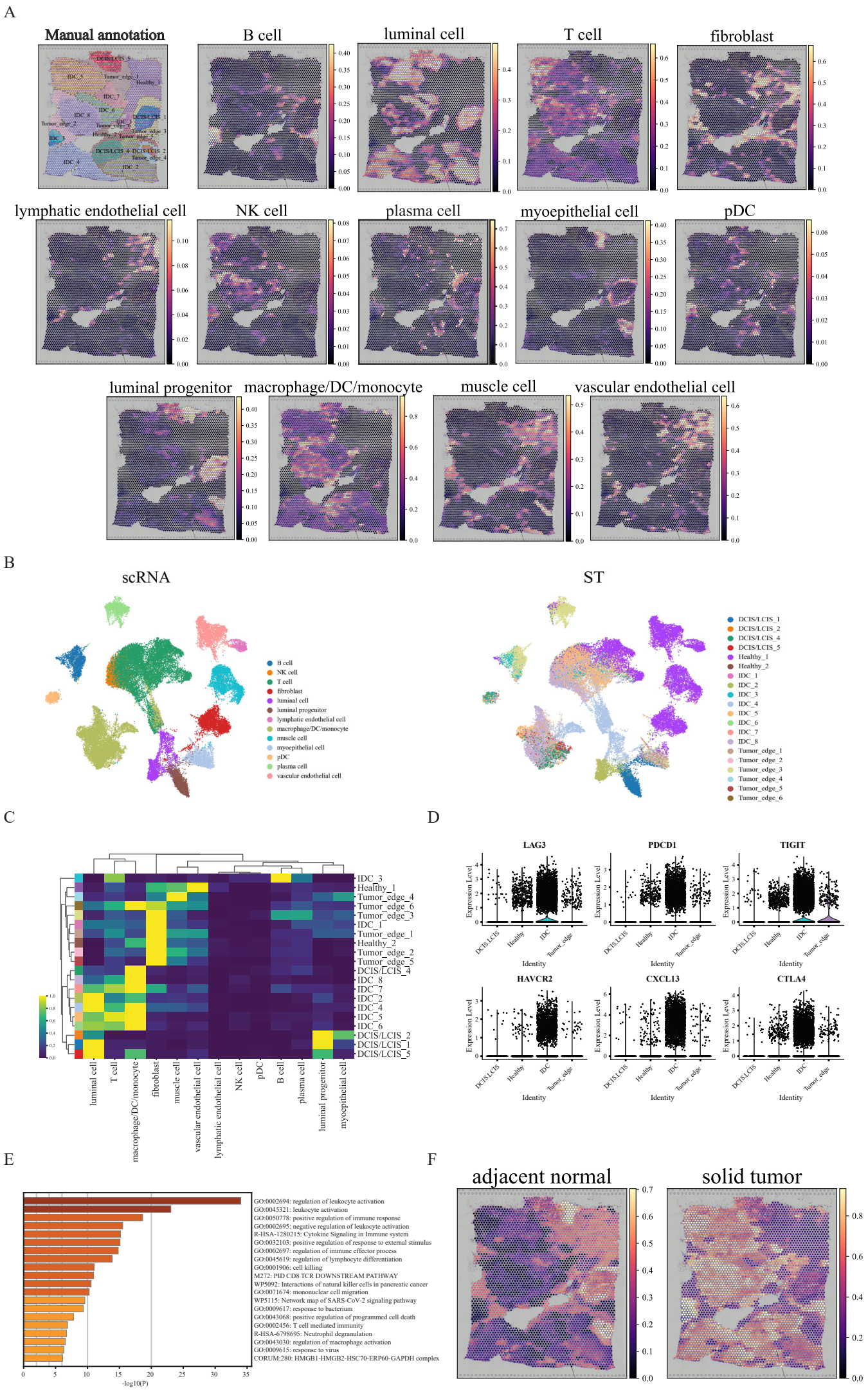

A
Spatial gene expression
Counts
Gene 1
Gene 2
Neighbor graph
Spatial location
G, F
G, F′
Augmentation
Encoder
GNN $\mathcal{F}(\cdot)$
Shared
GNN $\mathcal{F}(\cdot)$
Self-supervised contrastive learning
$Z_s$
$Z'_s$
Spatial neighbor
Latent representation
Read out
Local representation
$z_i$
$g$
$g'$
$z'_i$
Positive pair (maximize similarity)
Negative pair (minimize similarity)
Decoder for gene expression reconstruction
$Z_s$
GNN $\mathcal{F}'(\cdot)$
Decoder
$H_s$
Reconstructed gene expression
D
i) Spatial segmentation
ii) Multi-sample integration
Vertical Integration
Horizontal Integration
iii) ST and scRNA-seq integration
1) Spatial localization of scRNA-seq
Spot
Cell
0.3 0.4 0.18 0.12
0.5 0.08 0.12 0.3
0.7 0.2 0.08 0.02
2) Cell type deconvolution of ST
B
Uncorrected
Sample
Cell type
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Step 1: Spatial alignment
Sample 1
Sample 2
Step 2: Build integrated neighbor graph
Step 3: Spatially informed feature smoothing
Corrected
Cluster 1
Cluster 2
Cluster 3
C
Spatially informed contrastive learning
$M^T$ (spot × cell)
$H_c$ (cell × gene)
Reconstructed gene expression
$H'_s$ (spot × gene)
$H_s$ (spot × gene)
Spatial neighbor
Positive pair (maximize similarity)
Negative pair (minimize similarity)
$M^T$ (spot × cell)
Unsupervised Auto-encoder
scRNA-seq gene expression
Counts
Gene 1
Gene 2


A
DLPFC
DeepST
STAGATE
BayesSpace
SpaGCN
Giotto
Seurat
0.2
0.4
0.6
0.8
ARI
B
Histology
Manual annotation
Layer1
Layer2
Layer3
Layer4
Layer5
Layer6
WM
C
ARI=0.29
ARI=0.34
ARI=0.51
ARI=0.56
ARI=0.58
ARI=0.62
Seurat
Giotto
SpaGCN
BayesSpace
STAGATE
DeepST
D
EPL
GCL
RMS
GL
MCL
ONL
IPL
E
Seurat
STAGATE
DeepST
EPL
GL
RMS
ONL
MCL
GCL
IPL
F
RMS
GCL
IPL
MCL
EPL
GL
ONL
Mbp
Nrgn
Pcp4
Gabra1
Slc6a11
Cck
Apod
High
Low
G
Allen Reference Atlas
CA1
CA2
CA3
V3
MH
DG
LH
H
Seurat
STAGATE
DeepST
I
CA1
CA3
Dentate
V3
MH
LH
Wfs1
Cpne4
C1ql2
Enpp2
Nwd2
Gpr151
High
Low


A
Chen et al (2022)
AGM
Brain
Branchial arch
Cavity
Connective tissue
Dermomyotome
Heart
Liver
Mesenchyme
Neural crest
Notochord
Sclerotome
B
STAGATE
DeepST
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
C
Liver
Mesenchyme
Dermomyotome
Head mesenchyme
Sclerotome
Heart
Connective tissue
Afp
Myog
Meox1
Crym
Pax1
Nppa
Myl7
Regulon-Gata4
Postn
High
Low
D
Chen et al (2022)
Liver
Heart
Brain
Muscle
GI tract
Cavity
Epidermis
Meninges
Choroid plexus
Blood vessel
Adipose tissue
Smooth muscle
Olfactory epithelium
Mucosal epithelium
Connective tissue
Cartilage primordium
E
STAGATE
DeepST
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
F
Chen et al (2022)
Heart
Liver
Muscle
Epidermis
Meninges
Brain
Olfactory epithelium
DeepST
Heart
Liver
Muscle
Epidermis
Meninges
Cerebral cortex
Brain
Brain
Osteoblasts
G
Krt5
Krt14
Krt15
Neurod6
Tbr1
Eomes
Nov
Mfap5
High
Low


A
Aligned image
Section 1
Section 2
Uncorrected
Section 1
Section 2
B
Batch corrected
Color by domains
Section 1
Section 2 (ARI=0.27)
STAGATE
UMAP 2
UMAP 1
Spatial 2
Spatial 1
DeepST
UMAP 2
UMAP 1
Spatial 2
Spatial 1
ARI=0.38
C
Aligned image
Section 1
Section 2
Uncorrected
Section 1
Section 2
D
Batch corrected
Color by domains
Section 1
Section 2
STAGATE
UMAP 2
UMAP 1
Spatial 2
Spatial 1
DeepST
UMAP 2
UMAP 1
Spatial 2
Spatial 1
E
DeepST (section 1)
DeepST (section 2)
SpaGCN (section 1)
SpaGCN (section 2)
Domains
F


A
Simulated from seqFISH+ data
Simulated from STARmap data
Cell2location
DeepST
PCC
SSIM
RMSE
JSD
B
GC ground truth
Histology
C
B_Cycling
B_GC_DZ
B_GC_LZ
B_GC_prePB
B_naive
B_preGC
cell2location
DeepST
High
Low
D
B_GC_DZ
B_GC_LZ
B_GC_prePB
TPR
FPR
DeepST (ROC score = 0.98)
cell2location (ROC score = 0.96)
RCTD (ROC score = 0.70)
Seurat (ROC score = 0.94)
SPOTlight (ROC score = 0.78)
NNSL (ROC score = 0.90)
DeepST (ROC score = 0.98)
cell2location (ROC score = 0.97)
RCTD (ROC score = 0.87)
Seurat (ROC score = 0.93)
SPOTlight (ROC score = 0.94)
NNSL (ROC score = 0.54)
DeepST (ROC score = 0.97)
cell2location (ROC score = 0.96)
RCTD (ROC score = 0.87)
Seurat (ROC score = 0.52)
SPOTlight (ROC score = 0.92)
NNSL (ROC score = 0.48)
E
B_GC_DZ
B_GC_LZ
B_GC_prePB
Odds ratio
DeepST
cell2location
F
Ex_10_L2_4
Ex_7_L4_6
Ex_1_L5_6
Ex_8_L5_6
Ex_4_L_6
Oligos_1
cell2location
DeepST
High
Low


A
Manual annotation
B cell
luminal cell
T cell
fibroblast
lymphatic endothelial cell
NK cell
plasma cell
myoepithelial cell
pDC
luminal progenitor
macrophage/DC/monocyte
muscle cell
vascular endothelial cell
B
scRNA
ST
B cell
NK cell
T cell
fibroblast
luminal cell
luminal progenitor
lymphatic endothelial cell
macrophage/DC/monocyte
muscle cell
myoepithelial cell
pDC
plasma cell
vascular endothelial cell
DCIS/LCIS_1
DCIS/LCIS_2
DCIS/LCIS_4
DCIS/LCIS_5
Healthy_1
Healthy_2
IDC_1
IDC_2
IDC_3
IDC_4
IDC_5
IDC_6
IDC_7
IDC_8
Tumor_edge_1
Tumor_edge_2
Tumor_edge_3
Tumor_edge_4
Tumor_edge_5
Tumor_edge_6
C
D
LAG3
PDCD1
TIGIT
HAVCR2
CXCL13
CTLA4
Expression Level
Identity
DCIS/LCIS
Healthy
IDC
Tumor_edge
E
GO:0002694: regulation of leukocyte activation
GO:0045321: leukocyte activation
GO:0050778: positive regulation of immune response
GO:0002695: negative regulation of leukocyte activation
R-HSA-1280215: Cytokine Signaling in Immune system
GO:0032103: positive regulation of response to external stimulus
GO:0002697: regulation of immune effector process
GO:0045619: regulation of lymphocyte differentiation
GO:0001906: cell killing
M272: PID CD8 TCR DOWNSTREAM PATHWAY
WP5092: Interactions of natural killer cells in pancreatic cancer
GO:0071674: mononuclear cell migration
WP5115: Network map of SARS-CoV-2 signaling pathway
GO:0009617: response to bacterium
GO:0043068: positive regulation of programmed cell death
GO:0002456: T cell mediated immunity
R-HSA-6798695: Neutrophil degranulation
GO:0043030: regulation of macrophage activation
GO:0009615: response to virus
CORUM:280: HMGB1-HMGB2-HSC70-ERP60-GAPDH complex
-log10(P)
F
adjacent normal
solid tumor