RGAST: Relational Graph Attention Network for Spatial Transcriptome Analysis

Yuqiao Gong[1] and Zhangsheng Yu[1,2,3,4,*]

[1]Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, 200240, Shanghai, China

[2]SJTU-Yale Joint Center for Biostatistics and Data Science Organization, Shanghai Jiao Tong University, Shanghai, China

[3]Clinical Research Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Center for Biomedical Data Science, Translational Science Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]Corresponding author. yuzhangsheng@sjtu.edu.cn

Abstract

Recent advancements in spatially resolved transcriptomics have provided a powerful means to comprehensively capture gene expression patterns while preserving the spatial context of the tissue microenvironment. Accurately deciphering the spatial context of spots within a tissue necessitates the careful utilization of their spatial information, which in turn requires feature extraction from complex and detailed spatial patterns. In this study, we present RGAST (Relational Graph Attention network for Spatial Transcriptome analysis), a framework designed to learn low-dimensional representations of spatial transcriptome (ST) data. RGAST is the first to consider gene expression similarity and spatial neighbor relationships simultaneously in constructing a heterogeneous graph network in ST analysis. We further introduce a cross-attention mechanism to provide a more comprehensive and adaptive representation of spatial transcriptome data. We validate the effectiveness of RGAST in different downstream tasks using diverse spatial transcriptomics datasets obtained from different platforms with varying spatial resolutions. Our results demonstrate that RGAST enhances spatial domain identification accuracy by approximately 10% compared to the second method in 10X Visium DLPFC dataset. Furthermore, RGAST facilitates the discovery of spatially variable genes, uncovers spatially resolved cell-cell interactions, enables more precise cell trajectory inference and reveals intricate 3D spatial patterns across multiple sections of ST data. Our RGAST method is available as a Python package on PyPI at https://pypi.org/project/RGAST, free for academic use, and the source code is openly available from our GitHub repository at https://github.com/GYQ-form/RGAST.

Keywords: spatial transcriptomic, relational graph attention network, representation learning, spatial domain identification, RGAST

Introduction

Spatial transcriptomics is a breakthrough technology that allows researchers to investigate gene expression and tissue architecture simultaneously in a spatially resolved manner[1]. Technologies such as 10x Visium[2], Slide-seq[3], Stereo-seq[4], MERFISH[5], and seqFISH+[6] enable genome-wide profiling of gene expression in captured locations, or "spots," at the level of several cells or even subcellular structures. These technique provide crucial insights into the interactions between cells, tissues, and organs, and the underlying molecular mechanisms that govern cellular behavior[5]. Spatial transcriptomics has the potential to shed new light on various biological processes, including development, disease progression, and drug response. However, analyzing spatial transcriptome data remains a significant challenge due to its high dimensionality, noise, and complexity[7]. To understand the spatial patterns and relationships among different cells or spots, sophisticated computational methods are required[8].

Traditional pipeline methods like Seurat[9] and Scanpy[10], which primarily focus on gene expression data and ignore the structural relationship of the spatial neighborhood, have been shown to fail in spatial transcriptomics (ST) data analysis[11, 12]. To address this issue, several new computational methods have been developed that utilize the similarity between adjacent spots to better account for the spatial dependency of gene expressions. BayesSpace[13] and Giotto[14] are two such methods that use Markov random field (MRF) and hidden Markov random field (HMRF) models, respectively, to detect cell types and spatial domains based on gene expression patterns. However, these techniques using limited statistical methods are mainly used in ST data with lower resolution and unable to learn a universal embedding for downstream tasks. Recently, there are also some analytical tools that can generate low-dimensional embeddings. SpaGCN constructs a weighted graph of spots using spatial distances and histological dissimilarities and then integrates the graph with gene expression using a graph convolutional network (GCN) to cluster the spots[15]. stLearn defines morphological distance based on features extracted from a histology image and uses it along with spatial neighbor structure to smooth gene expressions[16]. SEDR employs a deep auto-encoder network to learn gene representations and uses a variational graph auto-encoder to embed spatial information[11]. conST uses a multi-modal, interpretable contrastive learning framework to learn low-dimensional embeddings of ST data for downstream analyses[17]. Although these methods consider the spatial structure of STs, they treat the spatial relationships in an explicit or predefined manner, which may not fully capture the complex and detailed spatial patterns in the data. A recent method called STAGATE is a graph attention autoencoder model that clusters spots/cells in ST data and detects spatial domains while allowing for adaptive updates to the spatial neighbor network (SNN) through a graph attention mechanism[12]. However, to consider the spatial similarity of spots at the boundary of spatial domains in more detail, the method prunes the network with pre-clustered gene expression, limiting its flexibility.

In this study, we introduce RGAST (Relational Graph Attention network for Spatial Transcriptome analysis), which constructs a relational graph attention network to learn the representation of each spot in the ST data. RGAST considers both gene expression similarity and spatial neighbor relationships to construct a heterogeneous graph network. The representation of each spot is adaptively updated using information from both spatial neighbors and spots with similar expression profiles through the cross-attention mechanism, resulting in a much more comprehensive and adaptive representation of the ST data. To demonstrate the superiority of RGAST's representation, we conduct extensive tests and comparisons with several existing methods on spatial transcriptome (ST) data generated by different platforms across multiple tasks. The results show that RGAST performs much better in all downstream analyses. Furthermore, we extend RGAST to analyze consecutive slices of spatial transcriptome data, allowing for the construction of a 3D model that reveals the 3D spatial patterns.

Results

Overview of the RGAST model

The spatial distribution of transcriptional expressions in tissues is crucial for understanding their biological functions and describing interactive biological networks[18]. Spatial transcriptomics (ST) data provides valuable location information that can be effectively utilized. Interestingly, certain spots within the tissue, despite being physically separated, may exhibit similar expression patterns due to their involvement in spatially regulated processes such as diffusion or signaling[1]. This implies that certain spots can share similar expression patterns even when located far apart. In the RGAST model, we consider two distinct relations to construct a heterogeneous graph (Methods). The first relationship is based on the similarity of gene expression between different spots. The second relationship is determined by the spatial neighborhood, which considers the relative spatial locations of the spots (Fig. 1). Taking both relationships into consideration, we further incorporate an attention mechanism across different relations that adaptively learns the edge weights between different spots, effectively determining the importance of each spot with respect to a target spot. These learned edge weights are then utilized by the target spot to update its representation by aggregating information from spots linked to it.

Using the relational graph attention auto-encoder, RGAST enables the learning of low-dimensional latent embeddings that incorporate both spatial information and gene expressions (Fig. 1). Initially, the expression profiles of each spot undergo dimensionality reduction using PCA. Subsequently, an encoder transforms the reduced expression into a d-dimensional latent embedding, which is then reconstructed back into an expression profile using a linear decoder (Methods). Additionally, an unsupervised deep clustering method can be optionally employed to enhance the compactness of the learned latent representation[19] (Methods). The resulting latent embeddings can be utilized to enhance the performance of various downstream analyses,

including data visualization, identification of spatial domains, trajectory inference, SVG detection, cell-cell interaction analysis, and extraction of 3D spatial domains.

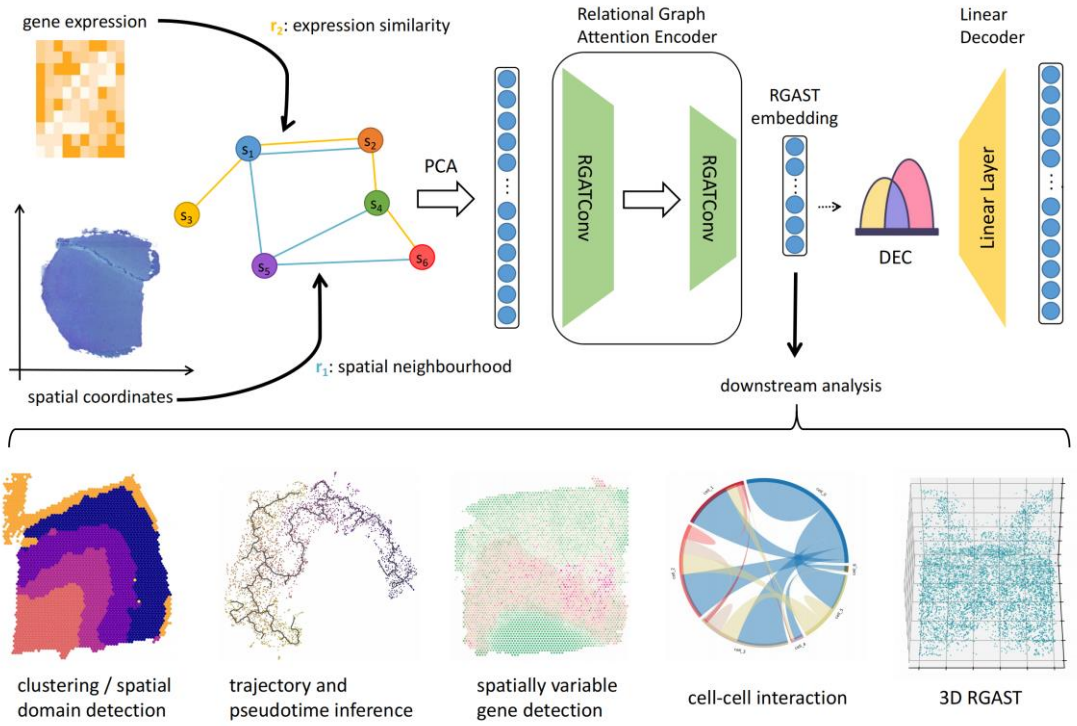

Fig. 1 | Overview of the RGAST model. RGAST learns low-dimensional latent embeddings with both spatial information and gene expressions. The expression after dimensionality reduction by PCA of each spot is first transformed into a d-dimensional latent embedding by an encoder and then reversed back into a reconstructed expression profile via a linear decoder. An optional unsupervised deep clustering method is then employed to enhance the compactness of the learned latent representation. The latent embeddings can be used to visualize the data and be applied to many downstream analysis like identifying spatial domains, trajectory inference, SVG detection, cell-cell interaction analysis and extracting 3D spatial domains.

RGAST improves spatial domain detection in various ST technologies with scalability

We assessed the spatial clustering performance of RGAST by applying it to ST data generated through various technologies. We first utilized the 10x Visium dataset, which included spatial expression profiles of 12 human dorsolateral prefrontal cortex (DLPFC) sections[20]. Maynard et al. manually annotated the DLPFC layers and white matter (WM) based on morphological features and gene markers, which we adopted as the ground truth. To compare the clustering accuracy of RGAST with other methods, we employed the adjusted Rand index (ARI) (Methods) and compared it with the non-spatial clustering method Seurat and five recently developed spatial clustering approaches: SpaGCN[15], BayesSpace[13], SEDR[11], STAGATE[12], and conST[17]. To ensure a fair comparison, we set the maximum number of clusters for each method the same as the ground truth label and used the recommended default parameters for all other methods. For methods that generated latent embedding (i.e., SEDR, STAGATE,

conST, and RGAST), we utilized the Leiden clustering algorithm (Methods). The results demonstrated that RGAST effectively identified the expected cortical layer structures and achieved the best performance in almost all slices (Fig. 2b and Supplementary Fig. S1-S11). For instance, in DLPFC section 151675, RGAST clearly delineated the layer borders and achieved the best clustering accuracy (ARI = 0.633) (bottom right of Fig. 1a). In contrast, the clustering assignment of the non-spatial method Seurat was discontinuous, with many outliers that impeded its clustering accuracy. Not surprisingly, algorithms that leveraged spatial information substantially outperformed the non-spatial clustering method Seurat (Fig. 2b). These findings underscore the superiority of RGAST in spatial domain identification and the necessity to incorporate spatial information in spatial clustering analysis.

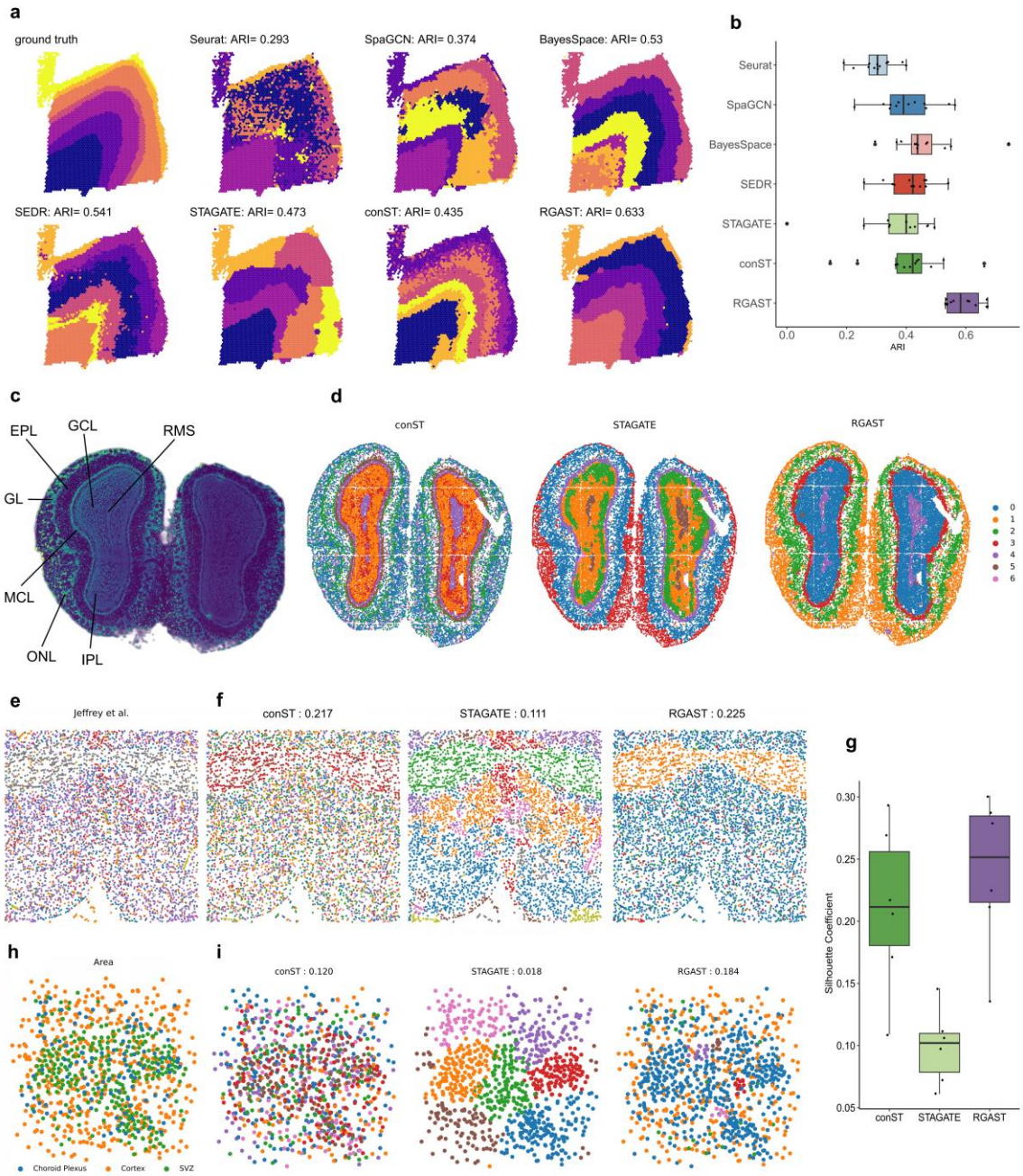

Fig. 2 | RGAST improves spatial domain detection in various ST technologies. a, Clustering results of different methods in the 10X Visium DLPFC section 151675 dataset. **b,** Boxplot of clustering accuracy for seven methods across all 12 sections of the 10X Visium DLPFC datasets, measured in terms of ARI scores. In the boxplot, the center line, box limits, and whiskers represent the median, upper and lower quartiles, and 1.5× interquartile range, respectively. These parameters are consistent across all boxplots in this paper. **c,** Laminar organization of the mouse olfactory bulb annotated in the DAPI-stained image generated by Stereo-seq. **d,** Spatial domains generated by Leiden clustering with a maximum of seven clusters on the low-dimensional conST, STAGATE, and RGAST embeddings in the Stereo-seq mouse olfactory bulb tissue section. **e,** Cell annotation by Jeffrey et al. based on scRNA-seq profiling. **f,** Spatial domains generated by different methods in the mouse hypothalamic preoptic MERFISH data ofsection Bergman_0.26. The number in the title represents the SC score of the different methods, which is consistent with i. **g,** Boxplot of clustering accuracy for different methods across all six sections of the MERFISH mouse hypothalamic preoptic datasets, measured in terms of SC scores. **h,** Cell annotation by area, including the Choroid Plexus, subventricular zone (SVZ), and cortex area in the mouse cortex SeqFISH+ dataset. **i,** Spatial domains generated by different methods in the mouse cortex SeqFISH+ dataset.

Next, we compared the performance of RGAST with conST and STAGATE, using the Stereo-seq Mouse olfactory bulb dataset. We selected conST and STAGATE for comparison since these two methods have been demonstrated to be superior to other techniques[12, 17]. This emerging spatial omics technology can achieve subcellular spatial resolution using DNA nanoball patterned array chips, and the data used here were binned into a cellular-level resolution of approximately 14 μm[4]. Fu et al. annotated the laminar organization of the coronal mouse olfactory bulb in the DAPI-stained image[11], which included the rostral migratory stream (RMS), granule cell layer (GCL), internal plexiform layer (IPL), mitral cell layer (MCL), external plexiform layer (EPL), and olfactory nerve layer (ONL) (Fig. 2c). Since this dataset contains approximately 20,000 spots, we employed a customized batch technique to prevent memory exhaustion during training, thereby enhancing the scalability of RGAST (Methods). Compared to the clusters identified by conST, those identified using both STAGATE and RGAST embeddings more accurately reflected the laminar organization and corresponded well to the annotated layers. However, STAGATE erroneously separated the GCL into two clusters, whereas RGAST did not (Fig. 2d). Furthermore, the five main layer structures identified by RGAST were all validated by the corresponding marker genes (Supplementary Fig. S12).

We also conducted experiments using Mouse hypothalamic preoptic MERFISH and Mouse cortex seqFISH+ data, which were captured at a higher resolution to achieve single-cell resolution, but no morphology data was available. Since no gold standard was available, we used Silhouette Coefficient (SC) to evaluate the clustering performance of each method. For the MERFISH datasets, Jeffrey et al. annotated the cell types based on scRNA-seq expression data[21] (Fig. 2e). The results showed that RGAST consistently outperformed the other methods in all six slices, with STAGATE far behind (Fig. 2f,g and Supplementary Fig. S13-S17). For example, in slice

Bergman_0.26, RGAST was more consistent with the clustering results generated by the expression profile (Fig. 2f). In the Mouse cortex seqFISH+ dataset, the spots were annotated as the Choroid Plexus, subventricular zone (SVZ), and cortex area [6] (Fig. 2h). RGAST still demonstrated better performance than the other two methods (Fig. 2i). It's noteworthy that STAGATE appeared to overuse the spatial neighborhood relationship, as it showed obvious clustering characteristics of block aggregation in such data with discrete cell type distribution (Fig. 2f,i).

RGAST assist identifying spatially variable genes

Spatially variable genes (SVGs) exhibit varying expression patterns across distinct spatial locations. Identifying SVGs can provide valuable insights into the biological processes that underlie the development and function of tissues and organs. To this end, we incorporated SpaGCN, a tool designed specifically for detecting SVGs, into our benchmarking alongside conST and STAGATE. We compared the performance of these methods to that of RGAST using datasets generated by different ST techniques. In order to ensure a fair and accurate comparison, we utilized identical detection methods and default parameters as those employed by SpaGCN (see Methods section).

We first utilized the section 151674 of DLPFC dataset from 10X Visium data as our benchmarking dataset. Our results demonstrated that SVGs exclusively detected by RGAST exhibited a clear spatial distribution pattern, whereas the spatial distribution patterns of SVGs exclusively detected by other three methods are less discernable (Fig. 3a). We employed Moran's I and Geary's C statistics to quantify the degree of spatial autocorrelation of gene expression. A higher Moran's I statistic and a lower Geary's C statistic indicate a more distinct spatial pattern. We can see the SVGs identified by RGAST are significantly more reliable than those identified by the other methods (Fig. 3b).

Due to the absence of pathological images in the MERFISH and SeqFISH+ datasets, SpaGCN was not included in our follow-up comparison. Interestingly, STAGATE failed to detect any SVGs in the mouse cortex SeqFISH+ dataset, also indicating an unclear spatial domain pattern in this dataset (Fig. 3c). To address this issue, we visualized the SVGs detected by RGAST and conST with the highest Moran's I, which revealed the superior performance of RGAST (Fig. 3c). In fact, RGAST outperformed conST even when we considered all the SVGs detected by each method (Fig. 3d). For the MERFISH data, we used slice Berman_0.26 as our benchmarking dataset and found that SVGs identified by RGAST displayed a clearer spatial distribution pattern, even when the target domain was similar (Fig. 3e). This was confirmed by our quantitative comparison as well (Fig. 3f). Lastly, to evaluate the representational ability of SVGs detected by each method, we used the SVGs as features to conduct clustering on the DLPFC data, with ARI as our evaluation criterion. For a fair comparison, we randomly selected four SVGs for each method, since STAGATE only detects four SVGs. Our results showed that RGAST and STAGATE performed similarly and significantly better than SpaGCN and conST (Fig. 3g).

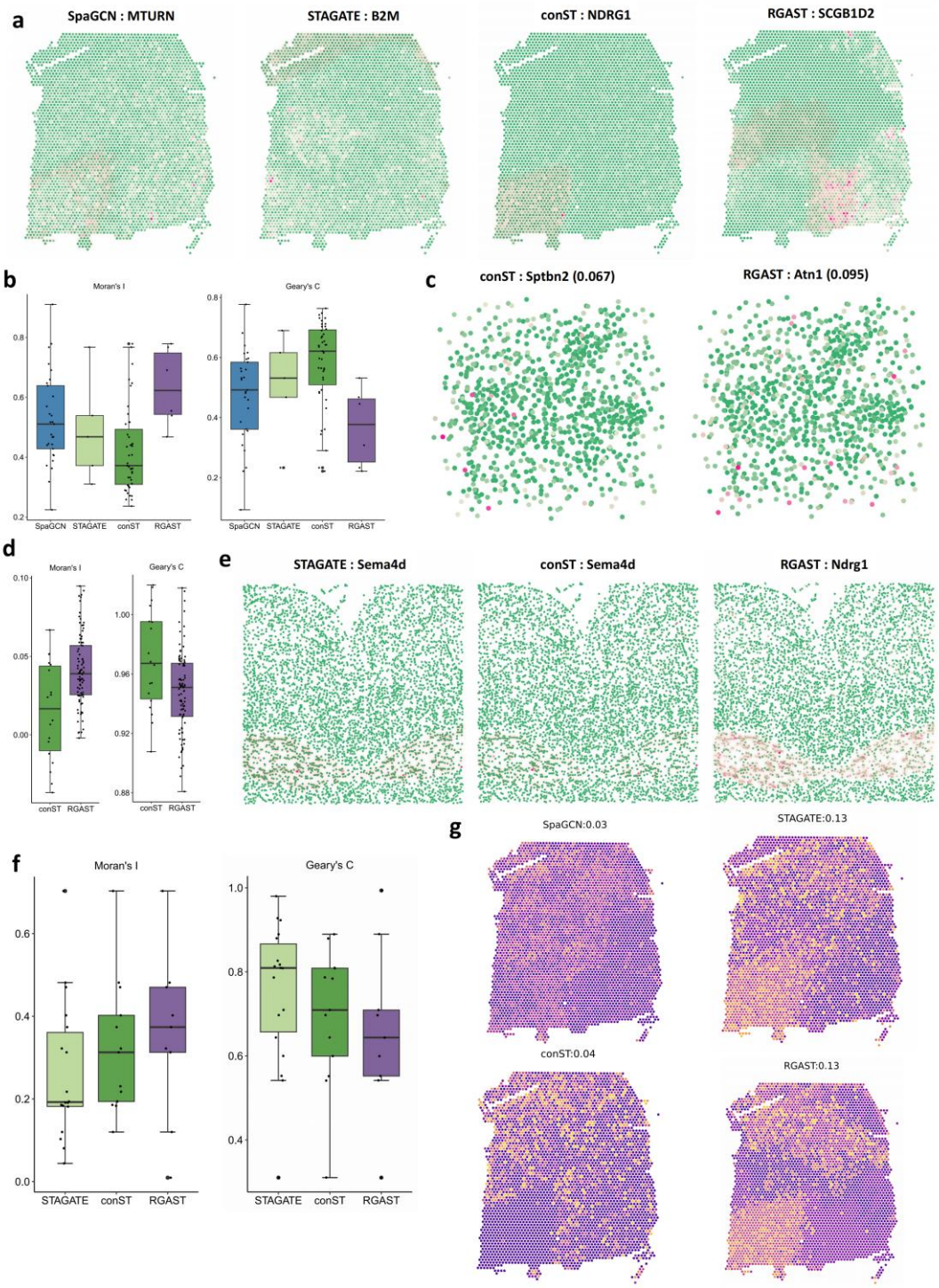

Fig. 3 | RGAST assist identifying spatially variable genes. a, Examples of the spatial distribution of SVGs exclusively detected by RGAST, SpaGCN, STAGATE, and conST, respectively, in the DLPFC dataset section 151674. Shades of red indicate the target domain used to detect the corresponding SVGs. **b,** Boxplot of Moran's I and Geary's C statistics for all SVGs detected by each method in the DLPFC dataset section 151674. **c,** Spatial distribution of SVGs detected by RGAST and conST with the highest Moran's I in mouse cortex SeqFISH+ data. STAGATE failed to detect any SVGs. **d,** Boxplot of Moran's I and Geary's C statistics for all SVGs detected by each method in mouse cortex SeqFISH+ data. **e,** Spatial distribution of SVGs detected by RGAST, STAGATE, and conST, respectively, in mouse hypothalamic preoptic MERFISH

data slice Berman_0.26 with a similar target domain. **f,** Boxplot of Moran's I and Geary's C statistics for all SVGs detected by each method in mouse hypothalamic preoptic MERFISH data slice Berman_0.26. **g,** Clustering results using only four SVGs detected by each method as features. The number in the title denotes the ARI score.

RGAST uncovers spatially resolved cell-cell interactions

Cell-cell interaction analysis plays a crucial role in unraveling the intricate mechanisms underlying biological processes[22]. The significance of studying cell-cell interactions has been further amplified with the emergence of spatial transcriptomic techniques, which provide spatially resolved gene expression data[23]. These techniques, for instance, enable the capture of gene expression patterns of individual cells in their native tissue context. By delving into cell-cell interactions, valuable insights can be gained regarding the communication and mutual influence between different cell types within specific tissue regions. This spatial information assumes vital importance in comprehending the intricacies of organization and functionality within complex tissues and organs.

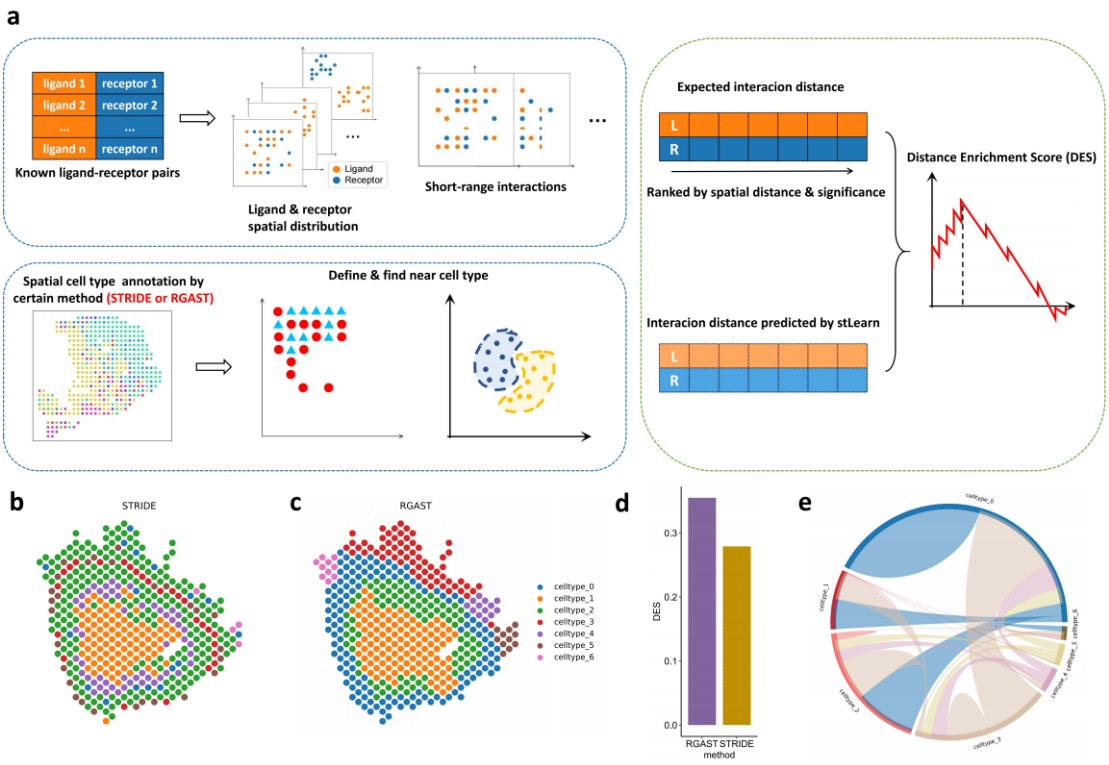

Fig. 4 | RGAST uncovers spatially resolved cell-cell interactions. a, Schematic overview of evaluation workflow. First, generate known ligand-receptor pairs from CellChatDB, then select short-range interactions from known pairs for the dataset (top left). Next, perform spatial cell type annotation on ST data using STRIDE and RGAST respectively and define near and far distributed cell type pairs based on the annotation (bottom left). Then feed annotated ST data to stLearn and extract predicted results and then evaluate the performances of different methods (right). See Methods for detail. **b,** Cell type annotated by STRIDE in human intestine A3 sample. **c,** Cell type annotated by RGAST in human intestine

A3 sample. **d,** Barplot of DES for RGAST and STRIDE in human intestine data. **e,** Visualization of CCI results predicted by RGAST.

Liu et al. developed a method to evaluate cell-cell interaction (CCI) tools by assessing the coherence between expected and observed spatial distance tendencies using ST data[24]. They conducted a comprehensive evaluation of various CCI tools. In this study, we followed the same evaluation workflow and benchmarked RGAST and STIDE, a deconvolution method for ST data that they identified as the most suitable method to be used in conjunction with a specific CCI analysis tool[24], on the human intestine dataset (Fig. 4a, Methods). For identifying short-term interactions, we employed stLearn as the downstream CCI analysis tool, as it has demonstrated superior performance in this dataset[24]. The distance enrichment score (DES), proposed by Liu et al., was utilized to measure the consistency between expected and observed spatial tendencies, thereby evaluating the occurrence of CCIs in the actual tissue environment (Methods). A higher DES value indicates a closer alignment between the inferred results and the real situation.

The results revealed that in this human intestine sample, the cell annotation performed by RGAST exhibited a smoother spatial context, and the cell types demonstrated a clearer layer structure compared to STRIDE (Fig. 4b,c). Quantitatively, the Distance Enrichment Score (DES) obtained using RGAST annotation outperformed the previously identified best-performing method, STRIDE [24] (Fig. 4d). Notably, the top interacting ligand-receptor pairs showed similarity between the two results, but the corresponding cell types differed (Supplementary Fig. S18). This suggested that the Cell-Cell Interactions (CCIs) predicted by stLearn, using RGAST annotation, aligned more closely with the actual scenario, further confirming the superiority of RGAST in spatial domain detection. Additionally, the CCI results generated by RGAST highlighted the complexity of cell type interactions within the human intestine developmental system (Fig. 4e).

RGAST reveals cell trajectories more accurately

Trajectory inference is a method used to infer the pattern of cells in dynamic developmental processes, with pseudotime representing the progression through this process. Improving the quality of the embedding method can significantly enhance the accuracy and resolution of trajectory inference results[25]. In this study, we selected the DLPFC dataset for trajectory and pseudotime inference, as this tissue represents a natural developmental progression. We used section 151676 as the benchmarking data and compared three methods - SEDR, STAGATE, and conST - with RGAST, as these methods have shown superior performance in trajectory inference analysis[11, 12, 17]. To validate the quality of the learned embeddings, we used Monocle3[26] and PAGA[27] to produce trajectory and pseudotime inference results with embeddings generated from different methods. For pseudotime inference, we selected white matter (WM) as the starting point (Methods).

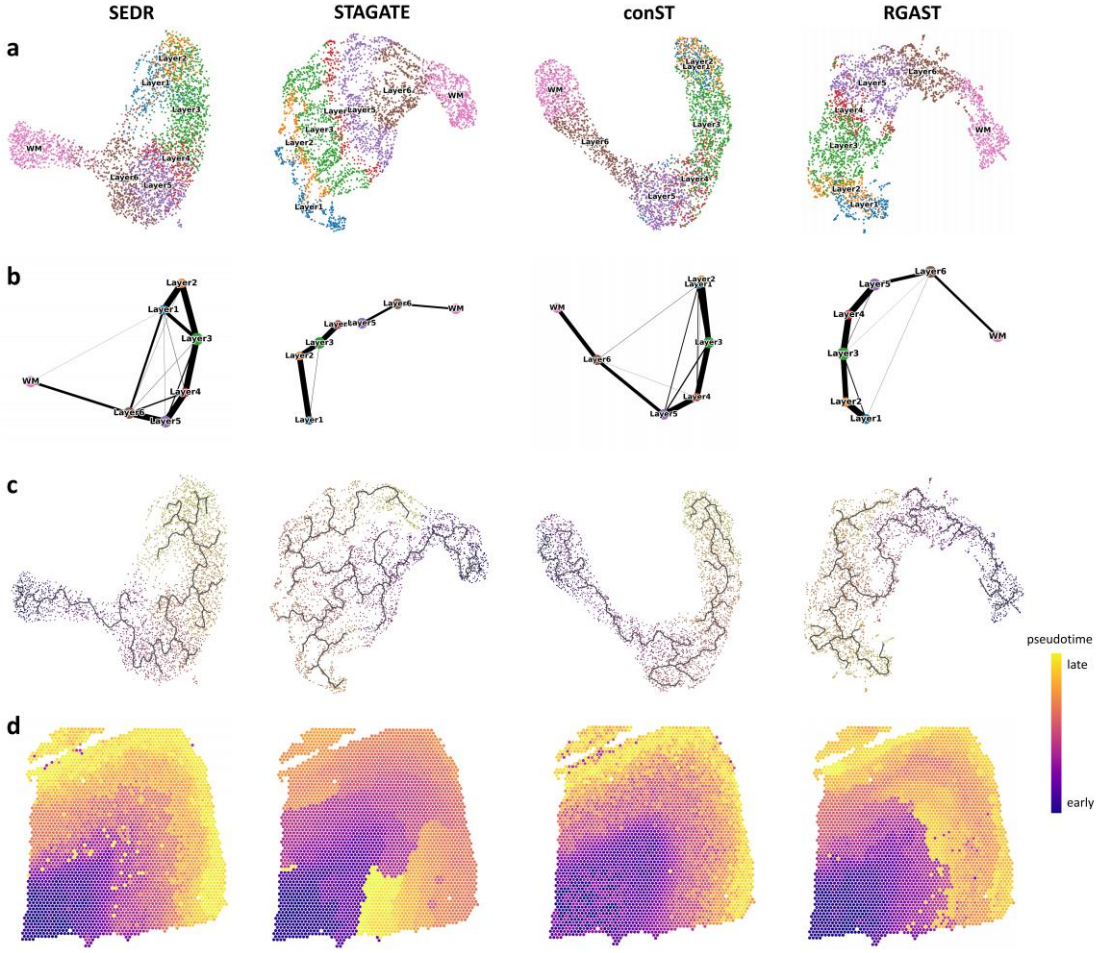

Fig. 5 | RGAST reveals cell trajectories more accurately. a, UMAP visualizations colored by different layers generated by SEDR, STAGATE, conST, and RGAST embeddings, respectively, in the DLPFC section 151676. The order of plots for different methods in b, c, and d is consistent with a. **b,** PAGA graph generated by different methods. **c,** UMAP visualizations colored by pseudotime inferred from SEDR, STAGATE, conST, and RGAST embeddings, respectively. **d,** Pseudotime inference results from different methods depicted on the spatial physical context.

Based on the results, the UMAP visualization generated by the SEDR embedding appeared as a circular structure in which layer 1 and layer 6 are connected to each other (Fig. 5a). In contrast, the conST and RGAST embeddings showed a streamlined expansion of different layers, while the UMAP visualization of the STAGATE embedding appears clumped together without an obvious linear structure (Fig. 5a). Regarding the PAGA graph, our results demonstrated that adjacent cortical layers tend to share greater similarity, with very weak connections between discontinuous layers in STAGATE and RGAST, which is consistent with the chronological order of cortex development. However, spurious connections between discontinuous layers are considerable in SEDR and conST (Fig. 5b). In the inferred development trajectory and pseudotime, STAGATE exhibited too many branching points, and the visualization in the spatial context confirmed the incorrect developmental trajectory inferred by STAGATE (Fig. 5c,d). While the pseudotime inferred by the other three methods

reflects the correct "inside-out" developmental ordering of cortical layers, the developmental pseudotime inferred by SEDR and conST is not as smooth as RGAST and exhibits more outliers in each layer structure (Fig. 5c,d). Therefore, our results suggest that RGAST can reveal cell trajectories more accurately than the other methods.

RGAST can reveal 3D spatial patterns with multiple ST sections

A single slice of ST data can provide a 2D spatial pattern of a specific organ. When multiple consecutive slices are available for the same organ, we can leverage the relationships between these slices to achieve a more precise representation of each spot. In this study, we extended RGAST to handle 3D ST data by simultaneously considering the 2D relationships within each section and the spatial neighborhoods between adjacent sections (Fig. 6a, Methods). To demonstrate the effectiveness of our approach, we applied RGAST to a pseudo-3D ST dataset obtained by aligning six mouse hypothalamic preoptic sections profiled using MERFISH (Fig. 6b).

Based on the clustering results in the spatial domain, we manually annotated the cell type of each domain using marker genes (Fig. 6c and Supplementary Fig. S19). Interestingly, although spots from different slices merged well in the UMAP visualization, there was a slight separation between the anterior sections (Bergman_0.26, 0.16, and 0.06) and posterior sections (Bergman_-0.04, -0.14, and -0.24) (Fig. 6d). This observation suggests that RGAST has the potential to eliminate technical noise, such as batch effects, while preserving biological differences. In the 3D spatial domain results detected by RGAST, a clear structure pattern was observed for mature oligodendrocytes and ependymal cells (Fig. 6e, f), which is consistent with previous studies[21]. Additional visualizations of other spatial domains can be found in Supplementary Fig. S20. These results illustrate that RGAST can facilitate the reconstruction of 3D tissue models and accurately capture 3D expression patterns by incorporating spatial information.

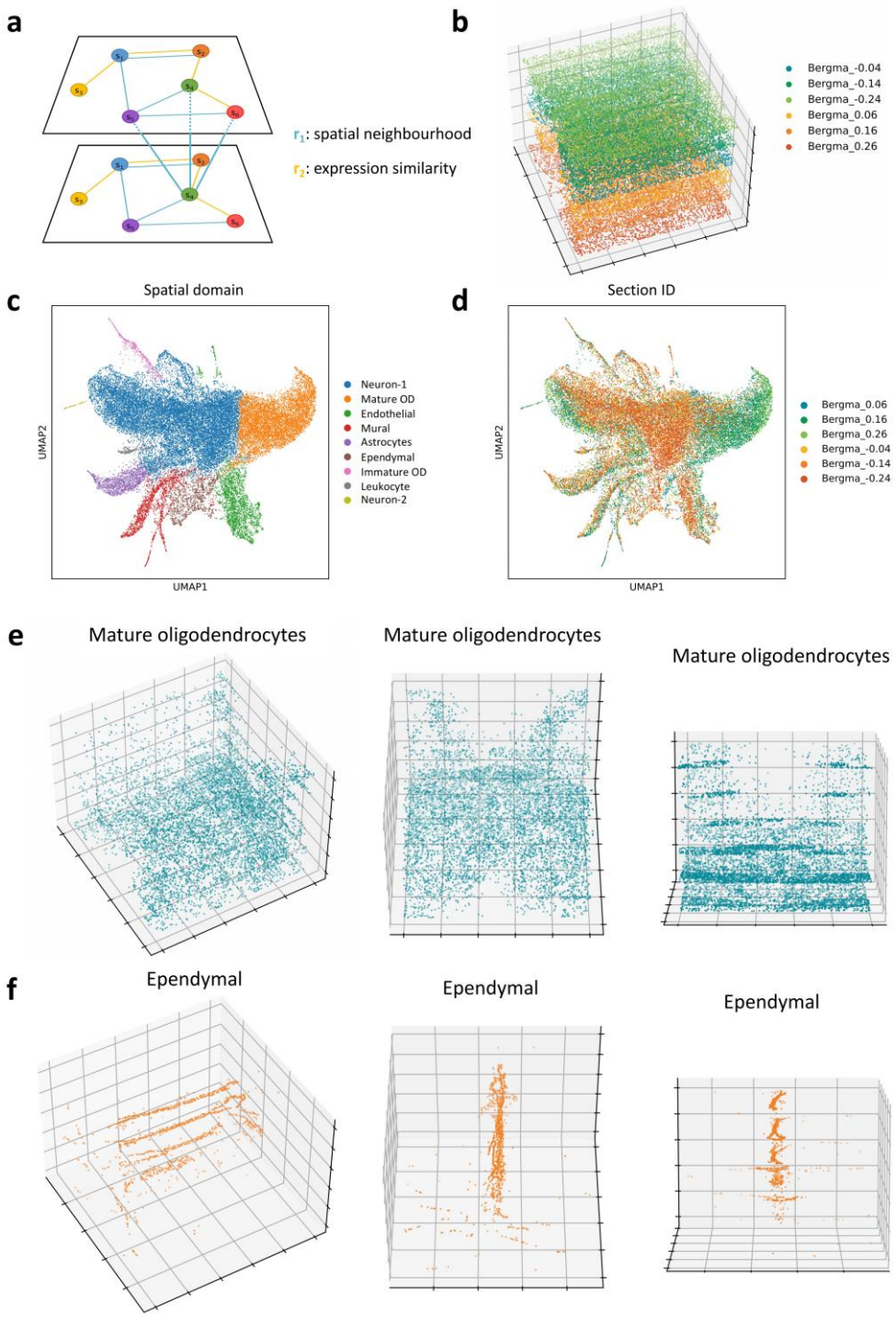

Fig. 6 | RGAST can reveal 3D spatial patterns with multiple ST sections. a, The 3D RGAST is constructed by by simultaneously considering the 2D relationships within each section and the spatial neighborhoods between adjacent sections. **b,** The alignied six mouse hypothalamic preoptic sections profiled using MERFISH. **c,** UMAP visualization of the RGAST embedding using the 3D model colored by manually annotated spatial domain. **d,** UMAP visualization of the RGAST embedding using the 3D model colored by different section. **e,** Visualization of the 3D structure pattern of mature oligodendrocytes from different angles. **f,** Visualization of the 3D structure pattern of Ependymal from different angles.

Discussion

Spatial transcriptomics is a groundbreaking technology that enables researchers to investigate gene expression and tissue architecture in a spatially resolved manner. Different ST technologies allow for genome-wide profiling of gene expression at the level of individual cells or even subcellular structures. These technologies hold immense potential to provide new insights into diverse biological processes, such as development, disease progression, and drug response. Although several spatial domain detection and representation learning methods have been developed recently, they often treat spatial relationships in an explicit or predefined manner, which may not fully capture the complex and subtle spatial patterns present in the data. In this paper, we introduced RGAST, a relational graph attention network-based model for spatial transcriptome analysis. RGAST constructs a relational graph attention network to learn the representation of each spot in the spatial transcriptome data. By incorporating an attention mechanism, RGAST considers both gene expression similarity and spatial neighbor relationships in constructing the heterogeneous graph network, facilitating a more comprehensive and flexible representation of the spatial transcriptome data. The learned RGAST embedding can be employed for clustering and various downstream analyses.

We first applied RGAST to ST data generated using various technologies with different resolutions to see its spatial clustering performance. RGAST outperformed state-of-the-art methods significantly, accurately detecting the laminar structure of different layers in the DLPFC and mouse olfactory bulb data. Even in datasets with discrete cell type distributions, such as mouse hypothalamic preoptic and mouse cortex data, RGAST demonstrated superior performance while STAGATE appeared to overuse the spatial neighborhood relationship, resulting in evident clustering characteristics of block aggregation. The batch technique used by RGAST also enable the scalability of RGAST when encountering dataset with high throughput and resolution. While the current RGAST methodology focuses on gene expression and spatial information, it does not utilize histological images. Contemporary methods like SpaGCN and stLearn incorporate histological images as input, but our study revealed suboptimal utilization of these images. SpaGCN uses pixel-level features by directly calculating mean color values from the RGB channels. However, pixel values are susceptible to noise and do not provide semantic features for cell analysis. On the other hand, stLearn employs a pre-trained deep learning model for image feature extraction from spots, integrating them with spatial location and gene expression. Nevertheless, stLearn does not fine-tune the network specifically for histological images, as it utilizes models pre-trained on natural images. In the future, the attention mechanism and heterogeneous graph of RGAST can be extended to adaptively integrate histological image features, offering a more comprehensive analysis.

When applying different latent embeddings for trajectory inference in the developmental system's DLPFC dataset, we observed that adjacent cortical layers

exhibited greater similarity in both STAGATE and RGAST. However, in terms of the inferred developmental trajectory and pseudotime, STAGATE displayed an excessive number of branching points. Visualization within the spatial context further confirmed the incorrect developmental trajectory inferred by STAGATE. These findings suggest that RGAST provides a more accurate representation of cell trajectories compared to other methods. We further validated the superiority of RGAST by applying it to two additional downstream analysis tasks: SVG detection and CCI analysis. Firstly, the SVGs detected by RGAST exhibited a clearer spatial distribution pattern and were more representative. Secondly, the CCI results based on RGAST demonstrated higher consistency with the ground truth compared to the recommended best deconvolution method from a previous study.

To enhance the analytical power of spatial omics, it is crucial to integrate multiple datasets obtained from the same tissue. In this study, we utilized six consecutive mouse hypothalamic preoptic MERFISH sections to construct a 3D RGAST model, aiming to achieve a more precise representation of each spot. Remarkably, while the spots from different slices merged well in the UMAP visualization, a slight separation between the anterior and posterior sections was observed. This finding suggests that RGAST has the potential to eliminate technical noise, such as batch effects, while preserving biological differences. Moreover, the 3D spatial domain detected by RGAST provided us with a clearer insight into the structural pattern of the hypothalamic preoptic region.

In summary, RGAST presents a promising novel approach that integrates transcriptomic data and spatial coordinates to construct a comprehensive representation of spots. The low-dimensional embedding derived from RGAST offers improved accuracy in spatial clustering, SVG detection, CCI analysis , trajectory inference, and 3D modeling. We firmly believe that RGAST, as a novel technique for representation learning in spatial transcriptomics (ST) data, provides an unprecedented opportunity to effectively unravel the underlying mechanisms of organ development, disease progression, and drug response within a spatially resolved context.

Methods

Datasets and preprocessing

We applied RGAST to spatial transcriptomics (ST) datasets generated by different platforms, including 10x Visium, Stereo-seq, MERFISH, and seqFISH+. Specifically, the 10X Visium DLPFC dataset included 12 human DLPFC sections sampled from three individuals. The number of spots ranged from 3,498 to 4,789 for each section, and the original authors manually annotated the areas of DLPFC layers and white matter (WM)[20]. The human intestine dataset was a developmental system dataset from human fetal intestine samples collected at 12 and 19 post-conception weeks (PCW)[28]. We used the A3 sample in our research, which contained only 434 spots. The Stereo-seq mouse olfactory bulb data was binned into a resolution of cellular levels (~14 μm) and contained 19,109 spots[11]. The MERFISH dataset was a spatially resolved cell

atlas of the mouse hypothalamic preoptic region, profiling ~1 million cells, each with only 155 genes[21]. We selected six consecutive slices from a single animal sample to conduct our experiments. After preprocessing, the number of spots ranged from 4,787 to 5,926 for each section. The mouse cortex seqFISH+ dataset contained mRNAs for 10,000 genes in single cells, with high accuracy and sub-diffraction-limit resolution. This dataset had a spot number of 913[6]. The data sources are summarized in in Supplementary Table S1.

For the mouse hypothalamic MERFISH data, we removed the putative doublets first. For all datasets, we conducted normalization and scaling, and then ran PCA using the Scanpy package[10]. The PCA representation of each spot was selected as the input of RGAST.

Construction of relational graph

The spatial neighborhood relationship is established by considering the Euclidean distance between the spatial locations of different spots. In the case of 10x Visium data, we set the graph to include the six nearest neighbors for each spot. For other datasets, we define the adjacency matrix, A, such that A_{ij}= 1 if and only if the Euclidean distance between spot i and spot j is less than a pre-defined hyperparameter, r. The expression similarity relationship is constructed by considering the Euclidean distance of the PCA representation of the gene expression spectra of different spots. Across all datasets, we connect each spot to the six spots exhibiting the most similar expression patterns. The specific hyperparameter settings used for analyzing different datasets in our research can be found in Supplementary Table S2.

Relational graph attention auto-encoder

The relational graph attention auto-encoder consists of an encoder of two relational graph attention layer and a linear layer decoder.

<u>*Encoder.*</u> The input of the encoder in our architecture is a graph with $R = 2$ relation types and N nodes (spots). The i_{th} node is represented by a feature vector of the PCA dimensional reduction representation of gene expressions $\boldsymbol{x}_i$. Query and key representations for the l_{th} encoder layer are computed for each relation type with the help of both query and key kernels, i.e.

$$q_i^{(r)} = \boldsymbol{W}_l^{(r)} \boldsymbol{x}_i \cdot \boldsymbol{Q}^{(r)} \tag{1}$$

$$k_i^{(r)} = \boldsymbol{W}_l^{(r)} \boldsymbol{x}_i \cdot \boldsymbol{K}^{(r)} \tag{2}$$

where $\boldsymbol{W}_l$ is the trainable weight matrix of layer l. Then additive attention[29] is applied to compute attention logits $a_{ij}^{(r)}$:

$$a_{ij}^{(r)} = LeakyRelu\left(q_i^{(r)} + k_j^{(r)}\right) \tag{3}$$

Then the attention coefficients for each relation type are then obtained via the across-relation attention mechanism:

$$\alpha_{ij}^{(r)} = \frac{exp\left(a_{ij}^{(r)}\right)}{\sum_{r' \in \mathcal{R}} \sum_{k \in \mathcal{N}_{r'}^{(i)}} exp\left(a_{ik}^{(r')}\right)} \tag{4}$$

where $\mathcal{R}$ denotes the set of relations, i.e. edge types. $\mathcal{N}_{r'}^{(i)}$ denotes the set of spots connected to node i under relation r'. To avoid overfitting, we employed the dropout strategy on the normalized attention coefficients with dropout rate=0.3. That is, $\alpha_{ij}^{(r)}$ will be set as 0 with a probability of 0.3.

Then spot i collectively aggregating information from spots connected to it with the neighborhood aggregation step to get the output of layer l. We denote the intermediate representation of $\boldsymbol{x}_j^{(r)}$ as $\boldsymbol{h}_j^{(r)}$: $\boldsymbol{h}_j^{(r)} = \boldsymbol{W}_l^{(r)} \boldsymbol{x}_j$. To enhance the discriminative power of the RGAST layer, we further implement the additive cardinality preservation mechanism[30]:

$$\boldsymbol{x}_i'^{(r)} = \sum_{j \in \mathcal{N}_r^{(i)}} \alpha_{ij}^{(r)} \boldsymbol{h}_j^{(r)} + \mathcal{W} \odot \sum_{j \in \mathcal{N}_r^{(i)}} \boldsymbol{h}_j^{(r)} \tag{5}$$

$$\boldsymbol{x}_i' = \sum_{r \in \mathcal{R}} \boldsymbol{x}_i'^{(r)} \tag{6}$$

where $\mathcal{W}$ is a non-zero vector $\in R^n$, n is the output dimension of layer l, $\odot$ denotes the elementwise multiplication. The output of encoder is considered as the final spot embedding.

<u>*Decoder.*</u> The decoder reverses the latent embedding back into the original PCA representation of the expression profile. The one-layer linear decoder treats the output of the encoder (denoted by $\boldsymbol{h}_i$) as its input and computes the reconstructed result. Specifically, the decoder computes the reconstructed result as follows:

$$\widehat{\boldsymbol{x}_i} = \boldsymbol{h}_i \boldsymbol{W}^T + \boldsymbol{b} \tag{7}$$

Where $\boldsymbol{W}$ and $\boldsymbol{b}$ are learnable weight matrix and bias vector.

<u>*Loss function.*</u> The objective of RGAST is to minimize the reconstruction loss of the original PCA profiles as follows:

$$\sum_{i=1}^{N} ||\widehat{\boldsymbol{x}_i} - \boldsymbol{x}_i||_2 \tag{8}$$

In all experiments we used Adam optimizer with learning rate=0.001 and weight decay=0.0001.

Deep embedding clustering (optional)

To enhance the compactness of the RGAST embedding, we use an unsupervised deep embedded clustering (DEC) method to iteratively group cells into different clusters[19].

To initialize the cluster centers, we first apply a clustering method to the learned latent representations (such as Louvain, Leiden, Kmeans, etc.). With this initialization, DEC improves the clustering using an unsupervised iterative method of two steps. In the first step, DEC calculates a soft assignment q_{ij} of the latent point h_i to the cluster center mui using the Student's t-distribution:

$$q_{ij} = \frac{\left(1+\left\|h_i-\mu_j\right\|^2\right)^{-1}}{\sum_{j'}\left(1+\left\|h_i-\mu_{j'}\right\|^2\right)^{-1}} \tag{9}$$

In the second step, we iteratively refine the clusters by learning from their high confidence assignments with the help of an auxiliary target distribution p based on q_{ij}:

$$p_{ij} = \frac{q_{ij}^2/\sum_i q_{ij}}{\sum_{j'}\left(q_{ij'}^2/\sum_i q_{ij'}\right)} \tag{10}$$

Based on the soft assignment q_{ij} and auxiliary target distribution p_{ij}, an objective function is defined using the KL divergence:

$$KL(P||Q) = \sum_i \sum_j p_{ij}\, log \frac{p_{ij}}{q_{ij}} \tag{11}$$

Then the overall loss function is the sum of (8) and (11). The RGAST parameters and cluster centers are then simultaneously optimized using Adam optimizer.

Batch technique for scalable training of RGAST

Training RGAST becomes challenging when the number of spots is large, as mini-batch training is not well-defined for graph neuron networks. However, in ST data, it is widely acknowledged that spots in close proximity exhibit higher similarity, and spatial neighborhood relationships are also established based on their spatial locations in our model. Our network can be partitioned into several subgraphs based on the quantile of spot locations along different axes with the price of getting rid of some long-term gene expression relationships. In our Stereo-seq mouse olfactory bulb dataset, we divided the graph into 2x2 subgraphs. The dividing point along the x-axis is determined by the median of spot locations, which is the same for the y-axis. These subgraphs were then fed into the RGAST model as mini-batch data.

Spatial domain detection / clustering analysis

Leiden algorithm

We used the Scanpy package to compute the Leiden clustering from the latent embeddings generated from different representation learning methods. Briefly, we used latent embeddings to compute neighborhood graphs with number of neighbors as the recommended default parameter and Euclidean distance as the distance measure. To set the maximum number of unique Leiden clusters, grid-searching on the Leiden clustering resolutions between 0.0 and 2.5 was performed at intervals of 0.02.

Evaluation metrices

Adjusted Rand index (ARI). The Rand Index computes a similarity measure between two clusterings by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clusterings. The raw RI score is then "adjusted for chance" into the ARI score using the following scheme:

$$ARI = \frac{RI - E(RI)}{max(RI) - E(RI)} \tag{12}$$

To calculate this value, first calculate the contingency table like that:

	Y_1	Y_2	...	Y_s	Sums
X_1	n_{11}	n_{12}	$\cdots$	n_{1s}	a_1
X_2	n_{21}	n_{22}	$\cdots$	n_{2s}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_r	n_{r1}	n_{r1}	$\cdots$	n_{rs}	a_r
Sums	b_1	b_2	$\cdots$	b_s	

each value in the table represents the number of data point located in both cluster (Y) and true class (X), and then calculate the ARI value through this table:

$$\overbrace{ARI}^{\text{Adjust Index}} = \frac{\overbrace{\sum_{ij}\binom{n_{ij}}{2}}^{\text{Index}} - \overbrace{\left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}^{\text{Expected Index}}}{\underbrace{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right]}_{\text{Max Index}} - \underbrace{\left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}_{\text{Expected Index}}} \tag{13}$$

The adjusted Rand index is thus ensured to have a value close to 0.0 for random labeling independently of the number of clusters and samples and exactly 1.0 when the clusterings are identical (up to a permutation). The adjusted Rand index is bounded below by -0.5 for especially discordant clusterings.

Silhouettes Coefficient (SC). SC is a metric used to evaluate the clustering effectiveness on dataset without ground truth label[31]. a_i is denoted as tightness, which represents the average distance between sample i and other samples within same cluster. The average distance between one sample and other samples from different cluster is defined as separation. $b_i = min[b_{i1}, b_{i2}, \dots, b_{ik}]$, which is the separation of sample i. Then the Silhouette Coefficient of sample i is defined as:

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)} \tag{14}$$

The SC score is the average of s_i ranging from -1 to 1 and higher score means higher performance.

Trajectory and pseudotime inference

PAGA graph and UMAPs

We also used the Scanpy package to compute the partition-based graph abstraction (PAGA), and uniform manifold approximation and projection (UMAP) from embeddings generated from different methods. After using latent embeddings to compute neighborhood graphs, UMAP algorithm was conducted based on the neighborhood graphs. Finally, PAGA was applied to quantify the connectivity between different layers based on the UMAP representations.

Monocle3

For SEDR, STAGATE, conST and RGAST, we first extracted the low-dimensional embedding and then used the embedding to replace the default PCs used in Monocle3 package. We then ran Monocle3 on UMAPs generated by the latent embedding using the recommended parameters and randomly set a spot in white matter (WM) as the starting point to generate the pseudo-time.

Spatially variable genes detection

Detection of SVGs

In this study, we employ the same detection method as that used in SpaGCN to identify spatially variable genes (SVGs) enriched in each spatial domain generated by different methods. We recognize that some genes may be expressed in multiple but disconnected domains. Therefore, instead of conducting differential expression (DE) analysis using spots from a target domain versus all other spots, we first select spots to form a neighboring set around the target domain. Our aim is to detect genes that are highly expressed in the target domain but not expressed or expressed at low levels in neighboring spots. To achieve this, we draw a circle with a predetermined radius around each spot in the target domain, and spots from non-target domains within the circle are considered its neighbors. The radius is set such that each spot in the target domain has an average of approximately ten neighbors. Next, we collect all the neighbors of spots in the target domain to form a neighboring set. For each non-target domain, if more than 50% of its spots are in the neighboring set, we select it as a neighboring domain. This criterion is in place to avoid cases where a domain is selected as a neighboring domain, but only a small proportion of its spots are adjacent to the target domain. Once neighboring domains are determined, we perform DE analysis between spots in the target domain and the neighboring domain(s) using the Wilcoxon rank-sum test. Genes with a false discovery rate (FDR)-adjusted P-value < 0.05 are selected as SVGs. To ensure only genes with enriched expression patterns in the target domain are selected, we require a gene to meet the following three criteria: (1) the percentage of spots expressing the gene in the target domain, that is, in-fraction, is $> 80\%$; (2) for each neighboring domain, the ratio of the percentages of spots expressing the gene in the target domain and the neighboring domain(s), that is, in/out fraction ratio, is > 1; and (3) the expression fold change between the target and neighboring domain(s) is > 1.5.

Evaluation metrices

Gene expressions at different locations may not be independent. For example, the expression levels of a gene at nearby locations may be closer in value than expression levels at locations that are farther apart. This phenomenon is called spatial autocorrelation, which measures the correlation of a variable with itself through space. To evaluate whether the detected SVGs exhibit an organized spatial expression pattern, we used Moran's I and Geary's C, two commonly used statistics to quantify the degree of spatial autocorrelation of gene expression.

Moran's I statistic. Moran's I metric is a correlation coefficient that measures the overall spatial autocorrelation of a dataset[32]. Essentially, it quantifies how similar one spot is to other spots surrounding it for a given gene. If the spots are attracted or repelled by each other, it suggests that they are not independent, and the presence of autocorrelation indicates a spatial pattern of gene expression. The Moran's I value ranges from -1 to 1, where a value close to 1 indicates a clear spatial pattern, a value close to 0 indicates random spatial expression, and a value close to -1 indicates a chessboard-like pattern. To evaluate the spatial variability of a given gene, we calculate the Moran's I using the following formula:

$$I = \frac{N}{W}\frac{\sum_i \sum_j [w_{ij}(x_i - \bar{x})(x_j - \bar{x})]}{\sum_i (x_i - \bar{x})^2} \tag{15}$$

where x_i and x_j are the gene expression of spots i and j, $\bar{x}$ is the mean expression of the gene, N is the total number of spots, w_{ij} is the spatial weight between spots i and j calculated using the 2D spatial coordinates of the spots, and W is the sum of w_{ij}. To select the k nearest neighbors for each spot, we use spatial coordinates. The Moran's I statistic is robust to the choice of k, and we set it to 5 in our analysis. We assign w_{ij}=1 if spot j is in the nearest neighbors of spot i, and w_{ij}=0 otherwise.

Geary's C statistic. Geary's C statistic is another measure of spatial autocorrelation, which is a statistical method used to examine the degree to which neighboring observations in a dataset are similar to each other for a given gene[33]. In other words, it measures the extent to which spots that are close to each other in space are more similar or dissimilar than those that are farther apart with respect to the gene expression. Geary's C is calculated using the following formula:

$$C = \frac{N}{2W}\frac{\sum_i \sum_j \left[w_{ij}(x_i - x_j)^2\right]}{\sum_i (x_i - \bar{x})^2} \tag{16}$$

The value of Geary's C ranges from 0 to 2, with a value of 1 indicating no spatial autocorrelation, values less than 1 indicating positive spatial autocorrelation (i.e., neighboring expressions are more similar than expected by chance), and values greater than 1 indicating negative spatial autocorrelation (i.e., neighboring expressions are more dissimilar than expected by chance).

Cell-cell interaction analysis

Spatial gene expression distribution distance

Using the physical locations of each spot as coordinates and gene expression levels as values, we constructed spatial gene expression distributions from the spatial transcriptomic (ST) data. The distance between the spatial expression distributions of two genes can be measured using the Wasserstein distance, as both distributions are discrete. The Wasserstein distance is a commonly used measurement for comparing two distributions, solving an optimal transport problem to evaluate the distance. In our case, assuming the ligand gene is expressed in m spots and the receptor gene is expressed in n spots of the ST data, we generated a matrix $D \in \boldsymbol{R}^{m\times n}$ to record the Euclidean distance between spots with expression of the ligand or receptor genes, based on their spatial coordinates. Next, the ligand and receptor gene expression distributions were transformed into one-dimensional vectors: $L \in \boldsymbol{R}^{m\times 1}$ and $R \in \boldsymbol{R}^{n\times 1}$, while preserving their spatial distances in matrix D. The Wasserstein distance from the ligand to the receptor can be determined by finding an optimal transport plan $\gamma' \in \boldsymbol{R}^{m\times n}$ that minimizes the total transport cost. The total transport cost is defined as the sum of the products of the transport value and the Euclidean distance between each spot. The optimal transport plan γ' can be calculated using the following formula:

$$\gamma' = argmin_{\gamma\in\Gamma(L,R)} < \gamma, D > \tag{17}$$

Furthermore, based on the optimal transport plan γ', the Wasserstein distance can be computed as follows:

$$W(L,R) = \min_{\gamma\in\Gamma(L,R)} < \gamma, D > = < \gamma', D > \tag{18}$$

The final distance between ligand and receptor distributions was defined as the mean of the Wasserstein distances from both directions:

$$W_{LR} = \big(W(L,R) + W(R,L)\big)/2 \tag{19}$$

where $W(L,R)$ represents the Wasserstein distances from the ligand gene distribution to the receptor gene distribution and the $W(R,L)$ represents the Wasserstein distances with the direction of receptor gene distribution to ligand gene distribution. The W_{LR} represents the final Wasserstein distance between ligand and receptor.

Ligand-receptor spatial distance ratio and P-value

We utilized the Wasserstein distance to quantify the distance between two gene expression distributions and further introduced another variable to represent the spatial interaction tendency. Specifically, for a ligand-receptor interaction, we denoted the ligand gene expression distribution as L and the receptor gene expression distribution as R. To facilitate convenience, we referred to the actual Wasserstein distance of this interaction as d_{real}. To gauge the significance of the spatial interaction, we constructed random gene expression distributions for the ligand and receptor, denoted as L_r and R_r, respectively, by permuting the coordinates for each spot in L and R. We calculated the Wasserstein distance between these two random distributions, referred to as $W_{L_rR_r}$, using the same approach, which we denoted as $d_simulation$ for convenience. By

repeating this permuting process a sufficient number of times (in our case, 1000 times), we generated a set of $d_simulation$ values. Consequently, we employed the ratio of d_real to the mean of the $d_simulation$ set, denoted as d_ratio, to quantify the spatial tendency of this interaction:

$$d_{ratio} = \frac{d_{real}}{\frac{1}{n}\sum_{i=1}^{n} d_{simulation_i}} \tag{20}$$

where n represents the number of permutations. Additionally, we constructed a null distribution of d_real using the $d_simulation$ set. Subsequently, we calculated a one-sided permutation test based on the null distribution to obtain a P-value, indicating the significance of the spatial interaction tendency. In our case, we calculated left-sided P-values to identify short-range interactions since stLearn is designed to find short-range interactions.

Filter short-range interactions

To filter short-range interactions, we ranked them based on their spatial distance tendencies (d_ratio) and significance (P-value). Initially, we extracted ligand-receptor interactions from CellChatDB, which provided interaction type annotations such as cell-cell contact and secreted signaling types, facilitating functional validation[34]. As CellChatDB includes multi-subunit complexes, we split them into single-subunit interactions to align with stLearn, which does not consider multi-subunit complexes. However, we retained the original multi-subunit interactions. The union of CellChatDB and the rearranged single-subunit interactions formed our known interaction list for subsequent steps. For the human intestine dataset, we filtered interactions where both ligand and receptor genes were expressed in at least 10% of the total spots from our known interaction list. Subsequently, we calculated the d_ratio and P-value for each filtered interaction, ranking them accordingly. To identify potential short-range interactions, we ranked interactions in ascending order based on their d_ratio and left-sided P-value, selecting the top 10% of these ranked interactions. Next, we filtered the potential interactions further by considering only those with a P-value less than 0.01, designating them as significant short-range interactions.

Defining near and far cell type pairs

Following spatial cell type annotation, we constructed spatial distributions for each cell type, representing the spatial locations of spots specific to each cell type. To simplify the process, we utilized the average Euclidean distance as the distance function. The formulas for calculating the distances are as follows:

$$dis(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$$

$$ct_{distance}(ct_a|ct_b) = \left(\frac{\sum_{i=1}^{n_a} \min_{j=1,2,\ldots,n_b} dis(a_i,b_j)}{n_a} + \frac{\sum_{j=1}^{n_b} \min_{i=1,2,\ldots,n_a} dis(a_i,b_j)}{n_b}\right)/2 \tag{21}$$

where ct_a and ct_b represent cell types a and b, respectively. n_a and n_b denote the number of spots corresponding to each cell type. a_i refers to the i_{th} spot in cell

type a. The $dis()$ function computes the Euclidean distance between two spots. The spatial distance between cell types can be quantified by calculating the mean of the minimum Euclidean distances between the spots.

We designated the combination of two specific cell types as a cell type pair and classified the cell type pairs in each ST sample into three categories: near, medium, and far, based on their spatial distances. To achieve this, we applied k-means clustering with k=3 to partition these cell type pairs into three clusters. We then annotated the clusters as near, medium, and far in the order of their average distances.

Distance enrichment score (DES)

The Distance Enrichment Score (DES) is utilized to quantify the consistency between expected and observed spatial distance tendencies. A higher DES indicates better consistency between the expected and observed spatial tendencies. Initially, we ranked the short-range interactions using their d_ratio and P-value, forming the expected ranked interaction list $L_s = \{lr_1, lr_2, \dots, lr_{n_s}\}$, where lr_i represents the i_th ligand-receptor interaction, and n_s is the total number of short-range interactions. Next, we extracted interactions from stLearn's results and created the observed interaction list S for each cell type pair. Specifically, we denoted the observed interaction list for the near cell type pair ct_n as S_n. The DES is computed by adding a weighted proportion of the P-value (P_{match}) when an interaction exists in S_n and subtracting an unmatched weight ($P_{unmatch}$) when an interaction is absent in S_n while traversing through L_s. The P_{match} and $P_{unmatch}$ for the j_{th} interaction in L_s are defined as follows:

$$P_{match}(S_n, j) = \sum_{lr_j \in S_n} \frac{1 - P_value_j}{sum_{lr_j \in L_s}(1 - P_value_j)}$$

$$P_{unmatch}(S_n, j) = \sum_{lr_j \notin S_n} \frac{1}{|n_s - n_m|} \tag{22}$$

where n_m is the total number of matched interactions between S_n and L_s. The DES is the maximum deviation of $(P_{match} - P_{unmatch})$ from 0.

3D RGAST model

All current ST technologies capture gene expression patterns within the context of 2D tissue sections, which limits the accurate representation of 3D spatial information in real-world samples. A conventional approach to address this limitation is by reconstructing gene expressions in 3D space through the stacking of consecutive 2D sections[3, 35]. However, the presence of batch effects between sections poses a challenge in extracting meaningful 3D spatial patterns. Here, we proposed a 3D RGAST model that incorporates both the 2D relationships within each section and the spatial neighborhood relationships between adjacent sections to mitigate the impact of batch effects. If the presence of minimal batch effects between different sections is confirmed, gene expression similarity can also be applied across sections. Specifically, the spatial neighborhoods between adjacent sections are constructed based on aligned coordinates and a pre-defined radius. The underlying principle behind the use of 3D spatial

neighborhoods lies in the assumption that biological differences between consecutive sections should demonstrate continuity. By enhancing the similarity between adjacent sections, we aim to eliminate discontinuous independent technical noises.

Declarations

Data and code availability

All data used in this research can be found in **Supplementary Table S1**. Our RGAST method is available as a Python package on PyPI at https://pypi.org/project/RGAST, free for academic use, and the source code is openly available from our GitHub repository at https://github.com/GYQ-form/RGAST.

Competing interests

The authors declare no competing interests.

Authors' contributions

YG performed the research, analyzed data and wrote the original manuscript, ZY supervised the research. YG and YZ discussed and revised the manuscript.

Acknowledgements

We are thankful for Xin Yuan for providing practical suggesting for the research. The computations in this paper were run on the π 2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University.

Reference

1. Stahl, P.L., et al., *Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.* Science, 2016. **353**(6294): p. 78-82.
2. Ji, A.L., et al., *Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma.* Cell, 2020. **182**(2): p. 497-514 e22.
3. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science, 2019. **363**(6434): p. 1463-1467.
4. Chen, A., et al., *Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays.* Cell, 2022. **185**(10): p. 1777-1792 e21.
5. Chen, K.H., et al., *RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.* Science, 2015. **348**(6233): p. aaa6090.
6. Eng, C.L., et al., *Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH.* Nature, 2019. **568**(7751): p. 235-239.
7. Svensson, V., et al., *Power analysis of single-cell RNA-sequencing experiments.* Nat Methods, 2017. **14**(4): p. 381-387.
8. Regev, A., et al., *The Human Cell Atlas.* Elife, 2017. **6**.
9. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.
10. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data*

analysis. Genome Biol, 2018. **19**(1): p. 15.
11. Fu, H., et al., *Unsupervised spatially embedded deep representation of spatial transcriptomics.* bioRxiv, 2021.
12. Dong, K. and S. Zhang, *Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder.* Nat Commun, 2022. **13**(1): p. 1739.
13. Zhao, E., et al., *Spatial transcriptomics at subspot resolution with BayesSpace.* Nat Biotechnol, 2021. **39**(11): p. 1375-1384.
14. Dries, R., et al., *Giotto: a toolbox for integrative analysis and visualization of spatial expression data.* Genome Biol, 2021. **22**(1): p. 78.
15. 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, 2021. **18**(11): p. 1342-1351.
16. 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.
17. Zong, Y., et al., *conST: an interpretable multi-modal contrastive learning framework for spatial transcriptomics.* bioRxiv, 2022.
18. Armingol, E., et al., *Deciphering cell-cell interactions and communication from gene expression.* Nat Rev Genet, 2021. **22**(2): p. 71-88.
19. Xie, J., R. Girshick, and A. Farhadi, *Unsupervised Deep Embedding for Clustering Analysis*, in *Proceedings of The 33rd International Conference on Machine Learning*, B. Maria Florina and Q.W. Kilian, Editors. 2016, PMLR: Proceedings of Machine Learning Research. p. 478--487.
20. Maynard, K.R., et al., *Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex.* Nat Neurosci, 2021. **24**(3): p. 425-436.
21. Moffitt, J.R., et al., *Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region.* Science, 2018. **362**(6416).
22. Zhou, X., et al., *Circuit design features of a stable two-cell system.* Cell, 2018. **172**(4): p. 744-757. e17.
23. Stuart, T. and R. Satija, *Integrative single-cell analysis.* Nat Rev Genet, 2019. **20**(5): p. 257-272.
24. Liu, Z., D. Sun, and C. Wang, *Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information.* Genome Biol, 2022. **23**(1): p. 218.
25. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nat Methods, 2017. **14**(9): p. 865-868.
26. Cao, J., et al., *The single-cell transcriptional landscape of mammalian organogenesis.* Nature, 2019. **566**(7745): p. 496-502.
27. Wolf, F.A., et al., *PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells.* Genome Biol, 2019. **20**(1): p. 59.
28. Fawkner-Corbett, D., et al., *Spatiotemporal analysis of human intestinal development at single-cell resolution.* Cell, 2021. **184**(3): p. 810-826 e23.
29. Busbridge, D., et al., *Relational graph attention networks.* arXiv preprint arXiv:1904.05811, 2019.
30. Zhang, S. and L. Xie. *Improving attention mechanism in graph neural networks via cardinality preservation*. in *IJCAI: proceedings of the conference*. 2020. NIH Public Access.
31. Rousseeuw, P.J., *Silhouettes: a graphical aid to the interpretation and validation of cluster*

analysis. Journal of computational and applied mathematics, 1987. **20**: p. 53-65.

32. Li, H., C.A. Calder, and N. Cressie, *Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model.* Geographical analysis, 2007. **39**(4): p. 357-375.
33. Geary, R.C., *The contiguity ratio and statistical mapping.* The incorporated statistician, 1954. **5**(3): p. 115-146.
34. Jin, S., et al., *Inference and analysis of cell-cell communication using CellChat.* Nature Communications, 2021. **12**(1): p. 1088.
35. Ortiz, C., et al., *Molecular atlas of the adult mouse brain.* Science advances, 2020. **6**(26): p. eabb3446.