Briefings in Bioinformatics, 2026, 27, bbag248
https://doi.org/10.1093/bib/bbag248
Published: 25 May 2026
Problem Solving Protocol

OXFORD

Identifying batch-integrated domains from spatial transcriptomics via graph autoencoder with contrastive learning based on cross-modality and data augmentation

Yexuan Mao[1,‡], Lijun Quan[1,2,3,*,‡], Xiangyu Chen[1], Siyuan Wang[1], Guozheng Zhang[1], Siqi Li[1], Yelu Jiang[1], Liangpeng Nie[1], Tingfang Wu[1,2,3], Lingkun Meng[4,*], Qiang Lyu[1,2,3,*]

[1]School of Computer Science and Technology, Soochow University, 333 Ganjiang East Road, Gusu District, Suzhou, Jiangsu 215006, China
[2]Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, 163 Xianlin Avenue, Qixia District, Nanjing, Jiangsu 210023, China
[3]Jiangsu Key Laboratory of Drug Discovery and Translational Research for Brain Diseases, Soochow University, 199 Renai Road, Suzhou Industrial Park, Suzhou, Jiangsu 215123, China
[4]China Mobile (Suzhou) Software Technology Company Limited, 58 Kunlunshan Road, Huqiu District, Suzhou, Jiangsu 215163, China
*Corresponding authors. Lijun Quan, School of Computer Science and Technology, Soochow University, 333 Ganjiang East Road, Gusu District, Suzhou, Jiangsu 215006, China. E-mail: ljquan@suda.edu.cn; Lingkun Meng,
China Mobile (Suzhou) Software Technology Company Limited, 58 Kunlunshan Road, Huqiu District, Suzhou, Jiangsu 215163, China. E-mail: menglingkun@cmss.chinamobile.com; Qiang Lyu, School of Computer Science and Technology, Soochow University, 333 Ganjiang East Road, Gusu District, Suzhou, Jiangsu 215006, China. E-mail: qiang@suda.edu.cn.
‡Yexuan Mao and Lijun Quan contributed equally to this work.

Abstract

Spatially resolved transcriptomics (SRT) allows for the comprehensive profiling of gene expression while preserving spatial context, advancing the study of tissue architecture. However, existing computational approaches still face key limitations, particularly the insufficient exploitation of histology information and the lack of cross-modal meaningful contrastive strategies for biological analyses. To overcome these challenges, we propose GCAST, a graph contrastive autoencoder framework for spatial transcriptomics that seamlessly integrates multimodal SRT data. GCAST adopts a self-supervised strategy to derive biologically meaningful representations directly from histology images when available. GCAST constructs dual graph views based on data augmentation and introduces a novel contrastive learning designed to leverage histology-weighted and gene-weighted features and improve biological interpretability. In addition, GCAST employs a block-diagonal graph construction to automatically align multiple datasets, achieving batch-effect correction without manual intervention. The framework not only captures spatial gene expression patterns to identify tissue domains but also adapts to datasets with or without histological images and supports the integration of multiple datasets for joint analyses. Overall, GCAST provides a unified and biologically informed framework that has the potential to facilitate deeper analyses of spatial transcriptomics.

Keywords spatially resolved transcriptomics, domain identification, batch integration, contrastive autoencoder

Introduction

Spatial organization of gene expression is essential for defining cell identity, mediating intercellular communication, and shaping tissue function. Recent spatial transcriptomics technologies (e.g. 10x Visium [1], Stereo-seq [2]) enable joint profiling of gene expression and spatial location at near-single-cell resolution. Unlike single-cell RNA sequencing, which loses spatial context during dissociation, spatially resolved transcriptomics (SRT) provides a powerful framework for dissecting tissue architecture, revealing spatial heterogeneity, and uncovering molecular mechanisms in health and disease [3–5]. A key analytical task in spatial transcriptomics is identifying spatial domains, which are contiguous tissue regions with coherent gene expression profiles that often correspond to specific structures or functions [6].

Computational methods for spatial domain identification can be broadly divided into non-spatial and spatial clustering approaches. Conventional non-spatial methods, such as KMeans [7], Louvain [8], and Mclust [9], rely solely on transcriptomic similarity and, while computationally efficient, often produce fragmented or biologically implausible clusters due to ignoring spatial context. To address this, spatially aware approaches have been developed. Graph-based deep learning methods [10, 11] represent spatial transcriptomics data as graphs, where nodes correspond to spatial spots and edges encode

Received: December 8, 2025. **Revised:** February 22, 2026. **Accepted:** April 25, 2026

spatial or transcriptomic similarity. Specifically, STLearn [12] integrates histological features with gene expression, while STAGATE [13] uses a graph attention autoencoder to adaptively aggregate neighbor information. GraphST [14] and SpaNCMG [15] employ self-supervised graph learning to preserve local and global structures. SEDR [16] uses a variational graph autoencoder (VGAE) to jointly model gene expression and spatial graphs, and stAA [17] adopts an adversarial graph learning framework based on the Wasserstein distance to learn latent representations. More recently, STAIG [18] captures complex tissue structures by integrating spatial information via adaptive graph attention. Notably, CellCharter [19] and BANKSY [20] offer complementary frameworks for downstream analysis of spatial domains by modeling cellular niches and transcriptomic microenvironments, respectively.

Despite their success, current methods still have several limitations. First, histological information is often underutilized: raw patch features extracted from pretrained models may fail to capture biologically meaningful patterns [12, 21, 22]. Additionally, many graph-based self-supervised methods rely solely on spatial information, which may introduce false negatives and reduce biological interpretability. Finally, batch-effect correction and multi-dataset integration remain insufficiently addressed, often requiring manual alignment or external tools [23].

To address these challenges, we introduce GCAST, a graph contrastive autoencoder that effectively integrates cross-modal spatial transcriptomics data via data augmentation. GCAST supports datasets with or without histological images. When available, images are segmented into spatially aligned patches, from which informative histological features are extracted via self-supervised learning. To obtain compact and high-quality spot embeddings, a graph augmentation strategy generates complementary views, enabling contrastive learning based on both histology-weighted and gene-weighted features and producing more discriminative representations. For multiple datasets, GCAST uses a block-diagonal matrix for batch-wise preprocessing, enabling automatic alignment and integration while preserving dataset-specific features. Through these strategies, GCAST produces biologically meaningful representations and consistently outperforms existing methods in spatial domain identification across diverse datasets.

Methods

Overview of GCAST

We propose GCAST, a synergistic framework that integrates graph-based contrastive autoencoding with cross-modal feature weighting to improve spatial domain identification (Fig. 1). The framework begins by preprocessing spatial transcriptomics data through highly variable gene (HVG) selection, expression normalization, and construction of a spatial adjacency matrix (Fig. 1(a)). In parallel, histology patches are extracted and encoded using the self-supervised DINO [24] to capture morphological context (Fig. 1(b)). A transcriptome-derived graph $G(\mathbf{X}, \mathbf{A})$ is then constructed; for multi-dataset analyses, feature and adjacency matrices are merged into a unified graph representation (Fig. 1(c)). To facilitate contrastive learning, GCAST creates an augmented view $G(\mathbf{X}', \mathbf{A}')$ via controlled perturbations to node features and edges. The model jointly reconstructs $\mathbf{X}$ and aligns the original and augmented views to learn semantically meaningful spot embeddings (Fig. 1(d)), which subsequently support downstream batch-integrated spatial domain identification (Fig. 1(e)).

Multimodal data preprocessing

The GCAST model takes as input three components derived from multimodal spatial transcriptomics data:

HVG expression matrix. To address the high dimensionality and sparsity of raw gene counts, we first normalized each spot/cell to a total count of 1e6 and identified the top 3000 HVGs using Scanpy [25]. The resulting expression matrix $\mathbf{X} \in \mathbb{R}^{C\times d}$ (with $d = 3000$) was then z-score standardized to zero mean and unit variance, where C denotes the number of spatial spots/cells.

Spatial adjacency matrix. Tissue topology was encoded using a k-nearest-neighbor graph. Euclidean distances were computed from the spatial coordinates $\mathbf{P} \in \mathbb{R}^{C\times 2}$, and each spot was connected to its top-n nearest neighbors to form the adjacency matrix $\mathbf{A} \in \mathbb{R}^{C\times C}$.

Histology patches. For datasets with histological images, patches centered at each spot were extracted with diameter $4d_{\text{full}}$, where d_{full} is the full-resolution spot diameter, providing sufficient morphological context.

Together, these components constitute the multimodal input used for GCAST training.

Graph construction for batch integration tasks

Given the preprocessed gene expression matrix $\mathbf{X}$ and the adjacency matrix $\mathbf{A}$, we constructed a spatial transcriptomics graph $G(\mathbf{X}, \mathbf{A})$. For batch integration tasks involving M datasets, each with C_i spots ($i = 1, \ldots, M$), we applied the same preprocessing steps independently to each dataset. This yielded a set of gene expression matrices $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M$, where $\mathbf{X}_i \in \mathbb{R}^{C_i\times d}$, and corresponding spatial coordinates $\mathbf{P}_i \in \mathbb{R}^{C_i\times 2}$.

We then computed an adjacency matrix $\mathbf{A}_i \in \mathbb{R}^{C_i\times C_i}$ for each dataset based on $\mathbf{P}_i$, using the same n-nearest neighbor strategy described above. To construct a unified representation, we concatenated the expression matrices along the row dimension to obtain

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix} \in \mathbb{R}^{C\times d}, \quad \text{where} C = \sum_{i=1}^{M} C_i.$$

Similarly, the adjacency matrices were assembled into a block-diagonal matrix:

$$\mathbf{A} = \text{block-diag}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_M) \in \mathbb{R}^{C\times C}.$$

This block-diagonal construction produced no connections between spots from different datasets. Thus, we obtained a spatial transcriptomics graph $G(\mathbf{X}, \mathbf{A})$ for batch integration tasks.

Extracting embeddings of histological patches

GCAST learns patch representations using the self-supervised DINO framework [24]. Student augmentations generate six local crops (5%–40%) with flipping, color jittering, grayscale, blur, and normalization, while teacher augmentations produce two global crops (40%–100%) with the same transformations plus an additional solarization step. Both networks use the ResNet-50 backbone pretrained on ImageNet [26]. The student is updated by SGD, while the teacher is updated via EMA [27]. Consistency between the teacher and student outputs is

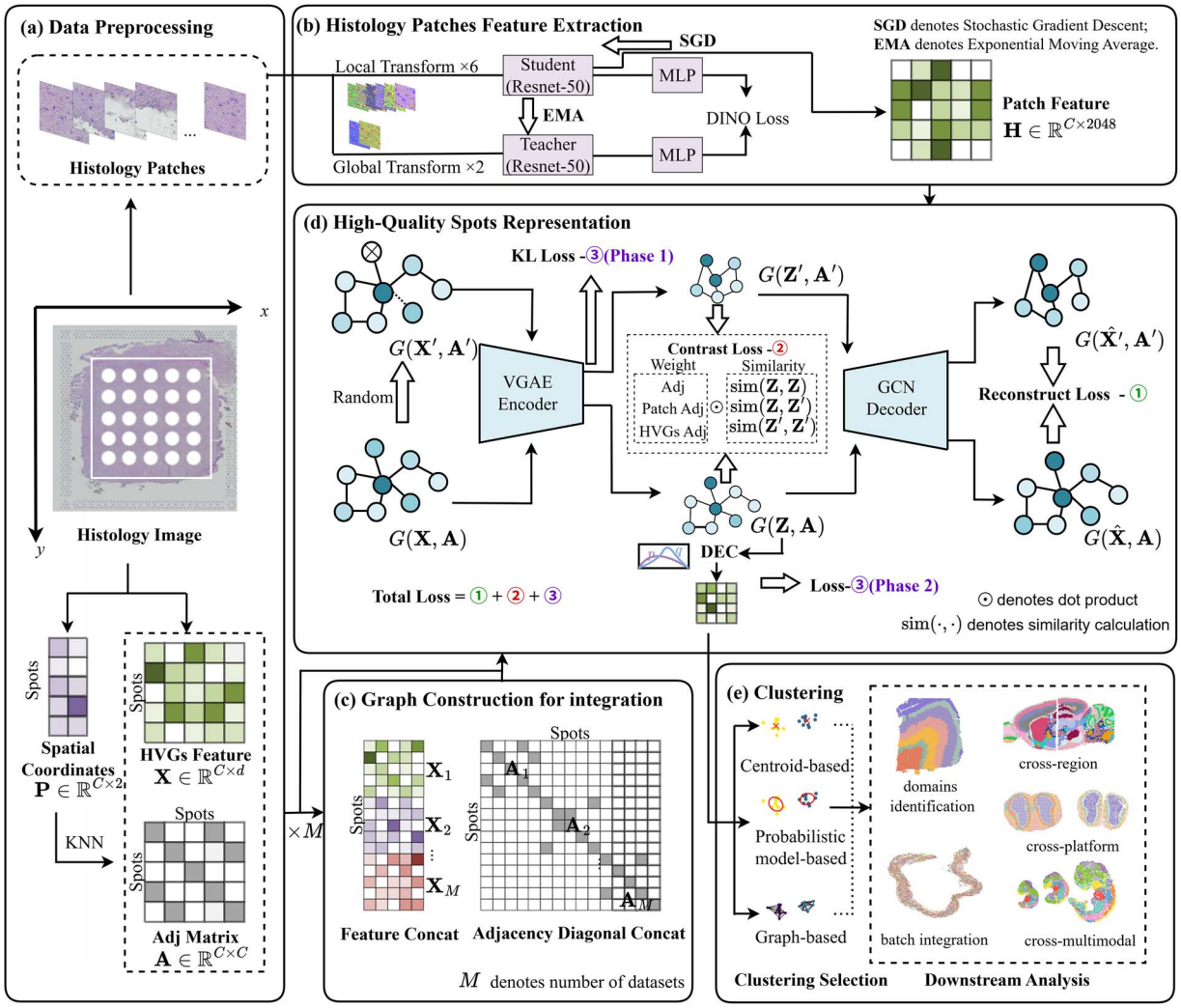

Figure 1 Overview of GCAST framework. **(a)** GCAST first preprocesses raw gene expression data, constructs the adjacency matrix from spatial coordinates, and divides histological images into patches. **(b)** Histological image patches are transformed to extract texture features via the DINO [24] framework. **(c)** For integration tasks, steps **(a)** and **(b)** are applied to each dataset; feature matrices are then concatenated and adjacency matrices are combined in a block-diagonal manner. **(d)** GCAST learns low-dimensional representations by constructing dual-view graphs and using a variational contrastive autoencoder. **(e)** The resulting embeddings can be used for downstream analysis of SRT data with various clustering methods.

optimized using the DINO loss:

$$\mathcal{L}_{\text{DINO}} = -\frac{1}{|V_t|\cdot|V_s|}\sum_{i=1}^{|V_t|}\sum_{j=1}^{|V_s|}\sum_{k=1}^{K}\sigma\left(\frac{z_{t,k}^{(i)}}{\tau_t}\right)\log\sigma\left(\frac{z_{s,k}^{(j)}}{\tau_s}\right), \tag{1}$$

where V_t and V_s denote teacher global views and student local views, respectively. Vectors $z_t^{(i)}, z_s^{(j)} \in \mathbb{R}^K$ are teacher and student outputs for the ith and jth views; τ_t and τ_s are temperature parameters; and $\sigma(\cdot)$ is the softmax over K classes.

The trained ResNet-50 encoder produces 2048D patch embeddings $\mathbf{H} \in \mathbb{R}^{C\times 2048}$, providing complementary spatial and morphological information.

High-quality spots representation

Variational encoder–decoder architecture

To the original graph $G = G(\mathbf{X}, \mathbf{A})$, we apply graph augmentation [28, 29] to generate an augmented view $G' = G(\mathbf{X}', \mathbf{A}')$ by randomly masking node features and dropping edges. These stochastic perturbations create diverse graph views that strengthen the contrastive learning process.

Then, GCAST employs a VGAE [30] with GCN layers to learn latent node embeddings for clustering. Given the original graph $G = (\mathbf{X}, \mathbf{A})$, an initial node representation is first obtained using a GCN layer:

$$\mathbf{Z}^{(1)} = \text{GCN}(\mathbf{X}, \mathbf{A}). \tag{2}$$

Based on $\mathbf{Z}^{(1)}$, the mean and variance of the latent Gaussian distribution are estimated, and the latent embedding is sampled via the reparameterization trick:

$$\begin{aligned} \boldsymbol{\mu} &= \text{GCN}_{\mu}(\mathbf{Z}^{(1)}, \mathbf{A}), \\ \log\boldsymbol{\sigma}^2 &= \text{GCN}_{\sigma}(\mathbf{Z}^{(1)}, \mathbf{A}), \\ \mathbf{z}_i &= \boldsymbol{\mu}_i + \boldsymbol{\sigma}_i \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \end{aligned} \tag{3}$$

producing latent node embeddings $\mathbf{Z} \in \mathbb{R}^{C\times d'}$ for downstream clustering, where d' denotes the latent dimension.

For the decoder, a single graph convolutional layer is used to reconstruct the original node features from the latent embeddings:

$$\hat{\mathbf{X}} = \mathrm{GCN}_{\mathrm{dec}}(\mathbf{Z}, \mathbf{A}). \tag{4}$$

To learn the intrinsic structure and feature representation of data, GCAST uses the feature reconstruction loss to measure the difference between the input feature and the reconstructed output. The reconstruction loss is defined as the mean squared error [31] between the original node features and their reconstruction from the latent embeddings:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{C}\sum_{i=1}^{C} \|\mathbf{X}_i - \hat{\mathbf{X}}_i\|_2^2, \tag{5}$$

where $\mathbf{X}_i$ is the original feature vector of node i and $\hat{\mathbf{X}}_i$ is the reconstructed feature vector produced by the GCN decoder. This loss ensures that the learned embeddings preserve the feature information of the input graph.

Then, the variational Kullback–Leibler (KL) divergence [32] can be used to encourage the learned latent distribution to be close to a standard Gaussian prior, which improves generalization and stabilizes training:

$$\mathcal{L}_{\mathrm{KL}} = -\frac{1}{2}\sum_{i=1}^{C}\sum_{j=1}^{d'}\left(1 + \log(\sigma_{ij}^2) - \mu_{ij}^2 - \sigma_{ij}^2\right), \tag{6}$$

where μ_{ij}, σ_{ij}^2 denote the mean and variance of the jth dimension of node i's latent embedding. The KL term regularizes the latent space toward the prior $p(\mathbf{Z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Similarly, by applying the same loss function strategy to the augmented graph $G' = G(\mathbf{X}', \mathbf{A}')$, we can obtain $\mathcal{L}'_{\mathrm{rec}}$ and $\mathcal{L}'_{\mathrm{KL}}$.

Contrastive learning strategy

Alongside reconstruction and variational objectives, GCAST employs a contrastive learning strategy to further enhance representation learning. Given intermediate embeddings $\mathbf{Z}$ and $\mathbf{Z}'$ from the original and augmented graph views, an InfoNCE-based contrastive loss [33] is applied over three complementary adjacency representations: the binary structural adjacency $\mathbf{A}$ (also denoted as $\mathbf{A}^{(s)}$), the gene expression-based probabilistic adjacency $\mathbf{A}^{(g)}$, and the histology-based probabilistic adjacency $\mathbf{A}^{(h)}$. Both $\mathbf{A}^{(g)}$ and $\mathbf{A}^{(h)}$ are constructed in their respective feature spaces using neighborhood sets defined by the binary adjacency $\mathbf{A}$. For nodes i and j, the probabilistic edge weight is

$$A_{ij}^{(*)} = \begin{cases} \dfrac{\exp[-\log(d_{ij}^{(*)} + \epsilon)]}{\sum_{k\in\mathcal{N}(i)} \exp[-\log(d_{ik}^{(*)} + \epsilon)]}, & j \in \mathcal{N}(i), \\ 0, & j \notin \mathcal{N}(i), \end{cases} \tag{7}$$

where $d_{ij}^{(*)}$ is the Euclidean distance in the corresponding feature space and $\mathcal{N}(i)$ is the neighbor set under $\mathbf{A}$. The superscript $(*)$ denotes the modality, where (s) refers to spatial coordinate information, (g) refers to gene expression features $\mathbf{X}$, and (h) refers to histological patch features $\mathbf{P}$, and $\epsilon = 10^{-6}$ ensures numerical stability.

To enforce consistency between $\mathbf{Z}$ and $\mathbf{Z}'$, we define a modality-specific contrastive loss for each adjacency $\mathbf{A}^{(*)}$:

$$\mathcal{L}_{\mathrm{GMMC}}^{(*)} = \sum_{i=1}^{C}\sum_{j=1}^{C} A_{ij}^{(*)}\left[\ell(\mathbf{z}_i, \mathbf{z}_j) + \ell(\mathbf{z}'_i, \mathbf{z}'_j) + 2\,\ell(\mathbf{z}_i, \mathbf{z}'_j)\right], \tag{8}$$

where the InfoNCE term [33] is

$$\ell(\mathbf{a}, \mathbf{b}) = -\log \frac{\exp(\mathrm{sim}(\mathbf{a}, \mathbf{b})/\tau)}{\sum_{k=1}^{N} \exp(\mathrm{sim}(\mathbf{a}, \mathbf{b}_k)/\tau)}. \tag{9}$$

Here $\mathrm{sim}(\cdot,\cdot)$ denotes the similarity function and τ is a temperature parameter.

The multimodal contrastive loss is the sum across all modalities:

$$\mathcal{L}_{\mathrm{contrast}} = \mathcal{L}_{\mathrm{GMMC}}^{(s)} + \mathcal{L}_{\mathrm{GMMC}}^{(g)} + \mathcal{L}_{\mathrm{GMMC}}^{(h)}. \tag{10}$$

When histology is unavailable, the term $\mathcal{L}_{\mathrm{GMMC}}^{(h)}$ is omitted. Applying the same procedure to the augmented graph G' yields the loss $\mathcal{L}'_{\mathrm{contrast}}$.

Objective function

The training process consists of two phases. In phase 1, GCAST is optimized using the reconstruction, contrastive, and KL regularization losses from the two graph views:

$$\mathcal{L}_{\mathrm{total1}} = \alpha(\mathcal{L}_{\mathrm{rec}} + \mathcal{L}'_{\mathrm{rec}}) + \beta(\mathcal{L}_{\mathrm{contrast}} + \mathcal{L}'_{\mathrm{contrast}}) + \gamma(\mathcal{L}_{\mathrm{KL}} + \mathcal{L}'_{\mathrm{KL}}), \tag{11}$$

where α, β, and γ balance the contributions of each loss component. When $\mathcal{L}_{\mathrm{total1}}$ fails to decrease for S consecutive steps, the training proceeds to phase 2. To further enhance cluster separability, GCAST incorporates the Deep Embedded Clustering (DEC) objective [34], denoted as $\mathcal{L}_{\mathrm{DEC}}$. The total loss in phase 2 becomes

$$\mathcal{L}_{\mathrm{total2}} = \alpha\mathcal{L}_{\mathrm{rec}} + \beta\mathcal{L}_{\mathrm{contrast}} + \lambda\mathcal{L}_{\mathrm{DEC}}, \tag{12}$$

where λ controls the contribution of the DEC term. Phase 2 employs the same early-stopping strategy as phase 1, and training continues until convergence, yielding the final embeddings $\mathbf{Z}$ for downstream clustering and analysis.

Experimental evaluations

Datasets

We evaluated GCAST on both annotated and unannotated spatial transcriptomics datasets. The annotated 10x Visium human DLPFC dataset [35] contains 12 slides labeled into six cortical layers and white matter, enabling quantitative assessment. Moreover, the STARmap mouse visual cortex dataset [36] is suitable for evaluating GCAST on labeled data. For unannotated analysis, we used the 10x Visium Mouse Brain Coronal dataset [37]. The DLPFC slides further support batch-integration evaluation, and cross-region integration was assessed using mouse anterior and posterior brain sections. Cross-platform robustness was examined using mouse olfactory bulb data profiled by 10x Visium and Stereo-seq. GCAST also integrates mouse embryo data across embryonic days E9.5-E11.5, demonstrating robust and consistent cross-temporal batch correction. Detailed information is provided in Supplementary Table S1.

Baselines and evaluation metrics

We compared GCAST with state-of-the-art deep learning methods (STAGATE, GraphST, DeepST, STAIG, SpaGCN, SEDR, stAA, and SpaNCMG) and machine-learning baselines such as NMF [38] and STLearn (Supplementary Notes 1.1).

For annotated datasets (e.g. DLPFC), clustering performance was quantified using adjusted rand index (ARI) [39], normalized mutual information (NMI) [40], adjusted mutual information (AMI), Purity [41], Homogeneity, Completeness, and V-measure [42] (Supplementary Notes 1.2.1). For unannotated datasets (e.g. Mouse Brain Coronal), we used the Silhouette Coefficient (SC) [43] and Davies–Bouldin (DB) index [44] (Supplementary Notes 1.2.2). To assess batch-effect correction, we employed the Inverse Local Simpson's Index (iLISI) and Cell-type LISI (cLISI) [45], which quantify technical mixing and biological identity preservation (Supplementary Notes 1.2.3). Additionally, to assess the spatial fidelity of clustering results and account for potential over-smoothing effects, we introduced spatial entropy as a supplement (Supplementary Notes 1.2.4). We further introduced the Cluster Count Match Score (CCMS) to evaluate whether the predicted cluster number aligns with biological expectations (Supplementary Notes 1.2.5).

Hardware requirements

In order to ensure the reproducibility and practical applicability of GCAST, we strongly recommend using a GPU with at least 16 GB of VRAM (e.g. NVIDIA V100 16GB/32GB, RTX 3090 24GB, or A100 40GB) to ensure stable and efficient training (Supplementary Notes 1.3). All experiments in this study were conducted using an NVIDIA Tesla V100 (32 GB VRAM).

Results

High performance in domain identification of GCAST with diverse clustering algorithms for individual SRT dataset

Performance on annotated human DLPFC data

Overall, GCAST consistently outperformed the existing methods on the human brain datasets. For each method, the best clustering result (selected from five clustering algorithms KMeans, KMedoids, Mclust, Louvain, and Leiden) was shown on each of the 12 DLPFC slides (Fig. 2(a); Supplementary Figures S1-S2); the statistics reported below are the medians computed over those best-per-slide results. GCAST achieved the highest median values on most evaluation metrics for the human brain datasets: ARI = 0.60, NMI = 0.70, AMI = 0.70, Purity = 0.81, Homogeneity = 0.72, Completeness = 0.68, and V-measure = 0.70 across the 12 DLPFC slides. By comparison, the baseline methods yielded lower median ARIs: 0.57 (STAIG), 0.52 (DeepST), 0.50 (GraphST), 0.33 (NMF), 0.56 (SEDR), 0.40 (SpaGCN), 0.54 (STAGATE), and 0.52 (STLearn), indicating comparatively reduced clustering performance. Beyond clustering accuracy, we added box plots of spatial entropy for all DLPFC slides. Although DeepST and STAIG have slightly lower median spatial entropy than GCAST (0.09, 0.10, versus 0.11), GCAST still maintains consistently low spatial entropy compared with most baselines, indicating improved local label consistency and reduced boundary diffusion.

For instance, GCAST successfully identified the spatial hierarchical structure of DLPFC on slide 151674, which not only achieved the highest performance with ARI of 0.64, NMI of 0.73, and Purity of 0.82 but also effectively recovered the hierarchical organization of cortical layers, preserving the continuity and integrity of adjacent domains. On the contrary, alternative methods tend to compromise the spatial coherence of the tissue structure, leading to varying degrees of boundary disruption, most notably between Layer_1 and Layer_2, where the separation becomes blurred or fragmented (Fig. 2(b)).

Stability evaluation across clustering algorithms

To further assess the quality of the embeddings generated by GCAST, we applied the five aforementioned clustering methods to evaluate the quality of embeddings. The y-axis reports seven evaluation metrics and the x-axis shows the corresponding score values. For each metric and compared method, five points are plotted, each representing the mean score across the 12 DLPFC slides for one specific clustering algorithm; the length of each bar reflects the mean of these five values. Across the five clustering algorithms, GCAST achieved average scores of 0.53 for ARI, 0.66 for NMI, 0.66 for AMI, 0.77 for Purity, 0.68 for Homogeneity, 0.64 for Completeness, and 0.66 for V-measure, all higher than those of the baseline methods. From the computational efficiency perspective, incorporating histological images increases runtime and GPU memory usage but yields substantially improved clustering performance, making the trade-off worthwhile for dataset DLPFC (Supplementary Table S2). In addition, the 95% confidence intervals derived from the five clustering algorithms—shown as horizontal lines—are narrower for GCAST, indicating that its embeddings are more stable across clustering methods than those of the other approaches (Fig. 2(c); Supplementary Table S3).

We further evaluated the clustering performance across all five methods for DLPFC slide 151674, where GCAST consistently outperformed the baselines, achieving an average ARI of 0.58, NMI of 0.71, and Purity of 0.78. As illustrated in Fig. 2(d) and Supplementary Figure S3, the embedding of slide 151674 produced by GCAST surpassed those of the baselines in most cases across the five clustering algorithms, providing additional evidence for the superior quality of the learned representations.

Performance on annotated mouse visual cortex data

Additionally, we evaluated GCAST on the STARmap mouse visual cortex dataset, whose structure is similar to that of DLPFC. GCAST effectively identified these spatial domains, achieving an ARI of 0.63, outperforming all baseline methods (Fig. 2(e); Supplementary Figures S4-S5). Although the differentiation between the L1 and L2/3 clusters was comparatively less distinct, which is consistent with their close biological relationship, the overall spatial clustering results exhibited a high concordance with the ground-truth annotations, demonstrating GCAST's capability to capture the hierarchical organization of cortical layers in the mouse visual cortex.

Performance on unannotated mouse brain coronal data

To evaluate GCAST on data without ground-truth, we applied it to the 10x Visium Mouse Brain Coronal dataset. GCAST effectively recovered the hierarchical organization of the coronal section, consistent with the histological image and tissue spatial arrangement (Fig. 2(f); Supplementary Figure S6). Based on the histological image and the Allen Reference Atlas [46], the hippocampal region of the mouse brain exhibits two darker band-like structures as highlighted by the black boxes, corresponding to the stratum pyramidale cell layer (SP) and the

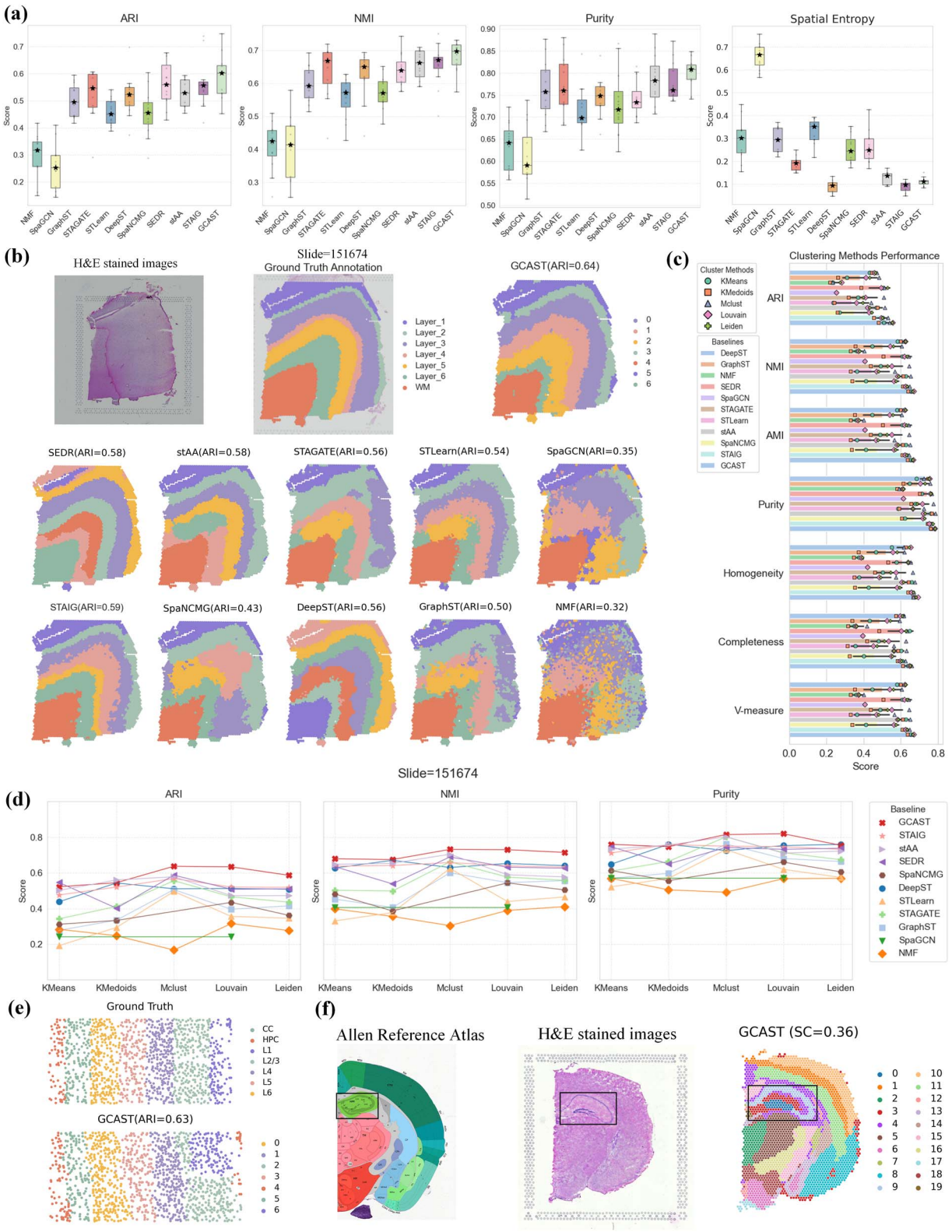

Figure 2 Spatial clustering performance comparison across 12 slides of DLPFC. **(a)** Boxplots illustrate ARI, NMI, Purity and spatial entropy scores across all 12 DLPFC slides for baseline methods and GCAST, where the central star denotes the median, box edges represent upper and lower quartiles, and whiskers extend to extreme data points within 1.5 times the interquartile range. **(b)** H&E-stained image, ground truth annotation, and spatial cluster visualizations for DLPFC slide 151674. **(c)** Bar chart with overlaid scatter points showing the average performance of five clustering algorithms applied to embeddings from GCAST and baseline methods, with confidence interval bar representing the 95% confidence interval obtained from the five clustering methods. **(d)** Line chart with overlaid scatter points showing ARI, NMI, and Purity scores across five clustering methods DLPFC slide 151674. **(e)** Ground truth annotation and spatial cluster visualization of GCAST for mouse visual cortex dataset from STARmap. **(f)** Annotations from Allen Reference Atlas, H&E-stained image and GCAST spatial cluster visualization on 10x Visium mouse brain coronal.

dentate gyrus granule cell layer (DG-GCL), respectively. Our clustering results show that cluster-3 closely resembles the stratum pyramidale cell layer, while cluster-9 aligns with the dentate gyrus granule cell layer, supporting the validity of the clustering.

Quantitatively, the SC for GCAST was 0.36 versus 0.30 for the runner-up, corresponding to a relative improvement of 20%. The DB index for GCAST was 1.0, which was lower than those of the baseline methods, indicating improved cluster compactness and separation (Supplementary Figure S7). Together, these qualitative and quantitative results indicate that GCAST produces anatomically interpretable and comparatively robust embeddings on datasets that lack manual annotations. Moreover, since this dataset is unlabeled, we select the number of cluster centers that yields the best results (Supplementary Figure S8).

GCAST achieves robust batch integration with effective spatial domain preservation

GCAST's leading performance of clustering and effective batch-effect correction for batched SRT data

To evaluate the performance of GCAST for batch-effect correction, all baseline methods were considered except SpaGCN, which does not support batch integration. DeepST leverages domain adversarial neural networks to align SRT data across multiple batches, while the remaining baseline methods adopt a simple spatial construction strategy that does not require prior alignment between datasets similar to GCAST, which ensured fairness to a certain extent.

GCAST conducted experiments on DLPFC across three subjects. Overall, GCAST achieved the highest aggregate clustering scores across all evaluation metrics, including the highest total ARI of 1.58, outperforming STAIG and DeepST across the three DLPFC subjects (Fig. 3(a); Supplementary Figure S9). This highlights GCAST's robust clustering capability for batch integration of SRT data. The only notable exception was on DLPFC-subject2, where GCAST's ARI and Purity were slightly lower than those of some competing methods. This is likely because subject2 is annotated into only five cortical layers, and the class sizes among these layers are highly imbalanced, which can reduce clustering scores even when biological structure is preserved. For batch-effect correction, GCAST achieved the best overall cLISI score (Fig. 3(b)), indicating strong preservation of cell-type integrity after integration. Its iLISI score was marginally lower than those of GraphST, STAIG, and DeepST, but remained well within the range that indicates effective batch mixing, suggesting a good balance between technical correction and biological fidelity.

GCAST's performance was especially strong on DLPFC-subject3, reaching ARI = 0.57, NMI = 0.70, and Purity = 0.78, representing relative improvements of approximately 6% in ARI, 2.5% in NMI, and 8% in Purity compared with the second-best method. Spatial visualizations further corroborate these results. When compared with the ground truth, GCAST accurately delineated most hierarchical brain structures (Fig. 3(c); Supplementary Figure S10). UMAP projections showed only minor ambiguities at the boundary between Layer_1 and Layer_2 in the ground truth, while batch-effect correction remained effective (Fig. 3(d)). The residual overlap likely reflects the biological continuity and gradual transcriptional transitions between these adjacent cortical layers, which inherently preclude a perfectly sharp separation even when using high-quality embeddings [47].

GCAST assembled domains with spatial distinctness for multiple subjects of SRT data

To assess multi-subject integration of SRT datasets, we selected two representative slides from different donors—slide 151507 (DLPFC-subject1) and slide 151673 (DLPFC-subject3)—each containing seven clusters. Analysis of cross-subject similarity in the integrated embedding indicates that our method promotes alignment across subjects while maintaining cluster identity (Fig. 3(e),(f); Supplementary Figures S11–S13). The UMAP projection shows that, although GCAST occasionally assigns spots from Layer_1 or Layer_6 to other clusters, it still preserved the distinct structural features of each slide (Fig. 3(g)), with corresponding cLISI scores of 1.11 and 1.73. Furthermore, we performed batch-integrated spatial domain identification on all 12 DLPFC slides. GCAST achieved competitive clustering performance, with an ARI of 0.42, while effectively preserving spatial domain structures (Fig. 3(h); Supplementary Figure S14). For batch integration, although GCAST obtained a relatively lower iLISI score, it outperformed all competing methods in terms of cLISI (1.30), indicating superior preservation of biological consistency across slides. Moreover, from another perspective, to investigate the differences in clustering performance between the batch-integrated and single-slide modes, GCAST generated embeddings from the SRT data of four slides per subject processed with batch integration or individually. Then, we cluster the generated embeddings under the two modes for the original slides. The results showed that GCAST was able to preserve good clustering performance on the single mode, and notably, the overall clustering performance under the batch-integrated mode even surpassed that of the single mode in subject2 (Fig. 3(i); Supplementary Figure S15). Together, these observations indicated that our approach successfully balanced cross-subject integration with the preservation of slide-specific spatial architecture.

Cross-region integration of mouse brain anterior-posterior sections

High-quality integration and reconstruction of hippocampal formation and mouse cerebral cortex

To assess the cross-region integration capability of GCAST, we applied it to mouse brain anterior and posterior sections. The spatial clustering results of GCAST showed strong concordance with anatomical annotations from the Allen Reference Atlas and histological features in H&E-stained images (Fig. 4(a)). The embeddings generated by GCAST achieved an SC of 0.31 and a DB index of 1.19, indicating well-separated and compact clusters (Fig. 4(b)). Additionally, we evaluated clustering performance under different numbers of cluster centers and selected the most appropriate setting for downstream analyses (Supplementary Figure S8).

From the spatial cluster and UMAP visualizations (Fig. 4(c); Supplementary Figure S16), GCAST accurately reconstructed the hippocampal formation (cluster-5), faithfully reflecting its characteristic curved C-shaped architecture in real 3D space. Moreover, spatial cluster visualizations show that only STAIG and GCAST recover the cross-region hippocampal formation, consistent with the original STAIG study. Furthermore, the clustering results of GCAST revealed pronounced hierarchical patterns in clusters-1, clusters-2, clusters-10, clusters-15, and clusters-24. These patterns correspond well to the laminar organization of the mouse cerebral cortex based on the Allen Reference Atlas, and were not consistently captured by the baseline methods.

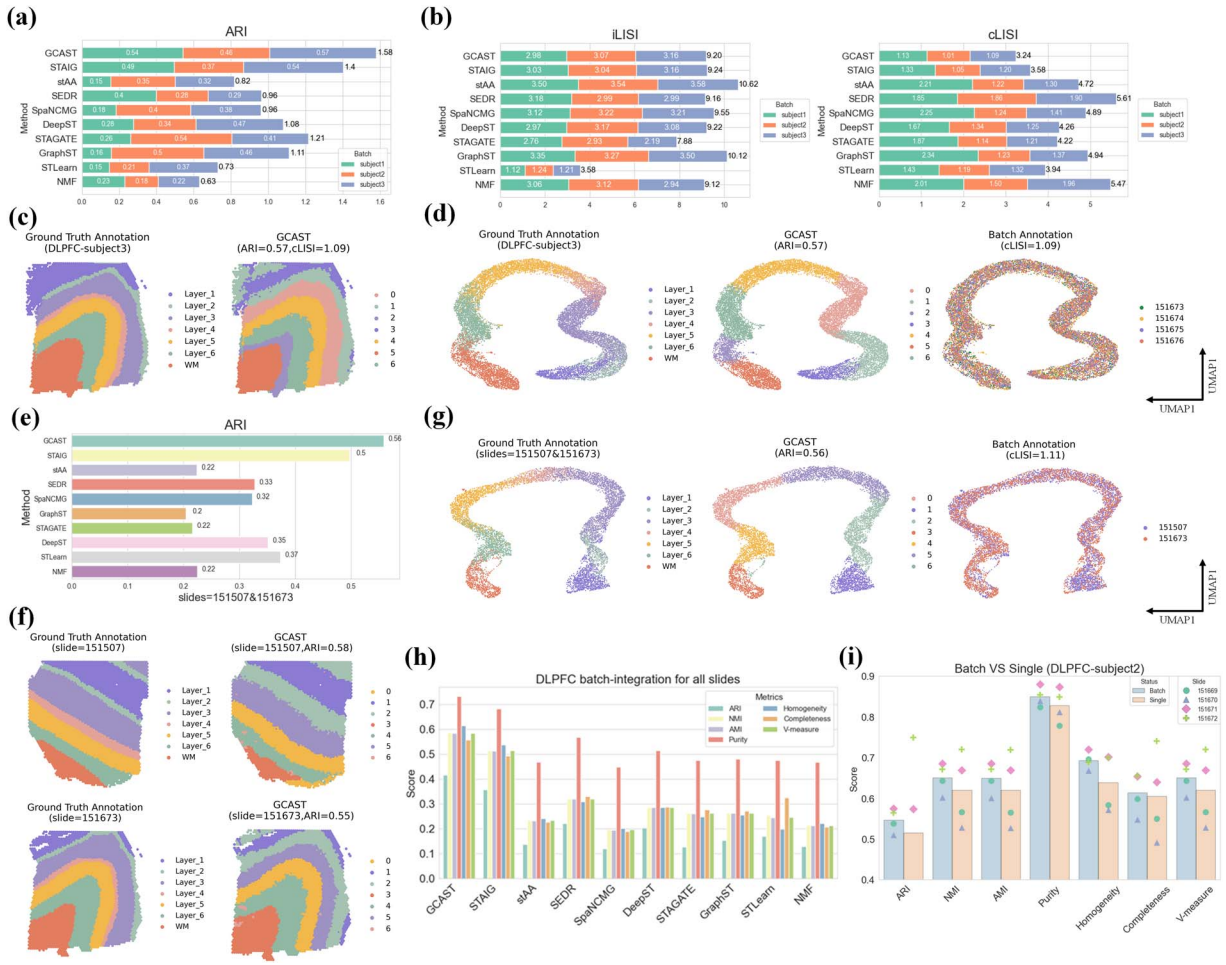

Figure 3 Batch integration of different slides for Human Dorsolateral Prefrontal Cortex (DLPFC). **(a)** Stacked-bar chart of ARI across three subjects of DLPFC. **(b)** Stacked-bar chart of iLISI and cLISI across three subjects of DLPFC. **(c)** Spatial visualizations showing ground truth annotations and clustering results of GCAST for batch-integrated DLPFC-subject3. **(d)** UMAP visualizations of ground truth annotations, batch annotations, and GCAST clustering results for batch-integrated DLPFC-subject3. **(e)** Bar chart of ARI for DLPFC slides 151507 and 151673 across all methods. **(f)** Ground truth annotations and GCAST clustering results visualized for cross-subject integration of DLPFC slide 151507 and slide 151673. **(g)** UMAP visualizations of ground truth annotations, batch annotations, and GCAST clustering results for cross-subject integration of DLPFC slide 151507 and slide 151673. **(h)** Bar plots illustrating clustering evaluation metrics for all-slides batch integration across all methods. **(i)** Bar-scatter plots comparing clustering evaluation metrics between single-processed slides and batch-integrated slides for DLPFC-subject2 integration.

Biological validation with specific marker genes of identified clusters

To biologically validate the identity of cluster-5, we performed Wilcoxon rank-sum tests to identify its marker genes. The top five markers—*Cnih2*, *Hpca*, *Camk2a*, *Cpne6*, and *Tk2b*—show well-established links to hippocampal biology (Fig. 4(d); Supplementary Table S4). In particular, *Hpca* is enriched in CA1 pyramidal neurons [48], *Camk2a* marks excitatory pyramidal cells, and *Cpne6* labels excitatory neurons in the cortex and hippocampus [49]. Further marker gene analysis shows that STAIG cluster-16 strongly overlaps with GCAST cluster-5, with 9 of the top 10 marker genes shared between the two methods (Supplementary Table S5), indicating consistent biological organization.

GCAST further identified marker genes reflecting the hierarchical organization of the mouse cerebral cortex. Cluster-2 expressed astrocyte-associated genes (*Slc1a2*, *Glul*, *Clu*, *Apoe*), corresponding to Layer_1. Cluster-10 showed high expression of excitatory neuron markers (*Camk2a*, *Camk2n1*, *Nrgn*, *Cux2*), indicative of Layer_2/3, while cluster-1 expressed neuronal differentiation markers (*Dkk3*, *Sncb*, *Mical2*, *Stmn1*, *Nrgn*), corresponding to Layer_4/5. Cluster-24 was characterized by deep-layer markers (*Tbr1*, *Slc17a7*, *Dkk3*, *Rprm*), consistent with Layer_6 [50], and cluster-15 exhibited oligodendrocyte markers (*Mbp*, *Plp1*, *Mog*, *Mag*, *Ermn*), corresponding to white matter. These results demonstrate that GCAST yields anatomically and biologically coherent clustering of the mouse brain.

Cross-platform batch integration of mouse olfactory bulb data

Integration of multi-technological olfactory bulb data recovers layered structures

To evaluate cross-platform integration, we applied GCAST to the 10x Visium and Stereo-seq mouse coronal olfactory bulb datasets, which capture similar anatomical regions despite being generated by different spatial transcriptomics technologies (Fig. 5(a)). Despite differences in sequencing technology and spot density, GCAST consistently identified matching domains across platforms, achieving a perfect CCMS score of 1 on both datasets (Fig. 5(b)). Clustering

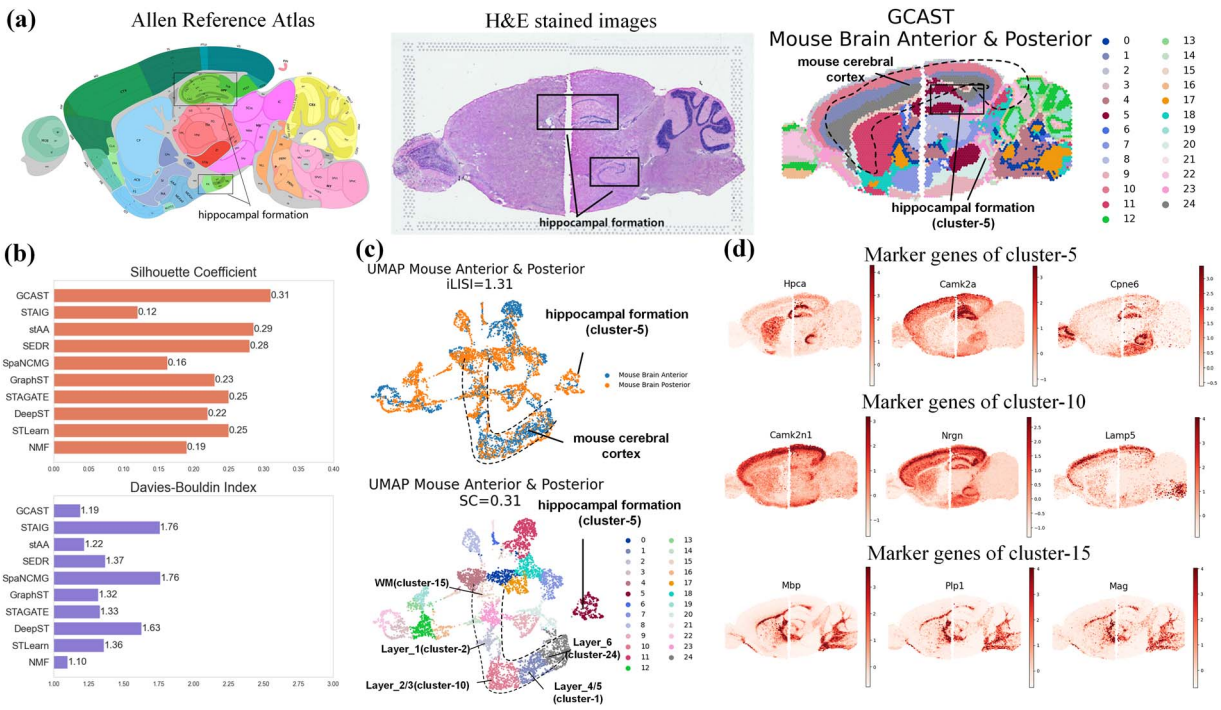

Figure 4 Performance of GCAST on cross-region integration of mouse brain anterior and posterior sections. **(a)** Allen Reference Atlas annotations, H&E-stained images of mouse brain anterior and posterior sections, and GCAST spatial visualizations for cross-region integration. **(b)** Bar plots of the SC and DB index for mouse cross-region integration. **(c)** UMAP visualizations of GCAST for mouse cross-region integration. **(d)** Visualizations of the marker genes associated with cluster-5, cluster-10, and cluster-15.

visualizations further confirmed that the hierarchical structure was preserved (Fig. 5(c)). Based on the hierarchical organization of the adult mouse olfactory bulb, the identified clusters corresponded to the following layers: ONL (cluster-1), GL (cluster-2), EPL (cluster-6), MCL (cluster-4), IPL (cluster-3), GCL (cluster-0), and RMS (cluster-5). In contrast, baseline methods exhibited poor cross-platform consistency and often sacrificed biological resolution. For example, STAGATE maintained a reasonable hierarchy on Stereo-seq but failed to align with 10x Visium, which was partitioned into only three clusters, two of which did not correspond to Stereo-seq. Similar issues were observed for other methods (Supplementary Figure S17).

Biological validation of identified clusters with specific marker genes

To further explore the hierarchical clustering results obtained by GCAST and support our hypothesis that marker gene spatial patterns reflect the laminar structure of the mouse olfactory bulb, we identified marker genes for all clusters using the Wilcoxon rank-sum test [51] and selected the top-ranked genes as cluster-specific markers. We then reviewed related literature and summarized the correspondence between each cluster and its potential laminar assignment (Supplementary Figure S18; Supplementary Tables S6-S7), providing evidence that supports our interpretations and validates the clustering results. In particular, marker genes of cluster-2 included *Apoe*, *Nxph1*, *S100a5*, *Ptn*, and *Calb2* (Fig. 5(d)). Among them, *Nxph1* and *Calb2* are well-known markers of periglomerular cells, an inhibitory interneuron population in the glomerular layer (GL) [52]. Although cluster-2 shared several markers with cluster-1 (e.g. *Apoe*, *Ptn*, and *S100a5*), suggesting the presence of glial cells, the co-expression of *S100a5* with *Nxph1* supported an interneuron identity within the GL. These observations align with the spatial visualization results and indicate that GCAST not only recapitulates known anatomical features but also reveals biologically meaningful domain organization.

Emerging developmental domains from cross-temporal SRT data

GCAST clearly revealed clustering patterns for cross-temporal analysis of mouse embryo data

GCAST analyzed spatial transcriptomic data from mouse embryos at embryonic days 9.5 (E9.5), 10.5 (E10.5), and 11.5 (E11.5) (Fig. 6(a)), which represent key stages of early organ development [53]. These time points capture the dynamic progression of tissue differentiation and organ formation, enabling the investigation of spatial gene expression patterns.

GCAST delivers comparable clustering performance to SpaNCMG and STLearn with a strong ARI of 0.31, while demonstrating superior batch integration capabilities, as evidenced by the significantly higher iLISI scores (Fig. 6(b),(c); Supplementary Figures S19-S20). Despite the morphological and compositional changes of certain organs during different embryonic stages, GCAST is still able to robustly capture specific organ-related or function-related domains, including heart and somite derivatives.

Marker gene analysis supports biologically meaningful cluster annotations

Cluster-7 and cluster-10 identified by GCAST exhibited highly similar spatial distributions across three developmental stages (E9.5, E10.5, and E11.5), suggesting that they represent conserved tissue domains during early organogenesis. Marker gene analysis showed that cluster-7 was enriched for *Mylpf*, *Myh3*, *Acta1*, *Myog*, and *Meox1*

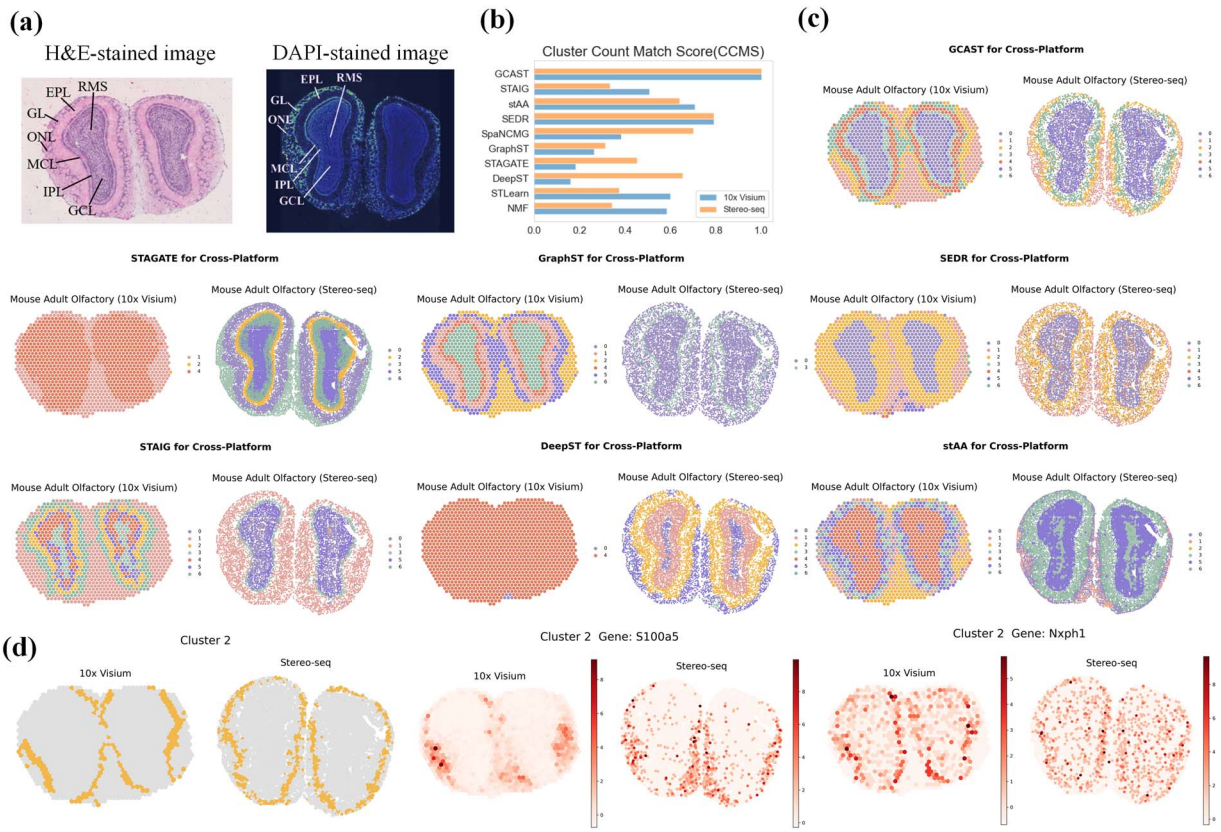

Figure 5 GCAST's performance for cross-platform integration of mouse olfactory bulb. **(a)** H&E-stained image and DAPI-stained image of mouse olfactory bulb. **(b)** CCMS for different methods across both platforms for mouse olfactory bulb. **(c)** Clustering visualizations of GCAST and baselines for mouse adult olfactory bulb data for two sequencing platforms: 10x Visium (left) and Stereo-seq (right). **(d)** Visualization results of cluster-2 and their corresponding marker genes across two sequencing platforms.

(Fig. 6(d); Supplementary Table S8), characteristic of somite-derived skeletal muscle progenitors, and was therefore annotated as somite derivatives [54]. Cluster-10 exhibited high expression of *Myl7*, *Myl3*, *Myl2*, *Tnnt2*, and *Actc1*, consistent with cardiomyocyte differentiation, and was annotated as the heart region. These results indicate that GCAST effectively captures biologically meaningful structures that are consistent across developmental stages and supported by known gene expression signatures.

Discussion

Recent advances in SRT have enabled deeper analysis of tissue organization and cellular heterogeneity, but data sparsity, technical noise, and complex spatial variation remain major challenges. To address these issues, we propose GCAST, a contrastive graph autoencoder that integrates gene expression, spatial coordinates, and histological images to generate robust low-dimensional embeddings for clustering and downstream analysis.

Model-wise, GCAST flexibly supports datasets with or without histology. For H&E images, it extracts patch features with a self-supervised DINO pipeline to provide complementary morphological context. GCAST further builds a variational dual-view graph using histology-gene weighted features: the decoder reconstructs gene expression to preserve biological information, while intermediate embeddings are contrasted against adjacency matrices derived from spatial, histological, and gene-expression similarities.

Across both labeled datasets (e.g. DLPFC) and unlabeled datasets (e.g. mouse brain coronal section), GCAST demonstrates strong and stable clustering performance, as supported by supervised and unsupervised evaluation metrics. It achieves accurate identification of spatial domains on individual datasets and remains robust under diverse batch-integration scenarios, effectively mitigating batch effects across datasets. Specifically, GCAST successfully identifies all clusters in cross-platform integration of the mouse olfactory bulb (10x Visium and Stereo-seq), accurately recovers the cross-regional organization of the hippocampal formation and the layered architecture of the mouse cerebral cortex, and uncovers consistent tissue organization patterns in cross-temporal mouse embryo datasets. Furthermore, the spatial entropy analysis across all datasets (Supplementary Figure S21) indicates that GCAST yields spatial fidelity levels essentially comparable to GNN-based models such as DeepST and STAIG, with its spatial entropy being slightly higher than these baseline methods. This reflects the core design philosophy of GCAST: its complex multi-loss architecture is primarily optimized for robust cross-dataset integration, rather than providing superior local boundary delineation. Consequently, while GCAST may not outperform simpler models in single-slice boundary sharpness, it successfully maintains comparable spatial coherence at a level broadly to existing GNN-based approaches while significantly enhancing cross-dataset alignment capabilities. Moreover, statistical tests show that GCAST significantly outperforms most baseline methods across multiple datasets, rather than by chance (Supplementary Table S9).

To verify the effectiveness of each module in GCAST, we conducted comprehensive ablation experiments on multiple datasets (Supplementary Figure S22). Overall, GCAST consistently outperforms

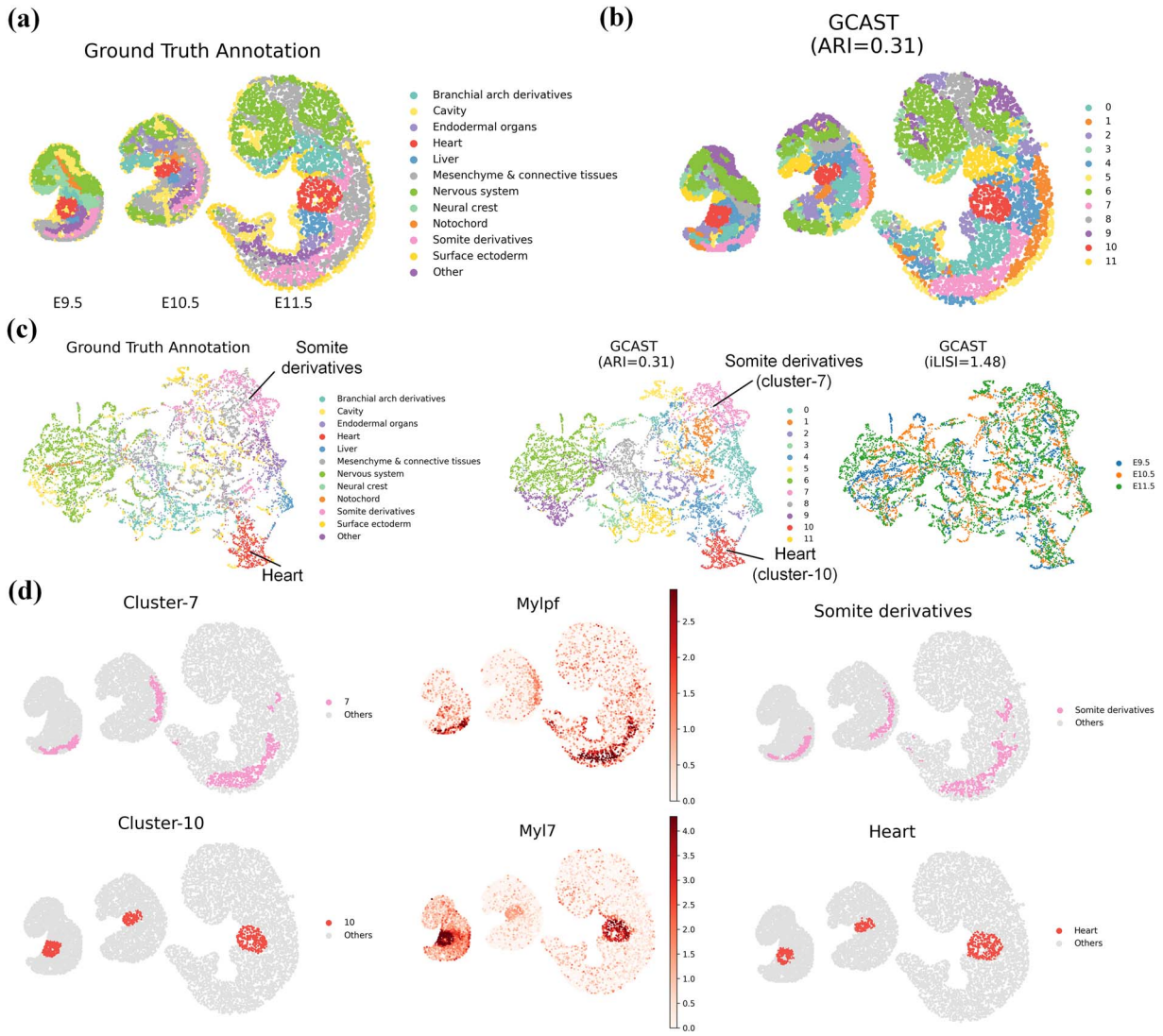

Figure 6 GCAST's capabilities for cross-temporal integration of mouse embryo datasets. **(a)** Ground truth annotations for mouse embryo datasets at E9.5, E10.5, and E11.5. **(b)** Clustering visualizations of GCAST across three time points of mouse embryo. **(c)** UMAP visualizations of ground truth annotations, time-points annotations, and GCAST clustering results for cross-temporal mouse embryo datasets. **(d)** Visualization results of cluster-7 and cluster-10 and their corresponding annotated categories and marker genes across three time points of mouse embryo.

its ablated variants, demonstrating that its performance gains arise from the meaningful contribution of each module rather than from a trivial aggregation of model components. Specifically, to ensure biological interpretability, we visualize the ablation results for mouse cross-region batch integration (Supplementary Table S10), showing that only GCAST recovers all cross-region correspondences, whereas other variants fail to capture the complete cross-region structure. Moreover, we analyze the loss weighting hyperparameters of GCAST under its two-phase training strategy (Supplementary Figure S23). The results show stable clustering performance across a wide range of loss weights in both phases, indicating low sensitivity to hyperparameter tuning, and the selected loss weight configuration is used throughout the experiments.

GCAST is a versatile and robust framework for spatial transcriptomics, producing high-quality embeddings that capture spatial structure and biological signals. Despite outperforming existing methods, its applicability to other sequencing technologies and to downstream tasks like cell–cell interaction and cancer heterogeneity requires further investigation.

Key Points

- GCAST integrates gene expression, spatial coordinates, and histological features into a unified contrastive graph autoencoder to produce biologically meaningful embeddings of spots/cells.
- It effectively captures spatial structure and tissue organization through a dual-view contrastive learning framework integrating three modalities, and employs a block-diagonal concatenation strategy to support batch-integration tasks without requiring prior spatial alignment.
- GCAST consistently achieves strong clustering performance across five clustering algorithms on DLPFC and delivers robust batch-integration results across cross-region, cross-platform, and cross-temporal spatial transcriptomics datasets.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

None declared.

Funding

This work was supported in part by the National Natural Science Foundation of China Fund under Grants (62272335 to L.Q.), in part by Jiangsu Colleges and Universities QingLan project (L.Q.), in part by University-Industry Collaborative Education Program (L.Q.), in part by Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, and in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization (L.Q., T.W., and Q.L.).

Data availability

The datasets used in this study are all publicly available: (1) the human dorsolateral prefrontal cortex (DLPFC) datasets: http://research.libd.org/spatialLIBD; (2) the mouse brain coronal section, the mouse brain anterior section, the mouse brain posterior section, and the mouse olfactory bulb of 10x Visium datasets: https://www.10xgenomics.com/datasets; (3) the mouse olfactory bulb of Stereo-seq datasets: https://github.com/JinmiaoChenLab/SEDR_analyses/tree/master/data; (4) the mouse embryo datasets: https://db.cngb.org/stomics/datasets/STDS0000058. The organized datasets can be found at: https://zenodo.org/records/18723302. More details are provided in Supplementary Table S1. Our code is available on the GitHub repository: https://github.com/ljquanlab/G-CAST/.

References

1. Rao N, Clark S, Habern O. Bridging genomics and tissue pathology: 10x genomics explores new frontiers with the visium spatial gene expression solution. *Genet Eng Biotechnol News: GEN* 2020; **40**:50–1.
2. Chen A, Liao S, Cheng M *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022; **185**:1777–1792.e21.
3. Asp M, Salmén F, Ståhl PL *et al.* Spatial detection of fetal marker genes expressed at low level in adult human heart tissue. *Sci Rep* 2017; **7**:12941.
4. Marx V. Method of the year: spatially resolved transcriptomics. *Nat Methods* 2021; **18**:9–14.
5. Eng C-HL, Lawson MJ, Zhu Q *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019; **568**:235–9.
6. Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. *Nat Methods* 2018; **15**:343–6.
7. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J, (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press; 1967. pp. 281–297.
8. Blondel VD, Guillaume J-L, Lambiotte R *et al.* Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008; **2008**:P10008.
9. Scrucca L, Fop M, Murphy TB *et al.* Mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *R J* 2016; **8**:289–317.
10. Kipf TN, Welling M. Semi-supervised classification withgraph convolutional networks. In: *Proceedings of the Fifth International Conference on Learning Representations (ICLR 2017)*. Toulon, France: OpenReview.net; 2017.
11. Chiang W-L, Liu X, Si S *et al.* Cluster-GCN: an efficient algorithm for training deep and large graph convolutional networks. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY: ACM; 2019. pp. 257–266.
12. Pham D, Tan X, Balderson B *et al.* Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat Commun* 2023; **14**:7739.
13. Dong K, Zhang S, Chen Z. Stagate: spatial transcriptomics analysis via graph attention auto-encoder. *Bioinformatics* 2022; **38**:i167–75.
14. Feng R *et al.* GraphST: unsupervised representation learning for spatial transcriptomics with graph neural networks. *Bioinformatics* 2023; **39**:btac740.
15. Si Z, Li H, Shang W *et al.* SpaNCMG: improving spatial domains identification of spatial transcriptomics using neighborhood-complementary mixed-view graph convolutional network. *Brief Bioinform.* 2024; **25**:bbae259.
16. Xu H, Fu H, Long Y *et al.* Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med* 2024; **16**:12.
17. Fang Z-Y, Liu T, Zheng R *et al.* stAA: adversarial graph autoencoder for spatial clustering task of spatially resolved transcriptomics. *Brief Bioinform* 2024; **25**:bbad500.
18. Yang Y, Cui Y, Zeng X *et al.* STAIG: spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration. *Nat Commun* 2025; **16**:1067.
19. Varrone M, Tavernari D, Santamaria-Martínez A *et al.* CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. *Nat Genet* 2024; **56**:74–84.
20. Singhal V, Chou N, Lee J *et al.* BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat Genet* 2024; **56**:431–41.
21. Jinyu H, Li X, Coleman K *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**:1342–51.
22. Gong E, Yao R, Chen D *et al.* DeepST: a deep learning-based framework for spatial transcriptomics analysis. *Brief Bioinform* 2022; **23**:bbac259.
23. Tran HTN, Ang KS, Chevrier M *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 2020; **21**:12.
24. Caron M, Touvron H, Misra I *et al.* Emerging properties in self-supervised vision transformers. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021)*. Montreal, QC, Canada: IEEE; 2021. pp. 9630–9640.
25. Alexander Wolf F, Angerer P, Theis FJ. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biol* 2018; **19**:15.
26. He K, Zhang X, Ren S *et al.* Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*. Las Vegas, NV, USA: IEEE; 2016. pp. 770–778.
27. Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: *Advances in Neural Information Processing Systems 31 (NeurIPS 2017)*. Red Hook, NY, USA: Curran Associates Inc.; 2017. pp. 1195–1204.
28. You Y, Chen T, Sui Y *et al.* Graph contrastive learning with augmentations. In: *Advances in Neural Information Processing Systems (NeurIPS*

2020). Red Hook, NY, USA: Curran Associates Inc.; 2020. pp. 5812–5823.
29. Zhu Y, Xu Y, Yu F *et al.* Graph contrastive learning with adaptive augmentation. In: Leskovec J, Grobelnik M, Najork M *et al.* (eds.), *Proceedings of the Web Conference 2021 (WWW 2021)*. Ljubljana, Slovenia. New York, NY, USA: Association for Computing Machinery; 2021. pp. 2069–2080.
30. Kingma DP, Welling M. Auto-encoding variational Bayes. In: Bengio Y, LeCun Y, (eds.), *Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014)*. Banff, AB, Canada: OpenReview.net; 2014.
31. Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2009.
32. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* 1951; **22**:79–86.
33. van den Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.
34. Guo X, Gao L, Liu X *et al.* Improved deep embedded clustering with local structure preservation. In: Sierra C, (ed.), *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization (IJCAI); 2017. pp. 1753–1759.
35. Maynard KE, Collado-Torres L, Weber LM *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 2020; **24**:425–36.
36. Wang X, Allen WE, Wright MA *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018; **361**:eaat5691.
37. Mahfouz A, Lelieveldt BPF, Grefhorst A *et al.* Genome-wide coexpression of steroid receptors in the mouse brain: identifying signaling pathways and functionally coordinated regions. *Proc Natl Acad Sci* 2016; **113**:2738–43.
38. Elosua-Bayes M, Nieto P, Mereu E *et al.* SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res* 2021; **49**:e50.
39. Hubert L, Arabie P. Comparing partitions. *J Classif* 1985; **2**:193–218.
40. Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J Mach Learn Res* 2010; **11**:2837–54.
41. Manning CD, Raghavan P, Schütze H. Cluster validity assessment: finding the optimal partitioning of a data set. In: *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press; 2008.
42. Rosenberg A, Hirschberg J. V-measure: a conditional entropy-based external cluster evaluation measure. In: Eisner J, (ed.), *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. Prague, Czech Republic: Association for Computational Linguistics; 2007. pp. 410–420.
43. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987; **20**: 53–65.
44. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979; **1**:224–227. Institute of Electrical and Electronics Engineers (IEEE), New York, NY, USA.
45. Korsunsky I, Fan J, Slowikowski K *et al.* Fast, sensitive, and accurate integration of single cell data with harmony. *Nat Methods* 2019; **16**:1289–96.
46. Seung Wook O, Harris JA, Ng L *et al.* A mesoscale connectome of the mouse brain. *Nature* 2014; **508**:207–14.
47. Lake BB, Ai R, Kaeser GE *et al.* Neuronal subtypes and diversity revealed by single-nucleus rna sequencing of the human brain. *Science* 2016; **352**:1586–90.
48. Park S-Y, Yoon SN, Kang M-J *et al.* Hippocalcin promotes neuronal differentiation and inhibits astrocytic differentiation in neural stem cells. *Stem Cell Rep* 2016; **8**:95–111.
49. Tan JZA, Jang SE, Batallas-Borja A *et al.* Copine-6 is a Ca2+ sensor for activity-induced AMPA receptor exocytosis. *Cell Reports.* 2023; **42**:113460.
50. Hevner RF, Shi L, Justice N *et al.* Tbr1 regulates differentiation of the preplate and layer 6. *Neuron* 2001; **29**:353–66.
51. Conover WJ. Practical Nonparametric Statistics. New York: John Wiley & Sons; 1980.
52. Wang I-H, Murray E, Andrews G *et al.* Spatial transcriptomic reconstruction of the mouse olfactory glomerular map suggests principles of odor processing. *Nat Neurosci* 2022; **25**: 484–92.
53. Theiler K, Westphal H. The House Mouse: Atlas of Embryonic Development. Tokyo:Springer-Verlag; 1989.
54. Scialdone A, Tanaka Y, Jawaid W *et al.* Resolving early mesoderm diversification through single cell expression profiling. *Nature* 2016; **535**:289–93.

Briefings in Bioinformatics, 2026, **27**, bbag248
https://doi.org/10.1093/bib/bbag248
Problem Solving Protocol