



## Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics

Zhen Li, Xiaoyang Chen, Xuegong Zhang, et al.

*Genome Res.* 2023 33: 1757-1773 originally published online October 30, 2023

Access the most recent version at doi:[10.1101/gr.277891.123](https://doi.org/10.1101/gr.277891.123)

---

**References** This article cites 60 articles, 7 of which can be accessed free at:  
<http://genome.cshlp.org/content/33/10/1757.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

## Method

# Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics

Zhen Li,<sup>1</sup> Xiaoyang Chen,<sup>1</sup> Xuegong Zhang,<sup>1</sup> Rui Jiang,<sup>1</sup> and Shengquan Chen<sup>2</sup>

<sup>1</sup>Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China; <sup>2</sup>School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China

Rapid advances in spatial transcriptomics (ST) have revolutionized the interrogation of spatial heterogeneity and increase the demand for comprehensive methods to effectively characterize spatial domains. As a prerequisite for ST data analysis, spatial domain characterization is a crucial step for downstream analyses and biological implications. Here we propose a prior-based self-attention framework for spatial transcriptomics (PAST), a variational graph convolutional autoencoder for ST, which effectively integrates prior information via a Bayesian neural network, captures spatial patterns via a self-attention mechanism, and enables scalable application via a ripple walk sampler strategy. Through comprehensive experiments on data sets generated by different technologies, we show that PAST can effectively characterize spatial domains and facilitate various downstream analyses, including ST visualization, spatial trajectory inference and pseudotime analysis. Also, we highlight the advantages of PAST for multislice joint embedding and automatic annotation of spatial domains in newly sequenced ST data. Compared with existing methods, PAST is the first ST method that integrates reference data to analyze ST data. We anticipate that PAST will open up new avenues for researchers to decipher ST data with customized reference data, which expands the applicability of ST technology.

[Supplemental material is available for this article.]

Recent innovations in spatially resolved transcriptomics grant us a novel perspective on the cellular transcriptome, and hence have stimulated efforts to unravel cell transcriptomes in the context of cellular organizations and catalyze new discoveries in different areas of biological research (Crosetto et al. 2015; Dries et al. 2021; Larsson et al. 2021; Zhuang 2021; Moses and Pachter 2022). Advanced spatial transcriptomic (ST) techniques, including Visium (10x Genomics), Slide-seq (Rodrigues et al. 2019), and Stereo-seq (Chen et al. 2022a), could print spatially barcoded and oligo(dT) probes as microarrays of spots on the surface of glass slides, measuring genome-wide expression levels in captured locations, referred to as spots. When performing exploratory analysis of ST data, one of the most critical steps is characterizing spatial domains, which display spatially organized and functionally distinct anatomical structures on tissues. Nevertheless, low capture efficiency is one of the key reasons for high dropouts in single-cell RNA sequencing, leading to sparse and noisy data (Haque et al. 2017; Qiu 2020), and the limitations are also inherited by spatial transcriptomics data, making characterizing ST data computationally challenging (Zeng et al. 2022). Also, ST data usually represents substantial spatial correlation across tissues, requiring advanced approaches to take spatial localization information into account (Cheng et al. 2022; Zeng et al. 2022; Jiang et al. 2023).

Several computational methods have been proposed to analyze ST data (Dries et al. 2021; Cheng et al. 2022; Zeng et al. 2022). BayesSpace performs spatial clustering via modeling the spatial correlation of gene profiles between neighboring spots (Zhao et al. 2021). SpaGCN identifies spatial domains by integrating gene expression, spatial location and histology through the

construction of an undirected weighted graph (Hu et al. 2021a). CCST is an unsupervised cell clustering method based on graph neural networks for ST data (Li et al. 2022). STAGATE integrates spatial information and gene profiles with an adaptive graph attention autoencoder to learn low-dimensional embeddings (Dong and Zhang 2022). DR-SC uses a hidden Markov random field to perform dimension reduction and spatial clustering within a unified framework (Liu et al. 2022b). Non-negative spatial factorization (NSF) is a spatially aware probabilistic dimension reduction method which assumes an exponentiated Gaussian process prior over the spatial locations (Townes and Engelhardt 2023). SpatialPCA integrates localization information with a kernel matrix and performs spatial-aware principle component analysis (PCA) for ST data (Shang and Zhou 2022). Also, SCANPY and Seurat, two widely-used single-cell analysis workflows in the Python and R communities, respectively, also provide vignettes tailored for ST data analysis (Wolf et al. 2018; Hao et al. 2021).

However, effective approaches for characterizing spatial domains in ST data need to overcome several challenges. First, existing methods failed to incorporate prior information from reference data to facilitate ST data analysis. Incorporating prior information in analyzing single-cell genomic data can better tackle the high level of noise and technical variation, and has been successfully applied in the analysis of various single-cell genomic data (Li et al. 2017; Lareau et al. 2019; Chen et al. 2021; Shengquan et al. 2021). With the advanced technologies, massive amounts of ST data have been generated by consortiums or individual research groups (Rodrigues et al. 2019; Vickovic et al. 2019; Stickels et al. 2021) and accumulated in repositories and databases (Fan et al. 2020; Moses and Pachter 2022; Xu et al. 2022; Zeng et al. 2022; Yuan et al. 2023), paving the way to taking full advantage of the

**Corresponding author:** [chenshengquan@nankai.edu.cn](mailto:chenshengquan@nankai.edu.cn)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.277891.123>. Freely available online through the *Genome Research* Open Access option.

© 2023 Li et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

existing data and leveraging the data as reference to characterize spatial domains. Second, local spatial patterns should be effectively captured. Although it has been observed that aggregating information from each spot's neighbors via a graph convolutional network (GCN) is a promising approach to identify localized gene expression patterns (Hu et al. 2021a; Li et al. 2022; Zeng et al. 2022), tailored model design for better characterization of local spatial patterns is still demanded (Cheng et al. 2022). Third, global spatial patterns should be carefully preserved. With the exponential growth of profiled spots and genes, the graph-based approach becomes cumbersome and time-consuming, requiring strategies that perform memory and time-efficient minibatch training and prediction while preserving global spatial patterns. Fourth, there is an emerging need for the automatic annotation of spatial domains, because advanced protocols constantly increase sequencing throughput, and the typical approach, which performs unsupervised spatial clustering and then assigns the putative domain to each cluster subjectively, would be cumbersome and irreproducible for large-scale ST data.

To this end, we propose PAST, a Prior-based self-Attention framework for Spatial Transcriptomics, which incorporates the graph residual network (Zhang and Meng 2019; Liu et al. 2021) to effectively characterize spatial domains. PAST is a variational graph convolutional autoencoder, and uses a Bayesian neural network (BNN) (Blundell et al. 2015) to incorporate prior information from reference data, self-attention mechanism (Liu et al. 2021) to capture spatial patterns, and a ripple walk sampler (Bai et al. 2021) to enable scalable subgraph-based training and prediction. Featured as the first ST method which integrates reference data from various sources, PAST aims to extract latent embeddings of ST data in a self-supervised manner. The latent embeddings obtained from PAST contain valuable spatial correlation patterns and can be directly fed into existing scRNA-seq computational tools for effective and novel downstream analyses.

## Results

### Overview of PAST

PAST is a self-supervised representation learning model based on a framework of variational graph convolutional autoencoder, taking transcriptomic profiles and spatial coordinates of the target ST data and transcriptomic profiles of the reference data as input. As shown in Figure 1B, the first layer of the encoder decomposes the total variation in ST data into two components (Li et al. 2017; Chen et al. 2021): the component that captures the shared biological variation from reference data based on the BNN module, and the component that captures the unique biological variation in target ST data based on an unconstrained fully connected network (FCN). The choice of the reference data is flexible and can be constructed from various sources (Fig. 1A; [Supplemental Text S1](#)), including: (1) external ST data of the same tissue and from the same protocol with the target ST data, (2) external ST data of similar tissues and from other protocols, (3) scRNA-seq data of similar tissues, or (4) the target ST data itself. In practice, the reference data can be incomplete: novel domains of biological variation that are not captured in the reference data can be present in target ST data. The second component uses an unconstrained FCN to adaptively capture the unique variation in target ST data that is not present in reference data.

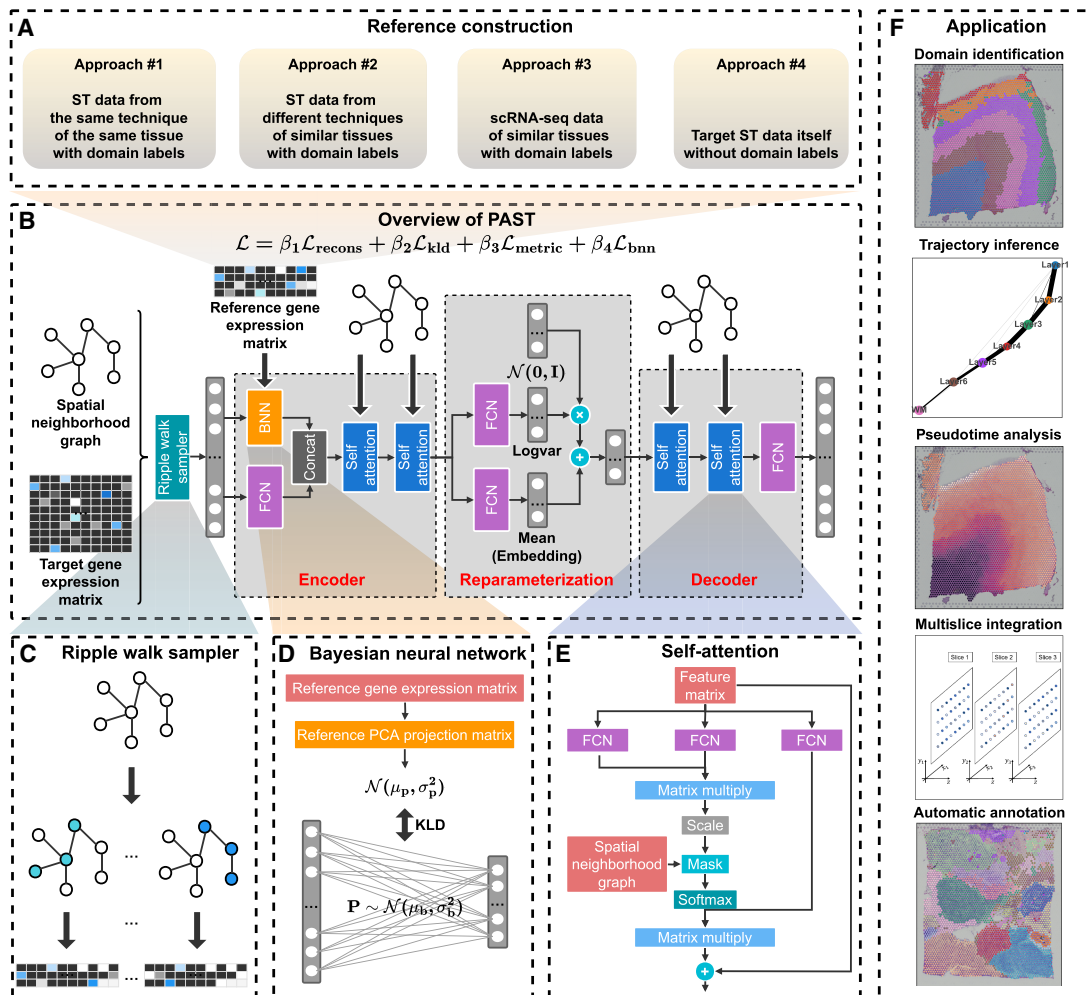
To integrate information of transcriptomic profiles and spatial coordinates, PAST identifies  $k$ -nearest neighbors ( $k$ -NN) for

each spot using spatial coordinates in a Euclidean space, and adopts the graph residual network (Zhang and Meng 2019) with a self-attention mechanism (Liu et al. 2021) to aggregate information from each spot's neighbors. Specifically, the self-attention mechanism, which fits the relationship between words well in machine translation tasks (Raffel et al. 2020), is used to model local spatial patterns between neighboring spots, whereas the ripple walk sampler (Bai et al. 2021), which enables efficient subgraph-based training for large and deep GCNs, is used to achieve better scalability on large-scale ST data and preserve global spatial patterns simultaneously. PAST also restricts the distance of latent embeddings between neighbors through metric learning (Hadsell et al. 2006), the insight of which is that spatially close spots are more likely to be positive pairs to show similar latent patterns, illustrated in detail in the Loss function part of the Methods section. After model training and prediction, the latent embeddings, that is, the mean matrix of the reparameterization module (Fig. 1B), can be applied to spatial domain characterization, spatial visualization, trajectory inference, pseudotime analysis, multislice joint embedding, and automatic annotation (Fig. 1F).

### PAST effectively leverages reference data from various sources for spatial domain characterization

We first use an example as a proof of concept to show the capability of PAST in incorporating reference data from different sources for spatial domain characterization. We collected ST data from a STARmap data set of the mouse primary visual cortex (MPVC) (Wang et al. 2018). We also constructed the reference data from the target ST data itself, a 10x Visium ST data set of the mouse brain coronal region (Palla et al. 2022), and a scRNA-seq data set of the mouse cortex (Tasic et al. 2016), which were referred to as PAST-S, PAST-E-Visium, and PAST-E-scRNA, respectively. The three variants of PAST were benchmarked against eight baseline methods, including two widely-used single-cell analysis workflows in Python and R community (SCANPY [Wolf et al. 2018] and Seurat [Hao et al. 2021]), and six methods for ST data analysis (SpaGCN [Hu et al. 2021a], CCST [Li et al. 2022], DR-SC [Liu et al. 2022b], NSF [Townes and Engelhardt 2023], SpatialPCA [Shang and Zhou 2022], and STAGATE [Dong and Zhang 2022]). We first show the advantages of PAST quantitatively from the following three perspectives. The detailed information about baseline methods and quantitative evaluation metrics are described in [Supplemental Text S2](#).

First, we quantitatively evaluated the performance of spatial domain characterization via supervised cross-validation (Cross-validation). The supervised cross-validation aims to assess the representation capabilities of the latent embeddings learned by various methods. Specifically, we treated the latent embeddings as the input and the labels of spatial domains as the output, adopted support vector machine (SVM) as the classifier as suggested by a benchmark study on supervised single-cell annotation (Abdelaal et al. 2019), and conducted fivefold cross-validation experiments. We evaluated the performance by the average score of accuracy (Acc), Cohen's kappa value ( $\kappa$ ), mean F1 score (mF1), and weighted F1 score (wF1), respectively, in the fivefold experiments. A higher score of the metrics suggests that the latent embeddings can better predict the spatial domains and, consequently, have superior representation capabilities. As shown in Figure 2A and [Supplemental Figure S1A](#), except SpatialPCA, PAST-S and PAST-E-Visium outperform the remaining seven baseline methods for excellent embedding capabilities. Although PAST-E-scRNA performed worse than

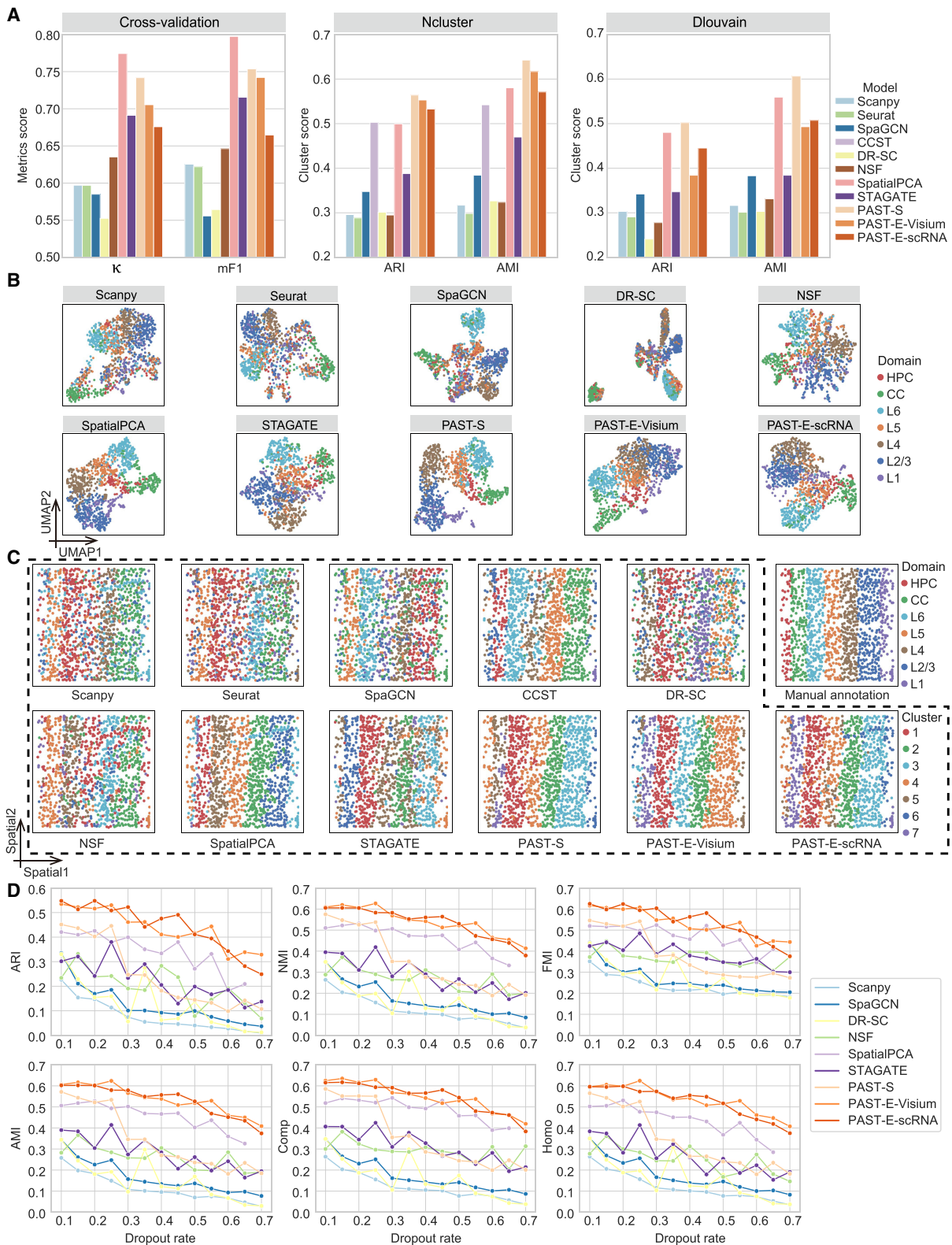


**Figure 1.** The PAST framework. (A) Four approaches for PAST to construct reference data, including three for external reference data (PAST-E) and one from the target data itself (PAST-S). (B) PAST is built on a variational graph convolutional autoencoder. The encoder of PAST consists of three layers, including the concatenation of a BNN and a FCN as the first layer and two self-attention modules as the subsequent two layers. The reparameterization module acquires the mean and log-variance matrix based on two FCNs, in which the mean matrix is the latent embedding obtained of PAST. The decoder of PAST is also a three-layer network, including a FCN layer and two stacked self-attention modules. The loss function of PAST consists of four parts, including reconstruction loss  $\mathcal{L}_{recons}$ , Kullback-Leibler divergence (KLD) loss  $\mathcal{L}_{kld}$ , metric learning loss  $\mathcal{L}_{metric}$  and loss of BNN module  $\mathcal{L}_{bnn}$ . (C) Ripple walk sampler samples high-quality subgraphs based on the spatial neighborhood graph and outputs minibatch gene expression matrices. (D) BNN module integrates shared biological variation through restricting the KLD distance between prior Gaussian distribution parameterized by reference data and Gaussian distribution parameterized by parameters of BNN. (E) The self-attention mechanism captures spatial correlation between neighbor spots, where FCN is used to generate queries, keys, and values for the calculation of attention weights. (F) The latent embeddings, that is, mean matrix in the reparameterization module, obtained by PAST facilitate various tasks including domain identification, trajectory inference, pseudotime analysis, multislice integration, and automatic annotation.

SpatialPCA and STAGATE, which may be caused by the inconsistency of tissues and protocols between ST data and the reference scRNA-seq data, it still outperforms the other six baseline methods. To qualitatively evaluate the representation capabilities of low-dimensional embeddings, we also visualized the spots using uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbor embedding (t-SNE) based on the latent embeddings learned by different methods (Fig. 2B; Supplemental Fig. S2A). Except L2/3, L6, and CC, SCANPY and Seurat could hardly distinguish other domains. SpaGCN, DR-SC, NSF provide relatively better performance. STAGATE succeeded to detect most of the domains but still failed to decipher the heterogeneity of HPC. The three variants of PAST consistently achieve satisfactory visualization results compared with other methods and only SpatialPCA,

PAST-S, and PAST-E-Visium successfully distinguished HPC from other domains.

Second, we compared the performance of spatial domain characterization through the unsupervised spatial clustering with a specified number of clusters (Ncluster), in which the specified number of clusters is set to the number of unique ground-truth spatial domains. We directly used the baseline spatial clustering methods, such as CCST (Li et al. 2022) and SpaGCN (Hu et al. 2021a), or the accompanying spatial clustering strategy of each of the other baseline methods on the learned latent embeddings. Following STAGATE, PAST performed clustering with a specified number of clusters by the mclust algorithm. For SCANPY, Seurat and NSF which perform clustering by Louvain or Leiden, we implemented a binary search to tune the resolution parameter in



**Figure 2.** PAST effectively leverages reference data from various sources for spatial domain characterization. (A) Quantitative performance evaluation of spatial domain characterization on STARmap MPVC data set via supervised cross-validation and unsupervised spatial clustering with a specified number of clusters (Ncluster) and with default resolution (Dlouvain). The cross-validation performance was evaluated by the average score of Cohen's kappa value ( $\kappa$ ) and mean F1 score (mF1), respectively, in the fivefold experiments. The spatial clustering performance was evaluated by adjusted rand index (ARI) and adjusted mutual information (AMI). (B) UMAP visualization results, (C) spatial visualization of clustering with a specified number of clusters and the manual annotation and (D) clustering evaluation of different methods under various dropout rates on MPVC data set.

clustering to make the number of clusters and the specified number as close as possible. We evaluated the clustering performance by six metrics: adjusted rand index (ARI), adjusted mutual information (AMI), normalized mutual information (NMI), Fowlkes-Mallows index (FMI), completeness (Comp), and homogeneity (Homo). As shown in Figure 2A and Supplemental Figure S1A, PAST-S and PAST-E-Visium provide superior performance than all eight baseline methods on all six metrics. PAST-E-scRNA only performs slightly worse than SpatialPCA even when integrating incompatible scRNA-seq data as reference. SCANPY and Seurat, two widely-used single-cell analysis workflows, provided a relatively poor performance even using the vignettes tailored for ST data, indicating the importance of utilizing spatial information. Furthermore, visualization of the spatial domains characterized by clustering with a specified number of clusters and the ground-truth spatial domains on the spatial coordinates of ST data also show consistent results with the quantitative evaluation (Fig. 2C). Note that for CCST, we only evaluated the clustering performance with a specified number of clusters because it is a spatial clustering method and does not provide latent embeddings.

Third, we test the performance of PAST and other methods for the unsupervised spatial clustering with default resolution (Dlouvain), which is useful for the general situation in which the number of unique ground-truth spatial domains is unknown (Cheng et al. 2022). We performed Louvain clustering algorithm, one of the most widely-used community-based clustering methods in single-cell studies, with default resolution based on the latent embeddings learned by different methods. As shown in Figure 2A and Supplemental Figure S1A, PAST-S again achieves overall the best performance whereas PAST-E-Visium and PAST-E-scRNA perform better than any other baseline methods except SpatialPCA. It is vital to emphasize that PAST-scRNA, which incorporates scRNA-seq data to analyze ST data, even achieved 28.2% and 32.0% improvement compared to a competitive ST method STAGATE in terms of ARI and AMI, respectively (31.6%, 18.3%, 29.1%, 34.1% for NMI, FMI, Comp, and Homo, respectively). The evaluation results are detailed in Supplemental Table S1.

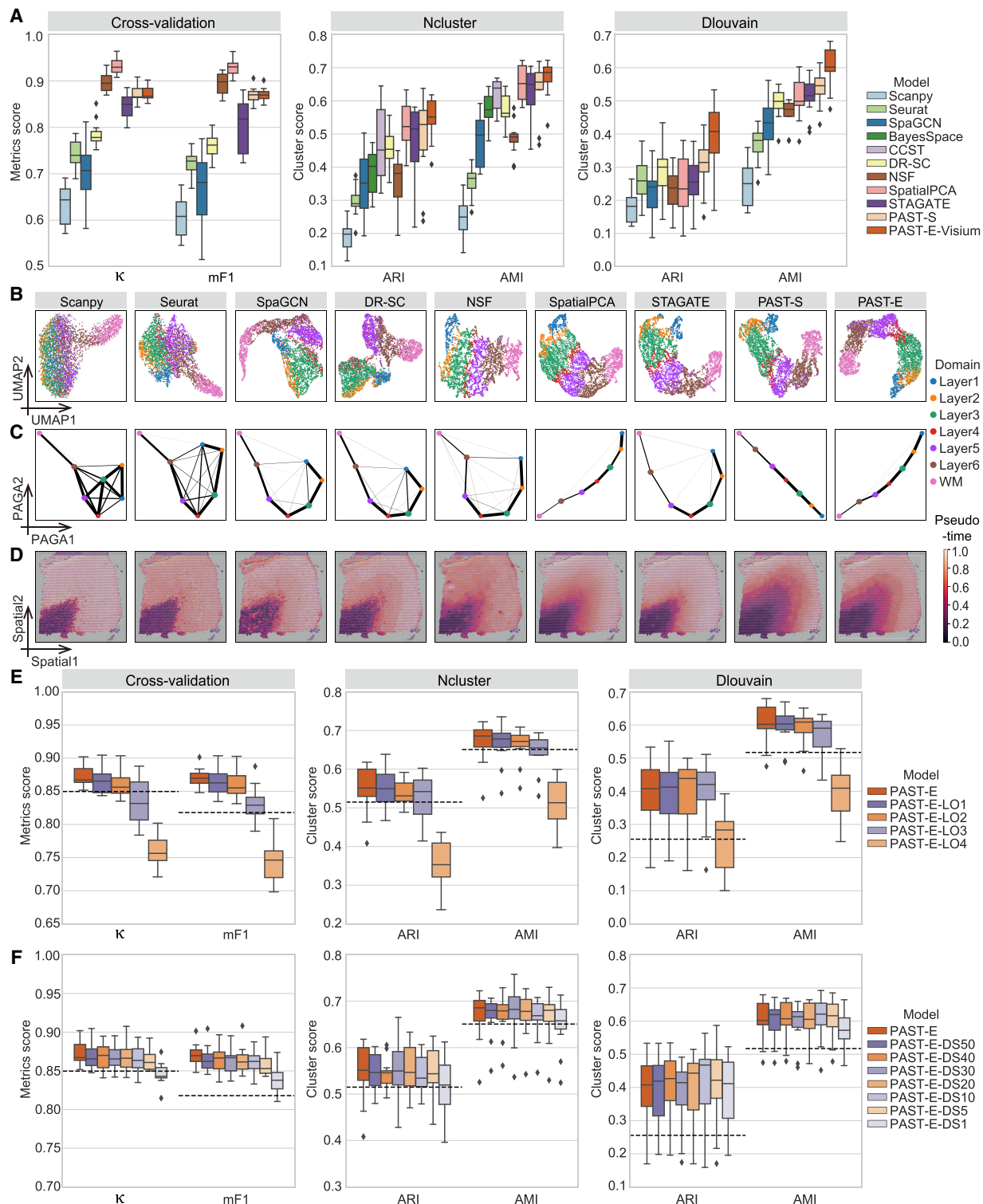
To mimic protocols that provide low coverage, we down sampled the reads in the target ST data by randomly dropping out the entries in the expression matrix to zero with a probability equal to the specified dropout rate, as suggested by a recent benchmark study (Cheng et al. 2022). As shown in Figure 2D, all methods including the three variants of PAST were affected by the increase in the dropout rate, however, PAST with external reference (i.e., PAST-E-Visium and PAST-E-scRNA) consistently outperform the remaining methods over all different dropout rates. Altogether, the results above suggest PAST effectively deciphers spatial patterns through integrating external reference data constructed from various sources and when target ST data has a higher degree of sparsity and noise, the benefit of utilizing external reference data increases. Furthermore, we also showed that PAST is robust to model hyperparameters including the number of neighbors and latent dimensions on the STARmap MPVC data set (Supplemental Figs. S3, S4).

### PAST robustly obtains informative low-dimensional embeddings and facilitates diverse downstream analyses

To test the capability of PAST in learning effective embeddings and accomplishing downstream tasks, we further collected 12 human dorsolateral prefrontal cortex (DLPFC) slices from a 10x Visium data set (Maynard et al. 2021). For each slice, we constructed the

reference data from the ST data itself and from external ST data, namely the remaining 11 slices. The two variants of PAST have different application scenarios and are referred to as PAST-S and PAST-E, respectively. First, we benchmarked PAST from the three quantitative perspectives against nine baseline methods, including the eight methods tested on the STARmap MPVC data set and a clustering method specifically designed for spatial transcriptomics (Ståhl et al. 2016) and 10x Visium data, namely BayesSpace (Zhao et al. 2021). Similar to CCST, we only evaluated the clustering performance with a specified number of clusters. As shown in Figure 3A and Supplemental Figure S5A, for supervised cross-validation, PAST-S and PAST-E performs better than other five baseline methods except NSF and SpatialPCA. Furthermore, PAST-S and PAST-E outperforms all the baseline methods in terms of unsupervised spatial clustering with a specified number of clusters and unsupervised spatial clustering with default resolution on the 12 DLPFC slices (Supplemental Table S2). One-sided paired Wilcoxon signed-rank tests showed that PAST-E showed much more advantages than PAST-S in terms of various clustering metrics in the two different clustering scenarios, with *P*-values of 0.0105 and 0.0002 for ARI, respectively (0.0061 and 0.0002 for AMI, 0.0061 and 0.0002 for NMI, 0.0105 and 0.0005 for FMI, 0.0105 and 0.0002 for Comp, and 0.0134 and 0.0002 for Homo), highlighting the benefit of integrating high-quality external reference data. Similarly, we visualized the spatial domains characterized by clustering with a specified number of clusters and the ground-truth spatial domains on the spatial coordinates of ST data to qualitatively evaluate the performance of different methods (Supplemental Fig. S6). Taking Slice 151,673 as an example, the spatial domain assignment of the nonspatial approaches, namely SCANPY and Seurat, provide poor continuity and smoothness, which is consistent with the above quantitative results. SpaGCN, NSF, and BayesSpace fail to decipher the expected layer pattern in layer 4–layer 6 of this slice, which impairs the spatial clustering accuracy, whereas CCST, DR-SC, SpatialPCA, STAGATE, PAST-S, and PAST-E successfully characterized the expected cortical layer structures, indicating the effective spatial domain characterization.

Second, taking Slice 151,673 as an example, we performed various downstream analyses based on the low-dimensional embeddings obtained by different methods. For low-dimensional visualization based on the UMAP and t-SNE algorithm (Fig. 3B; Supplemental Figs. S7, S8), except the white matter (WM), SCANPY and Seurat could hardly distinguish the cortical layers. SpaGCN DR-SC, and NSF provided relatively better performance, whereas SpatialPCA, STAGATE, PAST-S, and PAST-E successfully identified all the spatial structures and reconstructed the manifold of them. Also, UMAP and t-SNE visualization for the other 11 DLPFC slices provided similar results (Supplemental Figs. S7, S8). For spatial trajectory inference, based on the latent embeddings learned by different methods, we performed trajectory inference using PAGA on Slice 151,673 (Wolf et al. 2019). As shown in Figure 3C, PAST captured the well-known corticogenesis patterns (Gilmore and Herrup 1997; Agirman et al. 2017) and inferred a nearly linear development trajectory from inner to outer layers, indicating the utility of PAST in revealing development trajectory. For spatial pseudotime analysis using the DPT algorithm (Haghighverdi et al. 2016), PAST successfully characterized all the continuous development stages, whereas the other baseline methods only captured the pseudotime variation between WM and other cortical layers (Fig. 3D). Trajectory inference and pseudotime analysis for other slices also provided similar results (Supplemental Figs. S9, S10), highlighting the potential of PAST to extract informative



**Figure 3.** PAST robustly obtains informative low-dimensional embeddings and facilitates diverse downstream analyses. (A) Quantitative performance evaluation of spatial domain characterization on 12 DLPFC slices via supervised cross-validation and unsupervised spatial clustering with a specified number of clusters (Ncluster) and with default resolution (Dlouvain). The cross-validation performance was evaluated by the average score of Cohen's kappa value ( $\kappa$ ) and mean F1 score (mF1), respectively, in the fivefold experiments. The spatial clustering performance was evaluated by adjusted rand index (ARI) and adjusted mutual information (AMI). The center line, box limits and whiskers in the boxplots are the median, upper and lower quartiles, and 1.5 $\times$  interquartile range, respectively. (B) UMAP visualization, (C) PAGA trajectory inference results, and (D) DPT pseudotime analysis results of Slice 151,673. (E) The number of shared spatial domains in external reference data gradually decreased. (F) The spots in external reference data of PAST were down sampled with a stratified random sampling strategy to construct reference data with various scales. The horizontal dashed lines represent the corresponding median scores of STAGATE.

features and facilitate downstream analyses and biological implications. The detailed implementation of visualization, clustering and other downstream analyses are shown in Supplemental Text S3.

The previous example on DLPFC slices used external reference data constructed from the other slices, which contains more than 40,000 spots and heterogeneous information of almost all the domains (PAST-E). However, the prior information in external reference data is generally incomplete for the target ST data and the scale of high-quality reference data is generally limited. Therefore, we further carefully examined the robustness of PAST-E to the number of shared domains and scale of reference data, using STAGATE, a competitive method on DLPFC data set, as the baseline. We gradually left out different number of spatial domains in external reference data, and obtained PAST-E-LO1, PAST-E-LO2, PAST-E-LO3, and PAST-E-LO4, which denote leaving out layer 1, layers 1 and layer 2, layers 1 through layer 3, and layers 1 through layer 4, respectively. The results shown in Figure 3E and Supplemental Figure S11A were in line with intuition because as the spatial domains are continuously left out, the effective prior information in the reference data is decreased. Except PAST-E-LO4, the overall clustering performance of PAST-E when integrating reference with different number of shared domains is superior to STAGATE (the horizontal dashed lines), and PAST-E performs well even based on the reference data containing only nearly half of shared domains. Furthermore, another two different schedules removing the same number of shared domains in external reference data obtained similar results (shown in Supplemental Figs. S12, S13), suggesting the robustness of PAST to the number of shared domains in reference data. Meanwhile, to test the influence of limited scale of reference data without reducing the number of shared domains, we randomly down sampled the spots in reference data based on a stratified random sampling strategy, that is, sampling spots from each domain with a given rate. We obtained PAST-E-DS50, PAST-E-DS40, PAST-E-DS30, PAST-E-DS20, PAST-E-DS10, PAST-E-DS5, and PAST-E-DS1, which denote remaining sampled 50%, 40%, 30%, 20%, 10%, 5%, and 1% of the spots, respectively. Compared with PAST-E which integrated prior information from more than 40,000 spots, the data scale of reference of PAST-E-DS50 through PAST-E-DS1 was only 21,949 through 425. As shown in Figure 3F and Supplemental Figure S11B, PAST-E with down sampled reference data consistently achieves satisfactory performance for supervised cross-validation and spatial clustering, until only 1% of the spots in reference data remained. We then performed two-sided Wilcoxon signed-rank tests on the quantitative results of different reference data scales. As shown in Supplemental Figure S14, the scale of reference data, which may range from a few hundred to tens of thousands, has no significant impact on the performance of PAST-E, indicating that PAST-E is highly robust to the data scale of reference data when the number of shared domains in reference is sufficient.

### PAST captures local spatial patterns with self-attention mechanism

Next, we tested whether PAST could provide insights into local spatial patterns. We applied PAST to an osmFISH mouse somatosensory cortex data set (MSC) (Codeluppi et al. 2018), and used the target data itself as reference. As shown in Figure 4A and Supplemental Figure S15, PAST-S effectively characterizes spatial domains, and achieves significant improvement compared to the baseline methods in terms of cross-validation, and spatial clustering with a specified number of clusters or with default resolution

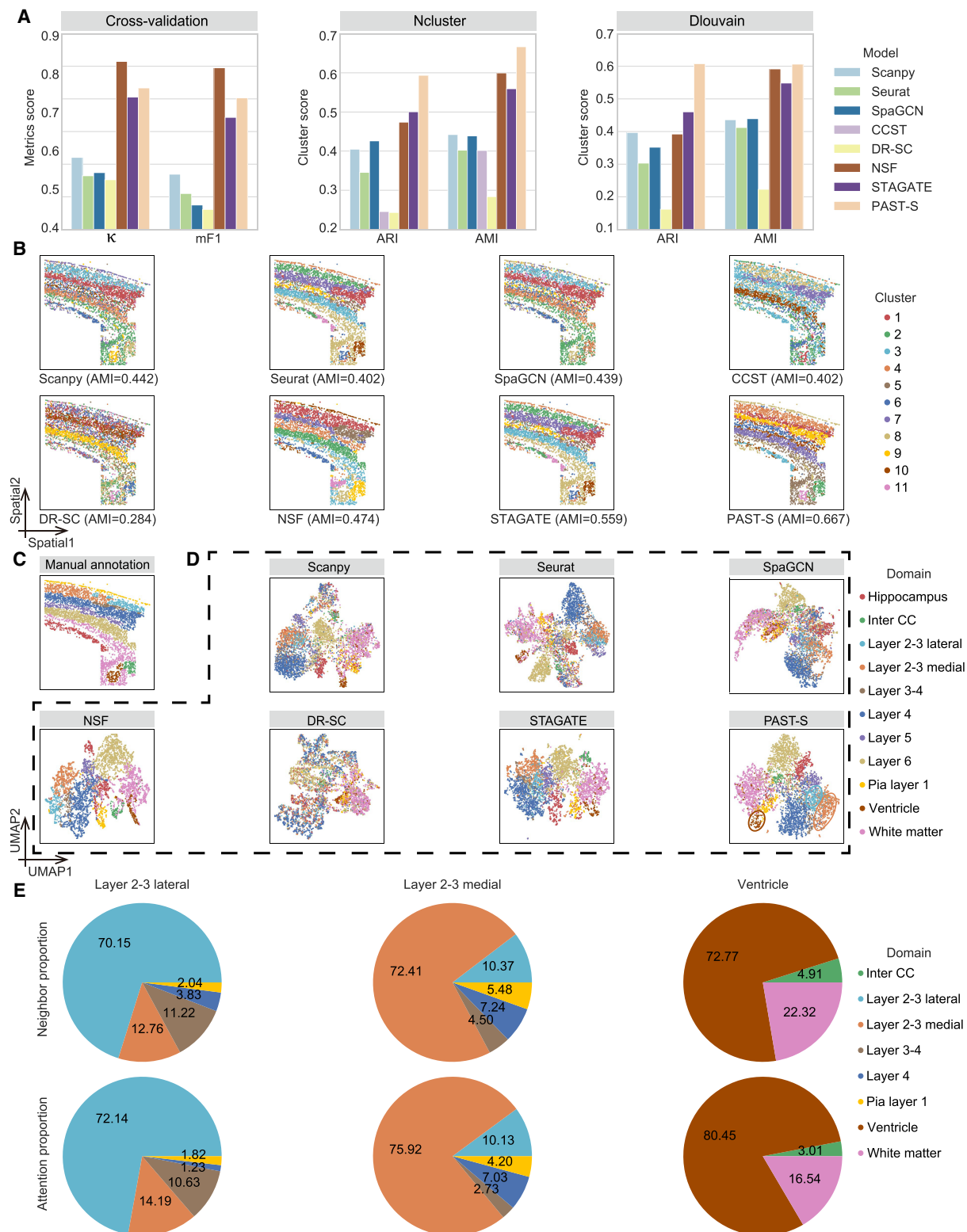
(Supplemental Table S3). Also, PAST with different numbers of neighbors also showed its consistent superiority (Supplemental Fig. S16). For spatial visualization, PAST-S successfully identifies the expected structures and delineated the layer borders clearly, and the identified domains are more consistent with the original annotation of the data set (Fig. 4B,C).

We then explored the capability of the self-attention module of PAST-S in capturing local spatial patterns. As shown in Figure 4D and Supplemental Figure S17, compared with baseline methods like SCANPY, Seurat, DR-S,C and SpaGCN, PAST could effectively identify domains like Layer 2–3 lateral, Layer 2–3 medial, and ventricle in UMAP and t-SNE visualization (blue, orange, and brown circles). For these three spatial domains, we selected the border spots whose spatial neighbors were originated from at least one other spatial domain, and calculated the neighbor proportion and attention proportion of each domain. More specifically, for each border spot of each domain, we recorded the proportion of different spatial domains in the six nearest neighboring spots, and calculated the average proportion for all the border spots of each domain, resulting in the neighbor proportion. In comparison with the neighbor proportion, the attention proportion recorded the attention-weighted proportion of different spatial domains in the six nearest neighboring spots. As shown in Figure 4E, the self-attention mechanism successfully identifies the neighboring spots from the same domain and improves the corresponding weights, that is, from 70.15% to 72.14% for Layer 2–3 lateral, from 72.41% to 75.92% for Layer 2–3 medial, and from 72.77% to 80.45% for ventricle, indicating that self-attention mechanism enables border spots to aggregate more information from neighbor spots belonging to the same domain. Taken together, PAST could effectively capture the local spatial patterns with the self-attention mechanism and characterize the spatial correlation of the spots in ST data, especially the easily confused border spots.

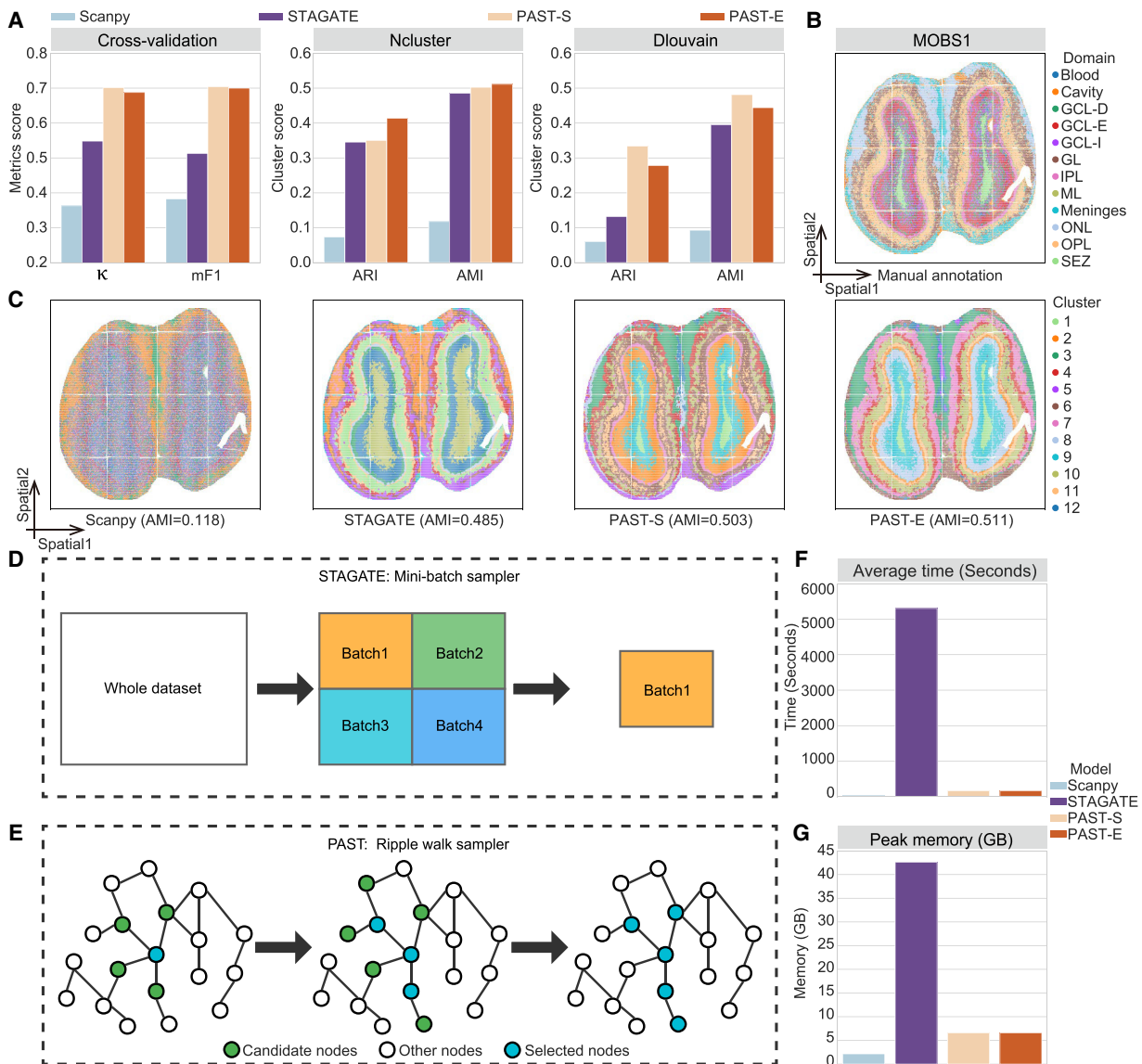
### PAST enables scalable training and prediction on large data while preserving global spatial patterns

With the development of ST technologies, the number of profiled spots and genes increases exponentially, requiring the scalability of computational methods. Traditional minibatch training strategy for deep learning, which partition entire data set into minibatches based on random sampling, is not suitable for graph neural network (GNN) because of the destruction to the connectivity and structure of graph (Bai et al. 2021; Zeng et al. 2022), making GNN-based methods such as SpaGCN (Hu et al. 2021a) and CCST (Li et al. 2022) not applicable to large-scale ST data. STAGATE partitioned the profiled slice and constructed a graph for each part to enable training on large data (Fig. 5D), which may lead to the loss of global spatial patterns (Dong and Zhang 2022). We thus introduced a subgraph-based strategy for scalable training and prediction that considers both the randomness and connectivity of the graph and samples high-quality subgraphs based on a ripple walk sampler (RWS). RWS was originally proposed only for training, and was extended by us from training to prediction by sampling subgraphs covering all spots of the target data set and aggregating the latent embeddings of different subgraphs in an ensemble manner (Methods). Based on the RWS strategy, PAST preserves global spatial patterns and enables efficient application simultaneously.

We collected a Stereo-seq mouse olfactory bulb section (MOBS1) with 107,416 spots and 26,145 genes and analyzed the data set on a standard desktop with an AMD Ryzen 7 Eight-Core



**Figure 4.** PAST capably captures local spatial patterns with self-attention mechanism. (A) Quantitative performance evaluation of spatial domain characterization on the osmFISH MSC data set via supervised cross-validation and unsupervised spatial clustering with a specified number of clusters (Ncluster) and with default resolution (Dlouvain). The cross-validation performance was evaluated by the average score of Cohen's kappa value ( $\kappa$ ) and mean F1 score (mF1), respectively, in the fivefold experiments. The spatial clustering performance was evaluated by adjusted rand index (ARI) and adjusted mutual information (AMI). (B) Visualization of the results of spatial clustering with a specified number of clusters on MSC. (C) The manual annotation of MSC. (D) UMAP visualization of spots in MSC. The blue, orange, and brown circles denote Layer 2–3 lateral, Layer 2–3 medial, and ventricle, respectively. (E) The neighbor proportion and attention proportion for border spots of domains Layer 2–3 lateral, Layer 2–3 medial, and ventricle, respectively.



**Figure 5.** PAST enables scalable training and prediction on large data while preserving global spatial patterns. (A) Quantitative performance evaluation of spatial domain characterization on the Stereo-seq MOBS1 data set via supervised cross-validation and unsupervised spatial clustering with a specified number of clusters (Ncluster) and with default resolution (Dlouvain). The cross-validation performance was evaluated by the average score of Cohen's kappa value ( $\kappa$ ) and mean F1 score (mF1), respectively, in the fivefold experiments. The spatial clustering performance was evaluated by adjusted rand index (ARI) and adjusted mutual information (AMI). (B) The manual annotation of MOBS1. (C) Visualization of the results of spatial clustering with a specified number of clusters on MOBS1. (D) Sampling strategy of STAGATE for minibatch training. (E) Ripple walk sampling strategy for scalable subgraph-based training and prediction. (F) Average time cost of different methods on MOBS1. (G) Peak memory usage of different methods on MOBS1.

3800X CPU, 64GB of RAM and an NVIDIA GeForce RTX 2080 Ti GPU. We collected another Stereo-seq mouse olfactory bulb section (MOBS2) with 104,931 spots and 23,815 genes to construct the external reference for PAST. Note that we encountered memory errors when performing Seurat, SpaGCN, CCST, NSF, and SpatialPCA on MOBS1. As shown in Figure 5A and Supplemental Figure S18, PAST-S and PAST-E, which incorporated self-prior information from MOBS1 itself and external-prior information from MOBS2, respectively, achieve overall the best performance for spatial domain characterization. The quantitative results are listed in Supplemental Table S4. Furthermore, PAST-S and PAST-E also achieve satisfactory performance in UMAP and t-SNE visualization (Supplemental Fig. S19). Also, the spatial domains detected

by PAST-S and PAST-E were spatially continuous and smooth, and the subependymal zone (SEZ) domain, denoted as Cluster 1, was only successfully characterized by these two variants of PAST (Fig. 5B,C). We performed differential expression gene (DEG) analysis for Cluster 1 of PAST-S and PAST-E, respectively. The results of both of these two variants showed that the top two DEGs were *Mbp* and *Ptp1*, which are classical genes encoding myelin proteins (Bitar et al. 2022). Myelin proteins were usually used as markers for striatal contamination or staining to facilitate accurate dissection of the SEZ (Friess et al. 2021; Bitar et al. 2022). We further performed Gene Ontology (GO) analysis for the top 10 DEGs with the lowest adjusted *P*-values of Cluster 1, and identified the most significant molecular function, that is, structural constituent of myelin

sheath, with false discovery rates (FDR) of  $1.28 \times 10^{-4}$  and  $9.33 \times 10^{-5}$  for PAST-S and PAST-E, respectively. The GO analysis results were well consistent with the molecular function of SEZ domain, that is, deriving oligodendrocytes to form mature myelin sheaths around axons (Menn et al. 2006; Kazanis 2009), suggesting the ability of PAST to facilitate functional implication.

We also recorded the computational time and peak memory usage of different methods, especially the two GNN-based methods, that is, STAGATE and PAST, which used different sampling strategies (Fig. 5D,E). As shown in Figure 5F, the average computational time of PAST-S and PAST-E (168 and 164 sec) in 10 replicated experiments were approximately 30× less than STAGATE (5304 sec). In terms of peak memory usage, PAST-S and PAST-E also consumed significantly less memory than STAGATE (Fig. 5G), indicating that the RWS of PAST has a significant advantage over the partition strategy of STAGATE. SCANPY adopts the straightforward representation learning methods, that is, principle component analysis, without incorporating spatial information of ST data, thus providing relatively inferior performance but consuming the lowest computation time (26 Sec). In summary, PAST enables scalable subgraph-based training and prediction while preserving global spatial patterns to better characterize spatial organizations.

### PAST facilitates multislice joint embedding and automatic domain annotation

Advanced ST sequencing technologies like 10x Visium, Slide-seq (Rodrigues et al. 2019), Slide-seq v2 (Stickels et al. 2021), and Stereo-seq (Chen et al. 2022a) generate multiple sequencing slices for the same tissue at one time. However, because of the difference in the dissection of the tissue slices and their placement on array, the spatial coordinate system of multiple slices are different and cannot be easily compared across slices (Zeira et al. 2022), urgently requiring computational methods for multislice joint embedding and automatic domain annotation. We hence proposed two strategies, namely a transfer learning (TL) strategy and a three-dimensional (3D) stacking (3DStack) strategy, based on current representation learning framework to acquire joint embeddings of multiple slices (Supplemental Text S4). Specifically, given multiple slices with inconsistent coordinates and a deep representation learning model, TL strategy chooses one among these slices to train the deep representation learning model and then acquires the joint embeddings of all slices based on the trained model. In comparison, the 3D stacking strategy adds a third dimension to stack the multiple slices in a 3D space and only considers spatial relationship between spots from the same slice by setting the distance between different slices to a large value (Supplemental Fig. S20). Based on the 3D coordinates reconstructed by 3D stacking strategy, current representation learning methods can be easily applied to multiple slices for joint embedding. After obtaining joint embeddings of multiple slices, automatic domain annotation can be easily implemented with a specified classifier (Supplemental Text S4). Specifically, we first split the data into a reference set with domain labels and a target set without labels, then based on the joint embeddings, we trained a SVM on the reference set with its embeddings and labels, and predicted the labels of target set with the trained SVM. Note that the official implementation of DR-SC, SpaGCN, SpatialPCA, and STAGATE cannot use 3D coordinates, whereas SCANPY and Seurat cannot integrate spatial information, thus not applicable to 3D stacking strategy. Nondeep learning models like SCANPY, Seurat, NSF, DR-SC and SpatialPCA are not applicable to TL strategy designed for deep

methods, and acquire the latent embeddings of each slice separately in the subsequent experiments.

We first adopted the 12 DLPFC slices to quantitatively show the integration and annotation performance of different methods based on the two strategies, denoting methods using the TL and 3DStack strategies as \*-TL and \*-3DStack, respectively. We split the DLPFC data set into a reference set, that is, Slice 151,673 with domain labels, and a target set, that is, remaining slices without labels. Naturally, PAST constructed reference data of the BNN module using Slice 151,673 with domain labels. As the UMAP and t-SNE visualization shown in Figure 6A,B and Supplemental Figure S21A–C, methods like SCANPY, Seurat, NSF, DR-SC, and SpatialPCA, which were directly applied to multiple slices and acquired the latent embeddings separately, obtain relatively poor performance, whereas STAGATE-TL, PAST-TL, and PAST-3DStack succeed in batch correction and domain identification, based on the joint embeddings of all the 12 slices obtained by TL and 3DStack strategy. We next used the embeddings and domain labels of Slice 151,673 to train a SVM classifier and annotated the spatial domains in other slices. For quantitative comparison of annotation results, PAST-TL consistently outperforms the baseline methods (Fig. 6C), and achieved significantly higher scores of Acc,  $\kappa$ , mF1, and wF1 than STAGATE-TL, the second-best method, with *P*-values of 0.0034, 0.0024, 0.0005, and 0.0005, respectively, in one-sided paired Wilcoxon signed-rank tests. PAST based on the 3D stacking strategy, namely PAST-3DStack, provides even better performance than PAST-TL and naturally outperforms any other baseline methods. Furthermore, clustering evaluation of the joint embeddings showed similar results (Supplemental Fig. S21D), highlighting the superiority of the TL and 3DStack strategies and PAST in multislice integration and automatic annotation.

We further collected another two 10x Visium human breast cancer sections (BCSs) for integration and annotation, where the first section (BCS1) of the data set is precisely annotated by (Fu et al. 2021) and the second section (BCS2) is not annotated. Naturally, we treated BCS1 (Fig. 6D) as reference set with domain labels and treated BCS2 as target set for annotation. PAST-TL and PAST-3DStack again successfully correct the batch effect between the sections and separate spots from different domains (Fig. 6E,F; Supplemental Fig. S22A,B,D). Based on joint embeddings acquired by different methods, we next trained a SVM classifier with embeddings and labels of BCS1, and annotated BCS2 (Fig. 6G; Supplemental Fig. S22C). For the evaluation of annotation results, we identified marker genes for each domain based on the ground-truth labels on BCS1 (referred to as BCS1-marker genes), and found that BCS1-marker genes of each domain were also differentially expressed in most of the corresponding PAST-TL and PAST-3DStack annotated tumor areas on BCS2 (Supplemental Fig. S23), indicating the ability of PAST-TL and PAST-3DStack to reveal biological variation among tumor areas. In addition, PAST successfully captured BCS2-specific spatial domains by separating the IDC\_3 region on BCS1 into IDC\_3 and Tumor\_edge\_3 on BCS2 (black circles in Fig. 6D,G). The BCS1-marker genes of Tumor\_edge\_3 also showed the unique spatial pattern of the newly annotated Tumor\_edge\_3 on BCS2, which further suggests the accuracy of annotation results on BCS2 (Fig. 6H). We also performed DEG analysis for Tumor\_edge\_3 region based on the annotation results of PAST-TL and PAST-3DStack on BCS2, and identified the same five most significant DEGs including *IGHG3*, *IGKC*, *IGLC2*, *IGHG4*, and *IGHG1* with adjusted *P*-values of  $1.11 \times 10^{-83}$ ,  $1.86 \times 10^{-78}$ ,  $9.18 \times 10^{-74}$ ,  $9.18 \times 10^{-74}$ ,  $1.66 \times 10^{-69}$  for PAST-TL and  $4.52 \times 10^{-74}$ ,  $2.60 \times 10^{-69}$ ,  $8.11 \times 10^{-66}$ ,  $1.75 \times 10^{-62}$ ,  $3.21 \times 10^{-60}$



for PAST-3DStack, respectively. Based on previous literature, the linkage of *IGHG1/3/4* provides information for prognosis of diseases like autoimmunity and malignancy (Oxelius and Pandey 2013), *IGKC*, serving as a robust immune marker, could predict metastasis-free survival and response to chemotherapy (Schmidt et al. 2012), and *IGLC2* was proven to be a significant prognostic biomarker for triple-negative breast cancer patients (Chang et al. 2021). The Gene Ontology enrichment for the top 10 DEGs of Tumor\_edge\_3 also significantly revealed lots of biological immune processes for both PAST-TL (Supplemental Fig. S24) and PAST-3DStack model (Supplemental Fig. S25).

PAST-TL and PAST-3DStack also successfully captured the biological difference between DCIS/LCIS\_1 and DCIS/LCIS\_2 (blue and orange circles in Fig. 6G) and accurately annotated DCIS/LCIS\_2 on BCS2, which was missed by other methods. The spatial expression visualization of BCS1-marker genes of DCIS/LCIS\_2 on BCS2 illustrated the accurate annotation of DCIS/LCIS\_2 by PAST and the two strategies (Fig. 6H). To summarize, based on the two strategies, PAST cannot only effectively facilitate multislice joint embedding and automatic domain annotation, but also identify test set-specific spatial patterns and reveal the corresponding biological implications.

## Discussion

In this article, we introduce PAST, a comprehensive and scalable framework that effectively integrates prior information via Bayesian neural networks, deciphers global and local spatial correlations based on graph residual network integrated with self-attention mechanism, and enables scalable training and prediction on large data sets with RWS strategy. With comprehensive experiments on multiple data sets generated by various technologies, we have shown that PAST capably incorporates prior information from various sources to better tackle the high-level technical variation and effectively integrates spatial information to characterize spatial domains, highlighting the superior information fusion capabilities of PAST. We have also showed the successful applications of PAST in downstream tasks including spatial domain characterization, ST visualization, spatial trajectory inference, and pseudo-time analysis. Experiments on large-scale data sets also proved the scalability and effectiveness of PAST. Furthermore, PAST not only enables multislice joint embedding and automatic annotation of spatial domains for multiple ST data, but also provides biological insights into the annotated domains.

PAST stands out as a representation learning method tailored for ST data by incorporating reference data, which can be constructed from four different approaches (Methods). This inherent property of PAST showcases remarkable potential in deciphering ST data, especially in light of the ever-growing reservoir of both ST data and scRNA-seq data, while it simultaneously poses challenges in selecting top-notch reference data for better performance. To offer practical suggestions and recommendations for users, we diligently constructed reference data encompassing all the four approaches for the data sets in the aforementioned experiments and reported their performance outcomes (Supplemental Text S5; Supplemental Fig. S26), which highlights the superiority of PAST even when integrating reference from completely different technologies. Nevertheless, owing to the potential uncertainty stemming from incongruity of external reference, it is advisable to prioritize reference data from the same technique, of the same tissue type, containing more shared domains, and of a larger scale. Furthermore, we also proposed three recommendation scores, that is, maxMAP,

maxSC, and maxCHS, to guide users to select high-quality reference data in an adaptive way (Supplemental Text S5). The three recommendation scores assess the latent embeddings acquired by PAST with different external reference data, signifying stronger recommendation levels through higher values. Finally, the scores have been proven to effectively steer users towards selecting the optimal approach for reference construction across all data sets, while maintaining satisfactory computational time (Supplemental Fig. S26; Supplemental Tables S5, S6), which provides an adaptive, accurate, and scalable approach for users to select optimal reference data in ST data analysis.

The self-attention mechanism of PAST is implemented to adaptively learn attention weights and proven to capably aggregates more information from neighbor spots belonging to the same domain. This capability can be translated into several practical benefits for researchers to analyze ST data, including interpretable visualization, effective clustering, and accurate annotation, which has been systematically showed in the aforementioned experiments. In addition, we further envision an extension of the self-attention mechanism, which models the relationship between spatially close spots based on the transcriptional information, to unravel intricate cell-cell interactions (CCI), because recent research has underscored the reliability of deciphering CCI from transcriptomic data (Armingol et al. 2021) and the strong relationship between spatial structures and CCI in the microenvironment (Asp et al. 2020; Liu et al. 2022c). Massively parallel sequencing-based technologies like 10x Visium, Slide-seq (Rodrigues et al. 2019), Slide-seq v2 (Stickels et al. 2021), and Stereo-seq (Chen et al. 2022a) generates multiple sequencing slices for the same tissue with inconsistent coordinates. We propose a 3D stacking strategy to stack the slices into a 3D coordinate system and only consider the attention weights between spots from the same slice (PAST-3DStack). A recent method, called PASTE (Zeira et al. 2022), has been proposed to align multiple adjacent tissue slices into a consistent spatial coordinate system. Based on the coordinates obtained by PASTE, PAST capably implements cross-slice attention weights and obtain joint embeddings of multiple slices, denoted as PASTE-PAST. We compared PAST-3DStack with PASTE-PAST and other baseline methods for joint embedding on three donors of the human DLPFC data set, highlighting the superiority of PAST-3DStack (Supplemental Text S6; Supplemental Figs. S27, S28). Nevertheless, compared with PAST-3DStack, the combination of PASTE and PAST enables the characterization of cross-slice attention and inspires us to explore the extension of self-attention mechanism to model cross-data-set attention for multislice integration.

Generalizability of computational methods across technologies, species, and tissue types holds immense significance. To test the consistency of PAST across various biological systems, we further extend our model to another two data sets of different species and tissue types, namely a Stereo-seq axolotl brain data set (Wei et al. 2022) and a Stereo-seq zebrafish embryo data set (Liu et al. 2022a) (Supplemental Text S7; Supplemental Figs. S29, S30). The experimental results showed the superiority of PAST in term of spatial clustering, UMAP, t-SNE, and spatial visualization. We also implemented PAST-3DStack to the Stereo-seq zebrafish embryo data set, which is composed of 12 slices, for joint embedding, illustrating the superiority of PAST and the 3D stacking strategy in multislice integration across different species (Supplemental Text S8; Supplemental Fig. S31). In summary, we have showcased the superior performance of PAST across four sequencing technologies, encompassing four species and spanning seven tissue types,

which underscores the broad applicability and unwavering consistency of PAST across various biological systems.

Certainly, there are several avenues for improving PAST. First, given that Chen et al. achieved satisfactory annotation and simulation performance in scATAC-seq data through assuming the latent embeddings follow a Gaussian mixture distribution (Chen et al. 2022b), we can explore the application of PAST in ST data simulation by adding more complex assumptions on latent distribution. Second, with the rapid development of ST sequencing technologies, many massively parallel sequencing-based methods are often complemented by high-resolution hematoxylin and eosin (H&E) histology images, demanding methods like PAST to incorporate cross-modality information which contains valuable heterogeneity and pathological information (Hu et al. 2021b). Third, the rapid development of graph neural network and large language model field provide advanced methodology to model the relationship between connected nodes in graph or related words in sentence, which also opens new horizons for better characterization of relationship between neighboring spots. Finally, Chen et al. have already proven the effectiveness of leveraging reference data in the analysis of scATAC-seq data (Chen et al. 2021), so we anticipate the extended application of PAST in other spatial omics data like spatial-ATAC-seq data (Deng et al. 2022; Gao et al. 2023).

## Methods

### The model of PAST

PAST is a self-supervised representation learning framework built upon variational graph convolutional autoencoder (Kingma and Welling 2014) and integrates a graph residual network (Zhang and Meng 2019; Liu et al. 2021) to model the node dependency in spatial graph. It takes three kinds of data as input, including pre-processed target gene expression matrix, target spatial coordinate matrix and reference gene expression matrix (Fig. 1B). The details about data preprocessing are shown in Supplemental Text S9.

#### Encoder

The encoder of PAST is a three-layer neural network. The first layer consists of a BNN module and a parallel FCN, and the output of the first layer is the concatenation of the two modules. It decomposes the total variation in ST data into two components (Li et al. 2017; Chen et al. 2021): the component that captures the shared biological variation from reference data with a BNN module, and the component that captures the unique biological variation in target ST data with an unconstrained FCN. This property has been validated through ablation study of BNN module and the FCN module, shown in Supplemental Text S10 and Supplemental Figures S32, S1B, S5B, and S15B. The second and third layers are two stacked self-attention modules. The input of self-attention module includes a feature matrix from the last layer and the  $k$ -NN graph constructed with spatial coordinates of the target data.

#### Reparameterization module

As suggested by variational autoencoder (VAE), we assume the prior of latent embeddings is centered isotropic multivariate Gaussian distribution, use two FCNs to obtain the mean and logarithmic variance matrices of latent space based on the output of encoder, and obtain normally distributed samples through the reparameterization trick. We emphasize that the mean matrix of this module is the final output of PAST, that is, latent embeddings.

#### Decoder

The decoder of PAST is also a three-layer network, in which the first and second layers are stacked self-attention modules and the third layer is a FCN. The decoder takes multivariate normally distributed samples from the reparameterization module as input and then outputs the reconstructed gene expression matrix of the target data.

#### Training and prediction workflow

In the forward propagation phase of training workflow, a  $k$ -NN graph is first constructed based on the target spatial coordinate matrix. Second, PAST samples high-quality subgraphs using the ripple walk training strategy based on the  $k$ -NN graph, providing subgraphs and minibatch gene expression matrices for encoder. The encoder of PAST takes the subgraphs and the target and reference gene expression matrices as input, in which the subgraphs serve as the spatial neighborhood graph for the self-attention module and the reference gene expression matrix provides prior parameters for BNN module. Third, reparameterization module uses the output feature matrix of encoder to acquire the parameters of latent Gaussian distribution based on two FCNs, implements the reparameterization trick (Kingma and Welling 2014) to sample multivariate normally distributed vectors, and then feeds the vectors and the subgraphs obtained from ripple walk sampler into the decoder to get a target reconstructed gene expression matrix. In the backward propagation phase, the gradient of each parameter in the model is calculated according to the loss function of PAST and is updated based on the Adam optimizer with default initial learning rate  $lr=0.001$ . The detailed model structure and hyperparameter settings of PAST are summarized in Supplemental Table S7. The prediction workflow only contains the forward propagation of the encoder to acquire the mean matrix of the reparameterization module, that is, the latent embeddings. Furthermore, PAST implements ripple walk prediction strategy instead of the ripple walk training strategy in the prediction workflow.

#### Ripple walk sampler

To extend our model to large-scale application scenarios, we adopt a subgraph-based sampling strategy called ripple walk sampler (Fig. 1C), which samples high-quality subgraphs to enable minibatch training (ripple walk training) under the premise of satisfying randomness and connectivity (Bai et al. 2021). Furthermore, we extend the ripple walk sampler for scalable prediction, called ripple walk prediction. The detailed explanations of ripple walk sampler, training, and prediction are shown in Supplemental Text S11 and Supplemental Algorithms S1, S2, S3.

#### Bayesian neural network

BNN introduces uncertainty into weights of the neural network to improve generalization in nonlinear regression problems (Blundell et al. 2015). Here we use BNN to integrate shared biological information from reference data (Fig. 1D).

Parameters of the BNN module could be represented by  $\mathbf{P} \in d_{in} \times d_{out}$ , and each element in  $\mathbf{P}$  is subject to a Gaussian distribution,

$$\mathbf{P}_{ij} \sim \mathcal{N}(\mu_b^{(ij)}, \sigma_b^{(ij)2}), \quad i = 1, \dots, d_{in}, \quad j = 1, \dots, d_{out},$$

where  $\mu_b^{(ij)}$  and  $\sigma_b^{(ij)2}$  are elements in the  $i$ th row and  $j$ th column of the mean matrix  $\mu_b$  and variance matrix  $\sigma_b^2$ , respectively. We get normally distributed parameters  $\mathbf{P}_{ij}$  through reparameterization trick in each forward propagation process. Given an input matrix  $\mathbf{M} \in B \times d_{in}$  in a minibatch, the output matrix of shape  $B \times d_{out}$  of

BNN module is

$$f_{bnn}(\mathbf{M}) = \mathbf{MP} = \mathbf{M}(\boldsymbol{\mu}_b + \boldsymbol{\delta} \odot \boldsymbol{\sigma}_b),$$

where  $f_{bnn}(\cdot)$  denotes the function of BNN module,  $\boldsymbol{\delta}$  is the standard gaussian noise matrix and  $\odot$  denotes element-wise product.

To integrate shared biological variation information in reference data, we assume elements in  $\mathbf{P}$  are subject to a prior Gaussian distribution parameterized by  $\boldsymbol{\mu}_p$  and  $\boldsymbol{\sigma}_p^2$ , where  $\boldsymbol{\mu}_p$  is the PCA projection weight matrix of the reference gene expression matrix and  $\boldsymbol{\sigma}_p$  contains elements equal to one-tenth standard deviation of  $\boldsymbol{\mu}_p$ . The step-by-step construction of reference gene expression matrix from various sources is illustrated in Supplemental Text S1 and Supplemental Algorithm S4. KLD is used to restrict the distance between the Gaussian distribution parameterized by the parameters of BNN module and the prior Gaussian distribution parameterized by reference data:

$$L_{bnn} = KLD(\mathcal{N}(\boldsymbol{\mu}_b, \boldsymbol{\sigma}_b^2), \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)) \\ = \frac{1}{d_{in}d_{out}} \sum_{i=1}^{d_{in}} \sum_{j=1}^{d_{out}} \left( \log \left( \frac{\boldsymbol{\sigma}_p^{(ij)}}{\boldsymbol{\sigma}_b^{(ij)}} \right) + \frac{\boldsymbol{\sigma}_b^{(ij)2} + (\boldsymbol{\mu}_b^{(ij)} - \boldsymbol{\mu}_p^{(ij)})^2}{2\boldsymbol{\sigma}_p^{(ij)2}} - \frac{1}{2} \right).$$

### Self-attention mechanism

Self-attention mechanism is the key mechanism of transformer for capturing the relationship between words in a given sentence in machine translation tasks, showing superior performance compared with other recurrent neural network-based models (Vaswani et al. 2017; Raffel et al. 2020). Recent advances have revealed that gene expression levels of neighboring spots in a tissue microenvironment tend to be correlated (Dries et al. 2021; Dong and Zhang 2022; Zeng et al. 2022). Inspired by the above understanding, PAST adopts a self-attention mechanism to model the relationship between neighboring spots (Fig. 1E).

The input of the self-attention module is a feature matrix  $\mathbf{M}$  and the corresponding  $k$ -NN graph  $\mathcal{G}$  is constructed using spatial coordinates with  $k=6$ . The input matrix  $\mathbf{M}$  is mapped to obtain query  $\mathbf{Q}$ , key  $\mathbf{K}$ , and value  $\mathbf{V}$  through three FCNs, and the adjacent matrix  $\mathcal{A}$  can then be calculated as follows:

$$\mathcal{A} = \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}},$$

where  $d_k$  is the feature dimension of query  $\mathbf{Q}$  and key  $\mathbf{K}$ . The attention weight matrix  $\mathbf{A}$  can be obtained by normalizing the adjacent matrix  $\mathcal{A}$ ,

$$\mathbf{A}_{ij} = \frac{e^{\mathcal{A}_{ij}}}{\sum_{k \in S_i} e^{\mathcal{A}_{ik}}},$$

where  $\mathbf{A}_{ij}$  is the attention weight from spot  $j$  to spot  $i$ , and  $S_i$  denotes the set of neighbors of spot  $i$  in  $\mathcal{G}$ . To avoid excessive spatial smoothing of hidden features, we add a shortcut in the self-attention module (Zhang and Meng 2019; Liu et al. 2021) and the output of self-attention module can be written as the subsequent form:

$$f_{attn}(\mathbf{M}) = \mathbf{V} + \mathbf{AV},$$

where  $f_{attn}(\cdot)$  denotes the function of self-attention module.

### Loss function

Analysis of omics data is susceptible to sequencing noise such as dropout, and recent studies have proven that it is beneficial to consider data sparsity and dropout in spatial transcriptomic data

(Shengquan et al. 2021). Here we design sparsity adaptive reconstruction loss based on the insight that the higher the data sparsity, the more serious the dropout phenomenon is, and the higher probability the zero gene expression observations are false negative. Given a minibatch containing  $B$  spots and  $M$  genes, the sparsity adaptive reconstruction loss is shown as

$$L_{recons} = \frac{1}{BM} \sum_{i=1}^B \sum_{j=1}^M I(\mathbf{x}_{ij})(\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij})^2,$$

where  $\mathbf{x}_{ij}$  denotes the profiled expression level of gene  $j$  in spot  $i$ ,  $\hat{\mathbf{x}}_{ij}$  denotes the reconstructed expression level of gene  $j$  in spot  $i$ , and the function  $I(\cdot)$  is ( $s_i$  represents the sparsity of spot  $i$ )

$$I(\mathbf{x}_{ij}) = \begin{cases} 1 & \mathbf{x}_{ij} \neq 0 \\ 1 - s_i & \mathbf{x}_{ij} = 0 \end{cases}$$

We assume the low-dimensional embeddings follow multivariate Gaussian distribution and the prior is the centered isotropic multivariate Gaussian distribution. The distance between the latent distribution and the prior distribution is restricted by KLD (Kingma and Welling 2014):

$$L_{kl} = KLD(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2), \mathcal{N}(0, \mathbf{I})) \\ = \frac{1}{2B} \sum_{i=1}^B \sum_{j=1}^{d_{lat}} (\boldsymbol{\mu}_{ij}^2 + \boldsymbol{\sigma}_{ij}^2 - \log(\boldsymbol{\sigma}_{ij}^2) - 1),$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}$  are the mean vectors and standard deviation vectors of latent space, respectively, whereas  $d_{lat}$  denotes the dimension of latent space.

Metric learning has been proposed to model the distance between samples, making the distance between positive sample pairs closer and the distance between negative sample pairs farther (Hadsell et al. 2006). Samples in spatial transcriptomics are of strong spatial autocorrelation, which means that spatial neighboring spots may have similar gene expression patterns and tend to belong to the same functional structure (Zeng et al. 2022). Here we design a metric learning loss function to restrict the distance of low-dimensional representations between neighbor samples:

$$L_{metric} = \sum_{i=1}^B \sum_{j=1}^B \mathbf{G}_{ij} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2,$$

where  $\mathbf{G}$  is the weighted spatial neighbor graph and the connection weights between samples are calculated using the Student's  $t$ -distribution kernel,  $\boldsymbol{\mu}_i$  and  $\boldsymbol{\mu}_j$  denote the latent embeddings of spot  $i$  and spot  $j$ , respectively. The calculation of  $\mathbf{G}$  can be denoted as

$$\mathbf{G}_{ij} = \frac{(1 + \|\mathbf{c}_i - \mathbf{c}_j\|_2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j \in S_i} (1 + \|\mathbf{c}_i - \mathbf{c}_j\|_2/\alpha)^{-\frac{\alpha+1}{2}}},$$

where  $\mathbf{c}_i$  denotes the spatial coordinate of spot  $i$ ,  $S_i$  denotes the set of neighbors of spot  $i$ ,  $\alpha$  is a scale factor that is set to 1 by default. We also tried to build the graph  $\mathbf{G}$  with gene expression information of spots, that is, denoting the above  $\mathbf{c}_i$  as gene expression vector of spot  $i$ , and compared the two different approaches (Supplemental Text S12; Supplemental Fig. S33), illustrating the effectiveness of building graph  $\mathbf{G}$  with spatial information.

As described in the above section, we integrate prior information through the BNN (Blundell et al. 2015) module and restrict the distance between parameter of the BNN module and prior distribution through  $L_{bnn}$  loss function. The objective function of

PAST is the combination of the aforementioned four parts:

$$L = \beta_1 L_{recons} + \beta_2 L_{kld} + \beta_3 L_{metric} + \beta_4 L_{bmi},$$

where  $\beta_1, \beta_2, \beta_3, \beta_4$  are coefficients served as a trade-off between the four parts and we set  $\beta_1 = 1, \beta_2 = 1, \beta_3 = 1, \beta_4 = 1$  by default. We have evaluated the impact of the four coefficients and the contribution of the four loss function parts through experimental analysis, which finally illustrates the robustness of PAST to different hyperparameter settings and the significance of each part of the loss function (Supplemental Text S13; Supplemental Figs. S34, S35).

### Multislice joint embedding and automatic annotation

We propose two strategies, namely a transfer learning strategy and a 3D stacking strategy, to apply representation learning methods designed for single slice to multiple slices for joint embedding. Furthermore, we annotate the domain labels of multiple slices in ST data based on the joint embeddings obtained by the aforementioned strategies. The detailed explanations of the joint embedding and automatic annotation strategies are shown in Supplemental Text S4.

### Baseline methods, downstream analyses, and evaluation metrics

Downstream applications including visualization, clustering, trajectory inference, and pseudotime analysis are detailed in Supplemental Text S3. To evaluate the performance of our model, we benchmarked PAST with nine baseline methods in term of 16 quantitative metrics and various downstream analyses. We describe the baseline methods, parameter settings, and the evaluation metrics in Supplemental Text S2. The details about data collection and preprocessing are shown in Supplemental Text S9. The information about all the target data sets and reference data sets is summarized in Supplemental Table S8.

### Data sets

The STARmap mouse primary visual cortex (MPVC) data set (Wang et al. 2018) is available at <https://kangaroo-goby.squarespace.com/data>. The 10x Visium dorsolateral prefrontal cortex (DLPFC) data set (Maynard et al. 2021) includes 12 slices from 3 neurotypical adult donors and is accessible through <http://spatial.libd.org/spatialLIBD>. The osmFISH mouse somatosensory cortex (MSC) data set (Codeluppi et al. 2018) is available at website <http://linnarssonlab.org/osmFISH/availability/>. The Stereo-seq mouse olfactory bulb (Chen et al. 2022a) Sections 1 (MOBS1) and Section 2 (MOBS2) are accessible at <https://db.cngb.org/stomics/mosta/download/>, where MOBS2 was taken as the reference data for MOBS1. The 10x Visium human breast cancer Section 1 (BCS1) was precisely annotated by Fu et al. (2021), whereas Section 2 (BCS2) is not manually annotated, available at the 10x Genomics website <https://www.10xgenomics.com/resources/datasets/human-breast-cancer-block-a-section-1-1-standard-1-1-0> and <https://www.10xgenomics.com/resources/datasets/human-breast-cancer-block-a-section-2-1-standard-1-1-0>, respectively. The three Stereo-seq axolotl brain data sets (Wei et al. 2022) identified by 2DPI are available through the website <https://db.cngb.org/stomics/artista/download/>. The Stereo-seq zebrafish embryo data set (Liu et al. 2022a) contains 12 slices of 12 h post-fertilization zebrafish embryo and can be downloaded from <https://db.cngb.org/stomics/zesta/download/>. The 10x Visium mouse brain coronal data set is accessible through [https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1\\_Adult\\_Mouse\\_Brain](https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Adult_Mouse_Brain). The Stereo-seq mouse brain data set (Chen et al. 2022a) can be downloaded from <https://db.cngb.org/stomics/mosta/download/>. The scRNA-seq mouse primary visual cortex data set was obtained from the NCBI

Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE71585.

### Software availability

The PAST algorithm is implemented in Python and the source code for reproduction is available at GitHub (<https://github.com/lizhen18THU/PAST>) and Supplemental Code. We also provided detailed documentation and step-by-step tutorials for applying PAST to ST data generated by different technologies at Read the Docs website (<https://past.readthedocs.io/en/latest>).

### Competing interest statement

The authors declare no competing interests.

### Acknowledgments

This work was supported by the National Key Research and Development Program of China Grant No. 2021YFF1200902 (R.J.), the National Natural Science Foundation of China Grants No. 62203236 (S.C.), No. 62273194 (R.J.), No. 61721003 (X.Z.), and the Fundamental Research Funds for the Central Universities, Nankai University, Grant No. 63231137 (S.C.).

*Author contributions:* S.C. conceived the study and supervised the project. Z.L. and S.C. designed, implemented, and validated PAST. X.C. and X.Z. helped with analyzing the results. S.C., Z.L., X.Z., and R.J. wrote the manuscript, with input from all the authors.

### References

- Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, Reinders MJ, Mahfouz A. 2019. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol* **20**: 194. doi:10.1186/s13059-019-1795-z
- Agirman G, Broix L, Nguyen L. 2017. Cerebral cortex development: An outside-in perspective. *FEBS Lett* **591**: 3978–3992. doi:10.1002/1873-3468.12924
- Armingol E, Officer A, Harismendy O, Lewis NE. 2021. Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet* **22**: 71–88. doi:10.1038/s41576-020-00292-x
- Asp M, Bergenstråhle J, Lundeberg J. 2020. Spatially resolved transcriptomes—next generation tools for tissue exploration. *Bioessays* **42**: 1900221. doi:10.1002/bies.201900221
- Bai J, Ren Y, Zhang J. 2021. Ripple walk training: a subgraph-based training framework for large and deep graph neural network. In *International Joint Conference on Neural Networks*. IEEE, Shenzhen.
- Bitar M, Weissleder C, North HF, Clearwater MS, Zalucki O, Halliday GM, Webster MJ, Piper M, Weickert CS, Barry G. 2022. Identifying gene expression profiles associated with neurogenesis and inflammation in the human subependymal zone from development through aging. *Sci Rep* **12**: 40. doi:10.1038/s41598-021-03976-4
- Blundell C, Cornebise J, Kavukcuoglu K, Wierstra D. 2015. Weight uncertainty in neural network. In *International Conference on Machine Learning*. PMLR, Lille.
- Chang Y-T, Tsai W-C, Lin W-Z, Wu C-C, Yu J-C, Tseng VS, Liao G-S, Hu J-M, Hsu H-M, Chang Y-J. 2021. A novel IGLC2 gene linked with prognosis of triple-negative breast cancer. *Front Oncol* **11**: 759952. doi:10.3389/fonc.2021.759952
- Chen S, Yan G, Zhang W, Li J, Jiang R, Lin Z. 2021. RA3 is a reference-guided approach for epigenetic characterization of single cells. *Nat Commun* **12**: 2177. doi:10.1038/s41467-021-22495-4
- Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S, et al. 2022a. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**: 1777–1792.e21. doi:10.1016/j.cell.2022.04.003
- Chen X, Chen S, Song S, Gao Z, Hou L, Zhang X, Lv H, Jiang R. 2022b. Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding. *Nat Mach Intell* **4**: 116–126. doi:10.1038/s42256-021-00432-w

- Cheng A, Hu G, Li WV. 2022. Benchmarking cell-type clustering methods for spatially resolved transcriptomics data. *Brief Bioinform* **24**: bbac475. doi:10.1093/bib/bbac475
- Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, Svensson CI, Linnarsson S. 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* **15**: 932–935. doi:10.1038/s41592-018-0175-z
- Crosetto N, Bienko M, Van Oudenaarden A. 2015. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16**: 57–66. doi:10.1038/nrg3832
- Deng Y, Bartosovic M, Ma S, Zhang D, Kukanja P, Xiao Y, Su G, Liu Y, Qin X, Rosoklija GB, et al. 2022. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**: 375–383. doi:10.1038/s41586-022-05094-1
- Dong K, Zhang S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention autoencoder. *Nat Commun* **13**: 1739. doi:10.1038/s41467-022-29439-6
- Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan G-C. 2021. Advances in spatial transcriptomic data analysis. *Genome Res* **31**: 1706–1718. doi:10.1101/gr.275224.121
- Fan Z, Chen R, Chen X. 2020. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res* **48**: D233–D237. doi:10.1093/nar/gkz934
- Friess C, Gotz M, Kjell J. 2021. Cryo-section dissection of the adult subependymal zone for accurate and deep quantitative proteome analysis. *J Vis Exp* **176**: e63047. doi:10.3791/63047
- Fu H, Xu H, Chong K, Li M, Ang KS, Lee HK, Ling J, Chen A, Shao L, Liu L. 2021. Unsupervised spatially embedded deep representation of spatial transcriptomics. *bioRxiv* doi:10.1101/2021.06.15.448542
- Gao Z, Chen X, Li Z, Cui X, Jiang Q, Li K, Chen S, Jiang R. 2023. scEpiTools: a database to comprehensively interrogate analytic tools for single-cell epigenomic data. *J Genet Genomics* doi:10.1016/j.jgg.2023.09.011
- Gilmore EC, Herrup K. 1997. Cortical development: Layers of complexity. *Curr Biol* **7**: R231–R234. doi:10.1016/S0960-9822(06)00108-4
- Hadsell R, Chopra S, LeCun Y. 2006. Dimensionality reduction by learning an invariant mapping. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, New York.
- Haghverdi L, Büttner M, Wolf FA, Büttner F, Theis FJ. 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**: 845–848. doi:10.1038/nmeth.3971
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM III, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* **184**: 3573–3587.e29. doi:10.1016/j.cell.2021.04.048
- Haque A, Engel J, Teichmann SA, Lönnberg T. 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* **9**: 75. doi:10.1186/s13073-017-0467-4
- Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, Li M. 2021a. 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. doi:10.1038/s41592-021-01255-8
- Hu J, Schroeder A, Coleman K, Chen C, Auerbach BJ, Li M. 2021b. Statistical and machine learning methods for spatially resolved transcriptomics with histology. *Comput Struct Biotechnol J* **19**: 3829–3841. doi:10.1016/j.csbj.2021.06.052
- Jiang R, Li Z, Jia Y, Li S, Chen S. 2023. SINFONIA: Scalable identification of spatially variable genes for deciphering spatial domains. *Cells* **12**: 604. doi:10.3390/cells12040604
- Kazanis I. 2009. The subependymal zone neurogenic niche: a beating heart in the centre of the brain: How plastic is adult neurogenesis? Opportunities for therapy and questions to be addressed. *Brain* **132**: 2909–2921. doi:10.1093/brain/awp237
- Kingma DP, Welling M. 2014. Auto-encoding variational Bayes. In *International Conference on Learning Representations, ICLR, Banff, Canada*.
- Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS, Pokholok D, Aryee MJ, Steemers FJ, Lebofsky R, et al. 2019. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* **37**: 916–924. doi:10.1038/s41587-019-0147-6
- Larsson L, Frisén J, Lundeberg J. 2021. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat Methods* **18**: 15–18. doi:10.1038/s41592-020-01038-7
- Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJJ, Kong SL, Chua C, Hon LK, Tan WS, et al. 2017. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* **49**: 708–718. doi:10.1038/ng.3818
- Li J, Chen S, Pan X, Yuan Y, Shen H-B. 2022. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat Comput Sci* **2**: 399–408. doi:10.1038/s43588-022-00266-5
- Liu A, Xu N, Liu H. 2021. Self-attention graph residual convolutional networks for event detection with dependency relations. In *Findings of the Association for Computational Linguistics*. Association for Computational Linguistics, Punta Cana, Dominican Republic.
- Liu C, Li R, Li Y, Lin X, Zhao K, Liu Q, Wang S, Yang X, Shi X, Ma Y, et al. 2022a. Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev Cell* **57**: 1284–1298.e5. doi:10.1016/j.devcel.2022.04.009
- Liu W, Liao X, Yang Y, Lin H, Yeong J, Zhou X, Shi X, Liu J. 2022b. Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data. *Nucleic Acids Res* **50**: e72. doi:10.1093/nar/gkac219
- Liu Z, Sun D, Wang C. 2022c. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol* **23**: 218. doi:10.1186/s13059-021-02568-9
- Maynard KR, Collado-Torres L, Weber LM, Uyttingco C, Barry BK, Williams SR, Cattallini JL, Tran MN, Besich Z, Tippani M, et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**: 425–436. doi:10.1038/s41593-020-00787-0
- Menn B, Garcia-Verdugo JM, Yaschine C, Gonzalez-Perez O, Rowitch D, Alvarez-Buylla A. 2006. Origin of oligodendrocytes in the subventricular zone of the adult brain. *J Neurosci* **26**: 7907–7918. doi:10.1523/JNEUROSCI.1299-06.2006
- Moses L, Pachter L. 2022. Museum of spatial transcriptomics. *Nat Methods* **19**: 534–546. doi:10.1038/s41592-022-01409-2
- Oxelius V-A, Pandey JP. 2013. Human immunoglobulin constant heavy G chain (IGHG) (Fcγ) (GM) genes, defining innate variants of IgG molecules and B cells, have impact on disease and therapy. *Clin Immunol* **149**: 475–486. doi:10.1016/j.clim.2013.10.003
- Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, Rybakov S, Ibarra IL, Holmberg O, Virshup I, et al. 2022. Slidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**: 171–178. doi:10.1038/s41592-021-01358-2
- Qiu P. 2020. Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun* **11**: 1169. doi:10.1038/s41467-020-14976-9
- Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res* **21**: 1–67.
- Rodrigues SG, Stickle RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. 2019. Slidpy: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**: 1463–1467. doi:10.1126/science.aaw1219
- Schmidt M, Mücke P, Gehrman M, Hengstler JG. 2012. Immunoglobulin kappa chain as an immunologic biomarker of prognosis and chemotherapy response in solid tumors. *Oncimmunology* **1**: 1156–1158. doi:10.4161/onci.21653
- Shang L, Zhou X. 2022. Spatially aware dimension reduction for spatial transcriptomics. *Nat Commun* **13**: 7203. doi:10.1038/s41467-022-34879-1
- Shengquan C, Boheng Z, Xiaoyang C, Xuegong Z, Rui J. 2021. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics* **37**: i299–i307. doi:10.1093/bioinformatics/btab298
- Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**: 78–82. doi:10.1126/science.aaf2403
- Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, Arlotta P, Macosko EZ, Chen F. 2021. Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqV2. *Nat Biotechnol* **39**: 313–319. doi:10.1038/s41587-020-0739-1
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, et al. 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**: 335–346. doi:10.1038/nn.4216
- Townes FW, Engelhardt BE. 2023. Nonnegative spatial factorization applied to spatial genomics. *Nat Methods* **20**: 229–238. doi:10.1038/s41592-022-01687-w
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. In *Conference on Neural Information Processing Systems*. Curran Associates, Long Beach, CA.
- Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, Åijö T, Bonneau R, Bergenstråhle L, Navarro JF, et al. 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* **16**: 987–990. doi:10.1038/s41592-019-0548-y
- Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**: eaat5691. doi:10.1126/science.aat5691
- Wei X, Fu S, Li H, Liu Y, Wang S, Feng W, Yang Y, Liu X, Zeng Y-Y, Cheng M, et al. 2022. Single-cell stereo-seq reveals induced progenitor cells

- involved in axolotl brain regeneration. *Science* **377**: eabp9444. doi:10.1126/science.abp9444
- Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**: 15. doi:10.1186/s13059-017-1382-0
- Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**: 59. doi:10.1186/s13059-019-1663-x
- Xu Z, Wang W, Yang T, Chen J, Huang Y, Gould J, Du W, Yang F, Li L, Lai T, et al. 2022. STOmicsDB: a database of spatial transcriptomic data. bioRxiv doi:10.1101/2022.03.11.481421
- Yuan Z, Pan W, Zhao X, Zhao F, Xu Z, Li X, Zhao Y, Zhang MQ, Yao J. 2023. SODB facilitates comprehensive exploration of spatial omics data. *Nat Methods* **20**: 387–399. doi:10.1038/s41592-023-01773-7
- Zeira R, Land M, Strzalkowski A, Raphael BJ. 2022. Alignment and integration of spatial transcriptomics data. *Nat Methods* **19**: 567–575. doi:10.1038/s41592-022-01459-6
- Zeng Z, Li Y, Li Y, Luo Y. 2022. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol* **23**: 83. doi:10.1186/s13059-022-02653-7
- Zhang J, Meng L. 2019. GResNet: graph residual network for reviving deep GNNs from suspended animation. arXiv:1909.05729 [cs.LG].
- Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uyttingco CR, Taylor SE, Nghiem P, et al. 2021. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* **39**: 1375–1384. doi:10.1038/s41587-021-00935-2
- Zhuang X. 2021. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat Methods* **18**: 18–22. doi:10.1038/s41592-020-01037-8

Received March 16, 2023; accepted in revised form September 19, 2023.