# TLAGC: Taylor Linear Attention-Guided Graph Convolutions for Revealing Spatial Domains in Spatial Multi-Omics Data

**Aoyun Geng, Chunyan Cui, Yunyun Su, Zhenjie Luo, Feifei Cui, Zilong Zhang***

School of Computer Science and Technology, Hainan University, Haikou, 570228, China
{aoyun, chunyancui, 23220854050043, luozhenjie1997, feifeicui, zhangzilong}@hainanu.edu.cn

## Abstract

With the rapid advance of spatial multi-omics technologies, it has become possible to simultaneously profile transcripts, proteins and chromatin states at their native spatial coordinates, thereby uncovering molecular architecture that transcends any single-omics perspective. However, the resulting data matrices are often highly sparse and suffer from unstable dimensionality. Graph-based neural methods capture only local neighborhood information, whereas conventional Transformers, although capable of modelling long-range dependencies, incur prohibitive computational costs on such data. To overcome these limitations, we propose **TLAGC** — a **T**aylor-**L**inear-**A**ttention-guided **G**raph **C**onvolutional framework that couples a Taylor-expanded linear attention (TLA) mechanism with graph convolutional networks. By eliminating the soft-max operation and linking the LocalGCN via residual connections, TLA preserves local structural information while enabling the integration of global and local contexts, thereby alleviating ineffective information propagation between spatially distant yet transcriptionally similar regions. Theoretical analysis confirms that TLA indeed reduces computational complexity, and extensive experiments on multiple spatial multi-omics benchmarks demonstrate that TLAGC consistently outperforms state-of-the-art baselines in delineating spatial domains.

**Code** — https://github.com/aythunder/TLAGC

## Introduction

With the advancement of sequencing techniques, recent progress in single-cell omics has significantly enhanced our understanding of cellular heterogeneity within cellular states and tissue structures. Traditional single-cell technologies, such as transcriptomics, proteomics, and epigenomics, enable exploration of disease mechanisms from multiple molecular perspectives. In recent years, single-cell multi-omics technologies have made it possible to simultaneously profile different cellular modalities. For instance, CITE-seq (Stoeckius et al. 2017) and REAP-seq (Peterson et al. 2017) measure both RNA expression and surface protein abundance in single cells, while SNARE-seq (Ma et al. 2020), SHARE-seq (Chen, Lake, and Zhang 2019), and 10x Multiome capture RNA expression and chromatin accessibility from the same cells. However, these methods often disrupt tissue architecture and overlook the spatial context of intercellular interactions and gene regulation within tissues.

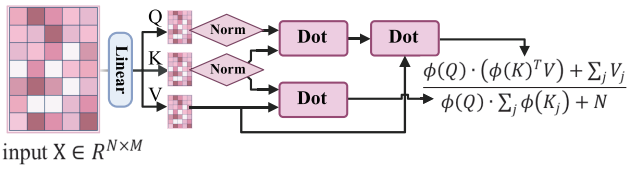


Figure 1: Derivation Diagram of TLA. Given an input matrix X with N spots and M omics features, we project X into query, key, and value matrices (Q, K, V) via linear transformations. Unlike conventional attention mechanisms, we adopt a series of normalized dot-product computations to optimize time complexity and better handle data sparsity.

Spatial transcriptomics technologies now allow the measurement of gene expression along with spatial location, offering a promising way to overcome this limitation (Sun et al. 2024). The complexity of biological systems extends beyond the transcriptional level; proteomics, epigenomics, and metabolomics also play pivotal roles in spatial regulation. The emergence of spatial multi-omics technologies, such as DBiT-seq (Liu et al. 2020), SPOTS (Ben-Chetrit et al. 2023), MISAR-seq (Jiang et al. 2023), and spatial-CITE-seq (Liu et al. 2023), enables the simultaneous capture of molecular expression across multiple omics layers. However, with the increase in data dimensions and complexity, effectively integrating and analyzing these multimodal data has become a core challenge in spatial multi-omics analysis.

One of the main challenges in spatial multi-omics integration is the problem of data sparsity. (Coleman et al. 2025). Although sequencing technologies have expanded the scope of measurable biological information, they often yield data with uneven coverage, weak signals, and limited observations, which substantially increase the complexity of data processing. Data sparsity not only increases the difficulty of analysis but also poses a significant challenge in extracting meaningful biological information from sparse observational data. Furthermore, most existing methods (Wang et al.

*Corresponding author
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2024b; Xu et al. 2024) use graph neural networks (GNNs) to process spatially constrained neighborhood graphs(Yang et al. 2025; Duan et al. 2024a). As a result, regions that share similar expression profiles but are far apart cannot communicate effectively, and GNNs tend to miss long-range global dependencies. Although graph transformers (Tang et al. 2023) can theoretically model global dependencies, they perform poorly when dealing with sparse data, often amplifying its inherent noise, and may encounter computational challenges when processing large-scale data.

To tackle these challenges, we introduce a novel spatial multi-omics integration framework called **TLAGC**. At its core is a **T**aylor **L**inear **A**ttention-guided **G**raph **C**onvolutional architecture that leverages a Taylor expansion to refine the attention mechanism and seamlessly fuses it with graph-convolution outputs via residual connections. Unlike conventional attention modules, TLA dispenses with the soft-max operation and instead employs a normalized dot-product, enabling the model to capture global dependencies with markedly reduced noise amplification (Figure 1).For clarity, we further outline the mathematical derivation of TLA. Our main contributions are as follows:

- Building on theoretical analysis, we propose Taylor Linear Attention, a lightweight kernel-based attention mechanism that replaces the soft-max operation with a first-order Taylor expansion and incorporates an L2-normalized Taylor kernel to constrain inner products within the range of [-1, 1], ensuring numerical stability while preserving contributions from weakly related nodes. This design effectively improves robustness under data sparsity and scales well to large datasets.
- We develop the TLAGC framework, which integrates TLA with GCN to jointly model global and local dependencies, thereby bridging long-range spatial gaps and enhancing biological interpretability.
- We adopt a Hybrid Spatial–Feature Graph Construction strategy to weight-fuse the spatial graph $G_s$ with the molecular feature graph $G_F$, enriching spatial context with complementary modality-specific information.
- Extensive experiments on spatial multi-omics benchmarks demonstrate that our method consistently surpasses state-of-the-art baselines and more accurately delineates functionally coherent spatial regions.

## Related Work

### Spatial Multi Omics Integration Methods

Most early approaches were designed for a single omics modality and therefore lack true cross-modal alignment (Wang, Liu, and Ma 2024; Duan et al. 2024b). SpatialMGCN (Wang et al. 2023) employs an attention-enhanced graph neural network to capture relationships among spatial domains, while GraphST (Long et al. 2023) and stDCL (Yu et al. 2025) leverage self-supervised contrastive learning for domain recognition. spaMultiVAE (Tian et al. 2024) extends the hierarchical Bayesian framework with a dependency-aware deep generative model, paving the way for spatial multi-omics applications. As spatial multi-omics technologies advance, SpatialGlue, the first dedicated method, constructs a K-nearest-neighbor graph from spatial coordinates and applies dual-attention mechanisms for cross-modal fusion (Long et al. 2024);however, it struggles to propagate information across distant but transcriptionally similar regions. SSGATE (Lv et al. 2024) introduces a bi-pathway graph attention auto-encoder strategy that accommodates both single-cell and spatial multi-omics data. COSMOS (Zhou et al. 2025) jointly embeds multiple omics layers and aggregates them via an attention module, but its integration capacity is likewise constrained by the graph radius. PRAGA (Huang et al. 2025) mitigates this limitation through a learnable spatial aggregation graph, albeit at the cost of considerable additional computation.

In summary, although existing methods can integrate spatial multi-omics data, they generally struggle with ineffective long-range spot communication and the inherent sparsity of high-dimensional data.

### Kernel/Linear Attention

To tackle the challenges that extreme data sparsity poses to attention mechanisms, several solutions have been proposed. Linformer (Wang et al. 2020) reconstructs the soft-max matrix through low-rank projection, while Nyströmformer (Xiong et al. 2021) employs landmark sampling to resample the sequence dimension—both approaches reduce complexity to near-linear. Performer (Choromanski et al. 2020) leverages the FAVOR+ scheme to kernelize dot-product attention, though its accuracy is limited by randomization error. Inspired by TaylorShift (Nauen, Palacio, and Dengel 2025), which applies Taylor expansions over tensor operations, we introduce TLA and adopt TaylorShift as the theoretical cornerstone of our framework. Existing Taylor-expanded attention variants are largely tailored to NLP (Dai et al. 2019) or computer-vision (Dass et al. 2023) tasks and overlook the sparsity inherent in spatial multi-omics data. Addressing this gap, we adapt TLA for spatial multi-omics data, enabling efficient and accurate attention computation for this domain.

## Theoretical Analysis: Taylor Linear Attention

In most spatial-omics frameworks, GNNs are the default tool for fusing cell–cell interactions with positional coordinates (Long et al. 2024). However,there may exist regions that are biologically similar yet spatially distant, and an adjacency matrix built purely from spatial proximity fails to connect them. Deepening a GNN increases the receptive field but quickly triggers over-aggregation. Appending a transformer encoder (Vaswani et al. 2017) remedies this by allowing every node to attend to every other one (as shown in Eq. (1)), but it struggles to handle sparse data effectively and incurs a quadratic cost with respect to the number of nodes.

$$Attention(Q, K, V) = softmax(QK^T)V \quad (1)$$

Inspired by TaylorShift (Nauen, Palacio, and Dengel 2025) and the extreme sparsity of omics matrices, we replace the soft-max kernel with its first-order Taylor term and obtain:

$$Attention(Q, K, V) = \frac{\phi(Q) \cdot \left(\phi(K)^\top V\right) + \sum_j V_j}{\phi(Q) \cdot \sum_j \phi(K_j) + N} \quad (2)$$

Algorithm 1: Derivation of our proposed TLA

**Input**: $X \in R^{N \times M}$

**Output**: $Attention(Q, K, V)$

1: **Step 1: Project**
Linearly map $X$ to $Q$, $K$, $V$.
2: **Add a global bias**
Apply $L_2$ normalization to obtain $\hat{Q}_i$ and $\hat{K}_i$, constraining inner-product ranges.
3: **Compute the kernel-attention numerator**
$K^\top V = \sum_{j=1}^{L} (\hat{K}_j^\top \otimes V_j), \quad A_{num} = \hat{Q} \cdot \hat{K}^\top V$
4: **Add a global bias**
$A_{num} \leftarrow A_{num} + \mathbf{1}_N \left( \sum_{j=1}^{L} V_j \right)^\top$, accelerate convergence
5: **Compute the denominator with the bias term**
$A_{num} = \hat{Q} \cdot \left( \sum_{j=1}^{L} \hat{K}_j \right) + N$, void division by zero
6: **Obtain TLA, as defined in Eq. (2)**

Where N is the spot count and $\phi(\cdot)$ is the Taylor-based kernel map. The derivation of TLA is provided in Algorithm 1.

We adopt a dot-product formulation that reflects the idea of a first-order Taylor expansion. Compared with the full attention expression, the second-order and higher-order terms are omitted:$\Delta A \approx \frac{1}{2}(q_i^\top k_j)^2$. Under L2 normalization, the approximation error is bounded by:

$$E\|\Delta A\|_F \leq \frac{1}{2}\rho\sigma^2 \tag{3}$$

Where $\rho$ denotes the correlation coefficient and $\sigma^2$ is the variance assuming a Gaussian distribution. This approximation prevents weak but biologically meaningful similarities from being excessively suppressed, thereby better accommodating the sparsity inherent in spatial omics data.

# Methodology

## Overview of TLAGC

In this paper, we propose the TLAGC architecture designed for multi-omics data integration, which initially merges data from two omics modalities and can be extended to more modalities if needed. As shown in Figure 2, the architecture centers on LocalGCN and the proposed TLA to extract local spatial neighborhood information and global dependencies for each modality, with weighted fusion via a learnable coefficient to obtain the final fused representation Z.

Our objective is to integrate the given multi-omics data with their corresponding spatial locations, aiming to ensure spatial consistency while learning a unified latent representation Z for subsequent spatial domain identification and downstream analyses. Since different omics modalities exhibit distinct characteristics and distributions, modality-specific preprocessing is required. Further preprocessing details can be found in the Appendix on GitHub.

## Hybrid Spatial–Feature Graph Construction

Assume the tissue section contains $N$ spots. For each spot $N_i$ we know its spatial coordinate $S_i = \{X_i, Y_i\}$ and a molecular feature vector $F_i^{(M)} \in R^d$, where $M$ denotes the omics layer—gene, protein, or chromatin in our data. For every omics $M$, we build an undirected $K$-nearest neighbor (KNN) graph that captures both spatial proximity and molecular similarity: $G^{(M)} = (V^{(M)}, E^{(M)}, W^{(M)})$ with nodes $V^{(M)}$, edges $E^{(M)}$, and weights $W^{(M)}$.

$$w_{ij}^{(x)} = \exp\left( -\frac{\|x_i - x_j\|_2^2}{\sigma_x^2} \right), \quad x \in \{S, F\} \tag{4}$$

$$W_{ij} = \alpha \cdot w_{ij}^{(S)} + (1 - \alpha) \cdot w_{ij}^{(F)} \tag{5}$$

Where $\sigma_x^2$ are scale parameters that normalize spatial and feature distances, respectively, and $\alpha \in [0, 1]$ balances the relative contribution of spatial and feature edge weights.

We first compute the Euclidean distance between spatial coordinates and, for each spot, select its $K_S$ nearest neighbors to construct the spatial graph. Likewise, we measure the cosine similarity between molecular feature vectors and select the $K_F$ nearest neighbors to build the feature graph. Finally, the two edge sets are merged, and each edge is assigned a hybrid weight. This construction of the adjacency matrix preserves spatial topological continuity while simultaneously capturing similarity at the molecular level.

## Encoder: Integration of Local Spatial and Global Dependencies via TLAGC

In the TLAGC module, we first use LocalGCN and TLA in sequence, and then fuse them using residual connections to simultaneously capture both local spatial neighborhood information and global dependencies. For each omics $M$, we encode the feature matrix $F^{(M)}$ and adjacency graph $G^{(M)}$ through TLAGC. In LocalGCN, we use a Hybrid Spatial–Feature Graph $G^{(M)}$ and the update at each layer in the weighted graph convolution is performed as follows:

$$H^{(l+1)} = \sigma\left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \tag{6}$$

Where $H^{(0)} = F^{(M)}$, $\tilde{A}$ is the weighted adjacency matrix of the graph (without adding extra self-loops), and $W^{(l)}$ is the learnable linear transformation. Using the adjacency matrix, we aggregate the molecular features from each spot's neighborhood through the weighted convolution operation, obtaining the local embedding $H_{loc}^{(M)}$. Subsequently, through the introduction of TLA, the local embedding is passed through equation (2) to compute the global representation $H_{glob}^{(M)}$, which captures long-range dependencies. Finally, the outputs are fused through a residual connection, yielding the low-dimensional latent representation $H^{(M)}$.

$$H^{(M)} = H_{loc}^{(M)} + H_{glob}^{(M)} \tag{7}$$

The final latent representations $H^{(M1)}$ and $H^{(M2)}$ for the two modalities are obtained from TLAGC. We fuse them using a learnable weighting coefficient $\alpha$, which is automatically adjusted during the training process, resulting in the final latent representation $Z$.

$$Z = \alpha \cdot H^{(M1)} + (1 - \alpha) \cdot H^{(M2)} \tag{8}$$

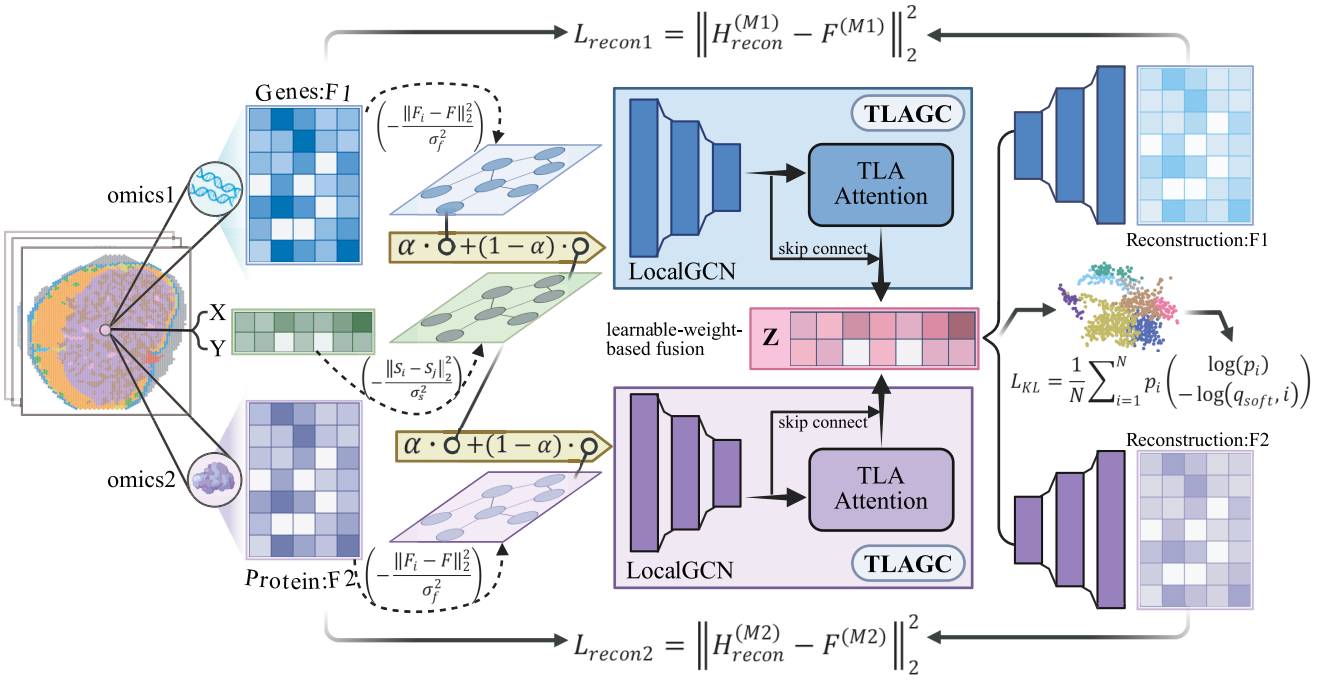


Figure 2: The overall framework of the proposed TLAGC. Each spot provides multi-omics profiles and spatial coordinates; here we illustrate two modalities. For each omics, a KNN graph is constructed from omics features and spatial positions and then fused by weighted combination. Then, the data are encoded by TLAGC, where LocalGCN and the proposed TLA jointly capture local and global dependencies, producing a latent representation per modality. These embeddings are fused by learnable weights, and a decoder reconstructs the omics data. The model is trained with reconstruction loss and KL divergence.

### Reconstruction-Regularized Clustering Objective

In multi-omics learning tasks (Ma et al. 2023), besides focusing on the similarity of latent representations, we also introduce reconstruction regularization to enhance the reconstruction capability of each modality's information (Wang et al. 2024a). Specifically, the final latent representation $Z$ is decoded into the original feature space via the LocalGCN decoder, yielding $H_{recon}^{(M)}$. The reconstruction loss is calculated using the mean squared error, as follows:

$$L_{recon} = \sum_{M} \lambda^{(M)} \left\| H_{recon}^{(M)} - F^{(M)} \right\|_2^2 \tag{9}$$

Where $H_{recon}^{(M)}$ denotes the reconstructed feature matrix for modality $M$, $F^{(M)}$ is the original input, and $\lambda^{(M)}$ is a modality-specific weighting factor that controls the contribution of the reconstruction loss for each omics modality.

To further constrain the distribution of the latent representations (Li et al. 2022), we introduce the KL divergence and map the final latent representation $Z$ to the clustering space through a new classifier $\Theta \in R^{d_z \times K}$. Using soft labels $q_{soft}$, we regularize the clustering process by calculating the KL divergence between the latent representation and the target distribution as the loss, as follows:

$$L_{KL} = \frac{1}{N} \sum_{i=1}^{N} p_i \left(\log(p_i) - \log(q_{soft}, i)\right) \tag{10}$$

Where $p$ is the predicted probability distribution for each sample obtained through the classification head.

The total loss is then:

$$L_{total} = L_{KL} + L_{recon} \tag{11}$$

## Experiments

### Experimental Setup

**Datasets** To comprehensively evaluate TLAGC on spatial multi-omics data, we employed the Human Lymph Node A1 and D1 datasets (spatial transcriptome–proteome) together with Mouse brain sections (spatial transcriptome–epigenome) (Long et al. 2024). These datasets originate from distinct sequencing platforms; detailed information is provided in the Appendix.

**Baselines** To highlight TLAGC's superiority, we compare it with three state-of-the-art algorithms designed for spatial multi-omics: (1) SpatialGlue (Long et al. 2024), the first framework for this type of data, which learns per-omics embeddings with graph neural networks and fuses them via a dual-attention mechanism; (2) PRAGA (Huang et al. 2025), which employs dynamic prototype contrastive learning to adapt to an unknown number of clusters and introduces a learnable graph that continuously updates spot adjacencies; and (3) COSMOS (Zhou et al. 2025), which integrates complementary signals across omics using a weighted k-nearest-

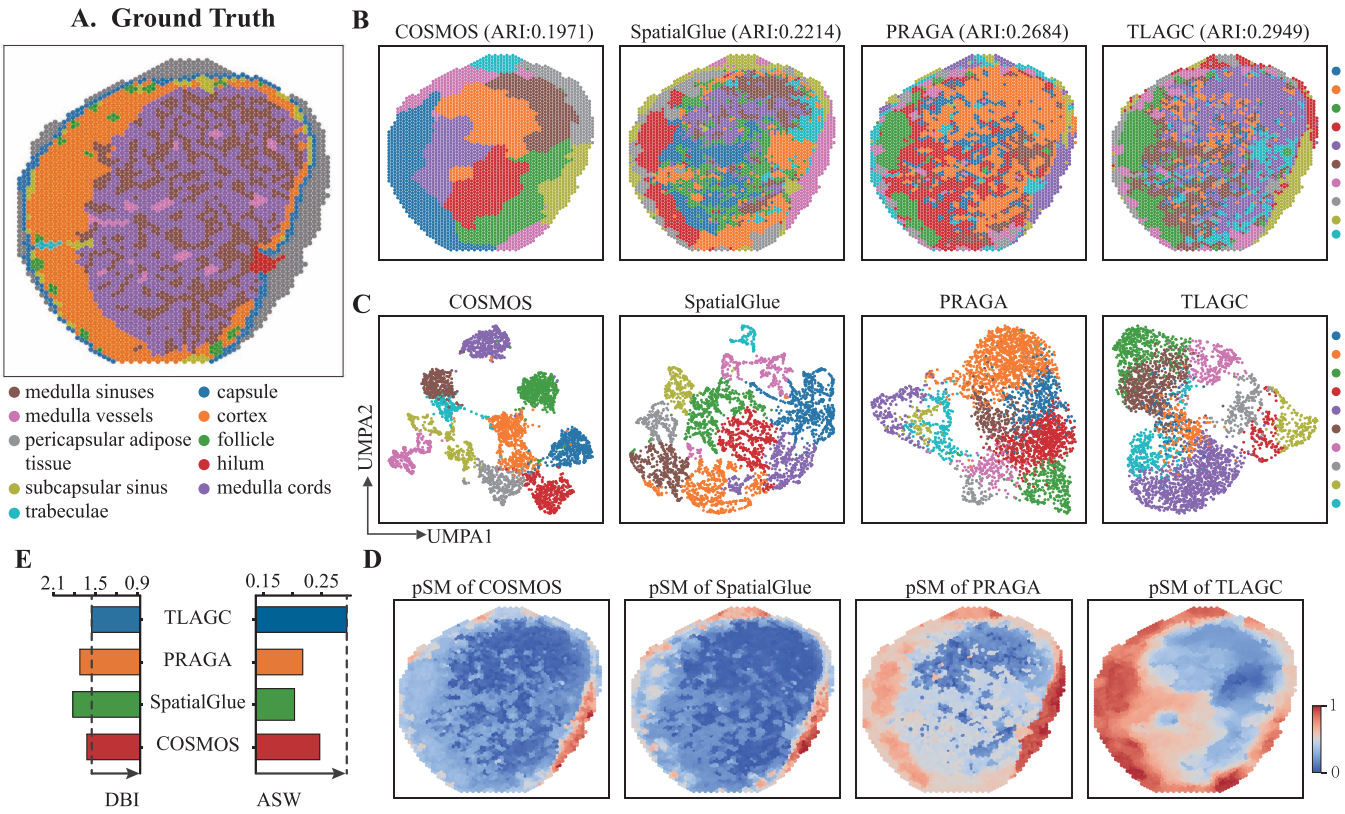


Figure 3: A. Ground truth annotation of the Human Lymph Node A1 tissue section. B. Spatial domain identification results on this section using baseline methods and our proposed approach. C. UMAP visualizations of the latent representations generated by different methods. D. Pseudo-spatiotemporal maps generated by various methods. E. Quantitative comparison based on DBI and ASW metrics.

neighbor strategy and is trained under a contrastive discriminator augmented with a spatial regularization term.

**Evaluation Metrics** We evaluate clustering performance using both internal and external metrics. Internal metrics—namely the Davies-Bouldin Index (DBI) (Davies and Bouldin 2009) and the Average Silhouette Width (ASW) (Rousseeuw 1987)—assess cluster compactness and separation without requiring ground-truth labels. External metrics, including ARI, AMI, F1 score, Homogeneity, and Completeness, quantify the agreement between predicted clusters and true labels. Further details are provided in the Appendix.

**Hyperparameters** For the Human Lymph Node A1 and D1 slices, we set the KNN graph parameters as follows: $K_S = 8$, $K_F = 8$, and $\alpha = 0.9$. The reconstruction loss weights were set to $\lambda_{r1} = \lambda_{r2} = 0.5$, and the number of clusters $K$ for soft label initialization was set to 10. For the Mouse brain slice, we configured $K_S = 10$, $K_F = 8$, and $\alpha = 0.7$. The reconstruction loss weights were set to $\lambda_{r1} = \lambda_{r2} = 0.7$, and the soft label $K$ was set to 20. The model was trained using the AdamW optimizer with a learning rate of $1 \times 10^{-3}$ and a weight decay of $1 \times 10^{-4}$. Please refer to the code for other implementation details.

## Experimental Result

**Quantitative Evaluation of TLAGC on the Human Lymph Node A1 Dataset** On the Human Lymph Node A1 slice, we benchmarked TLAGC against three leading baselines. As summarized in Table 1, TLAGC outperforms all competitors across every quantitative clustering metric. Comparison of the spatial domain segmentation in Figure 3B with the ground truth reveal that TLAGC yields the most coherent boundaries with minimal noise, surpassing SpatialGlue and PRAGA. Although COSMOS better preserves spatial coordinates, it over-merges several sub-regions—most noticeably in the medial cortex—resulting in substantial loss of local detail, whereas TLAGC remains closer to the reference segmentation.

In the UMAP projection (Figure 3C), SpatialGlue forms elongated stripes that fail to maintain local topology, and PRAGA's clusters are still loosely separated. By contrast, TLAGC produces compact, well-delineated clusters. Its Pseudo-Spatiotemporal Map (Figure 3D) displays a smooth cortex-to-medulla continuum that aligns with the lymph-node maturation trajectory; alternative methods show either overly monotonic gradients or patchy artifacts. Consistently, TLAGC achieves the lowest DBI and the highest ASW (Figure 3E), indicating superior inter-cluster separa-

| Methods | Internal metrics | | External metrics (%) | | | | |
|---|---|---|---|---|---|---|---|
| | DBI | ASW | ARI | AMI | F1 score | Homogeneity | Completeness |
| COSMOS | 1.63 | 0.197 | 19.71 | 30.49 | 31.57 | 34.59 | 28.34 |
| SpatialGlue | 1.82 | 0.154 | 22.14 | 32.11 | 33.12 | 43.60 | 33.19 |
| PRAGA | 1.72 | 0.168 | 26.84 | 38.61 | 41.80 | 41.39 | 36.21 |
| TLAGC* | **1.55** | **0.243** | **29.49** | **40.85** | **43.52** | **45.36** | **37.90** |

Table 1: Comparison of internal and external clustering metrics. The best results are highlighted in bold.

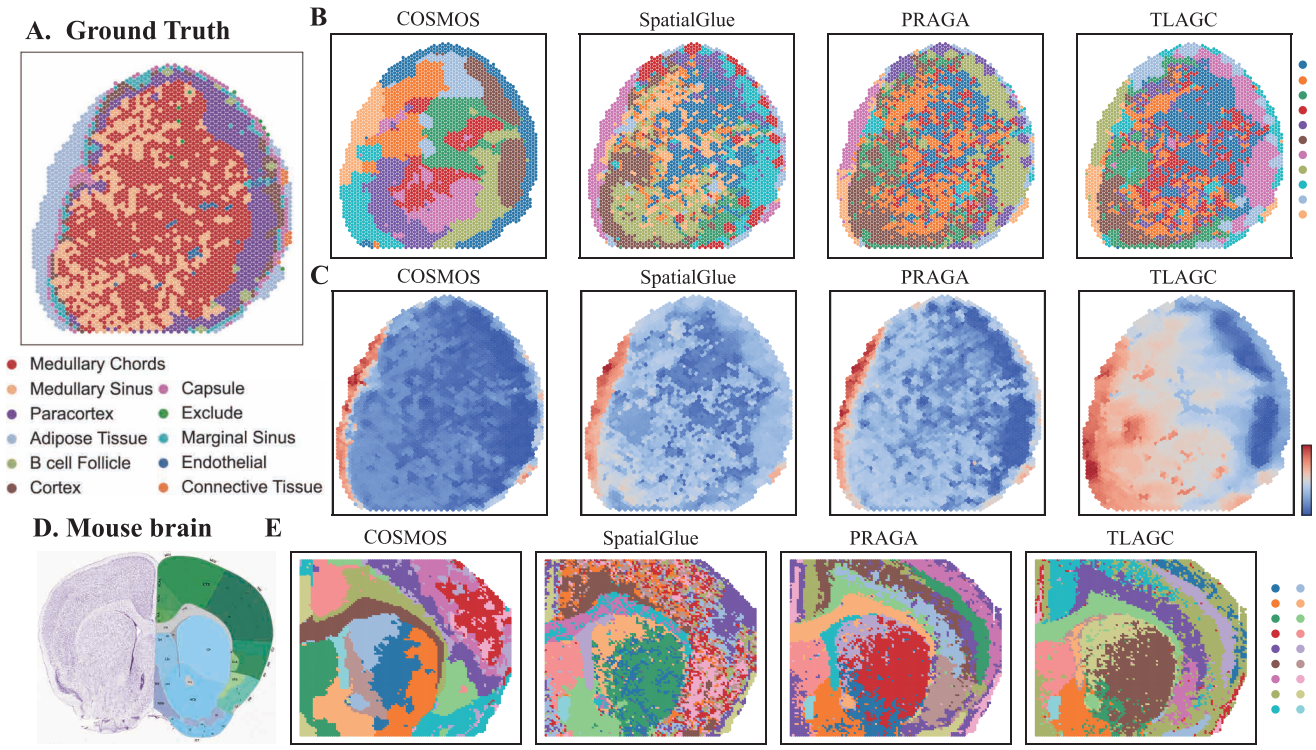


Figure 4: A. Ground truth annotation of the Human Lymph Node D1 section. B. Spatial-domain segmentation of the D1 by COSMOS, SpatialGlue, PRAGA and the proposed TLAGC. C. pSM generated by various methods. D. Ground truth annotation of the Mouse brain coronal section. E. Spatial domain identification results on this section by the four methods.

tion and boundary clarity.

These results demonstrate that the synergy between LocalGCN (capturing neighborhood structure) and TLA (modeling long-range dependencies) enables TLAGC to learn biologically coherent latent embeddings, facilitating deeper functional interpretation of tissue sections and advancing our understanding of tissue dynamics.

**Qualitative Analysis of TLAGC on the Human Lymph Node D1 and Mouse Brain Dataset** In Figure 4 we provide a qualitative comparison on the Human Lymph Node D1 slice and the Mouse brain coronal section. In Figure 4B, TLAGC markedly suppresses the spurious assignments that blur the boundary between Medullary Chords and Medullary Sinus. Although PRAGA produces a slightly cleaner morphology than the other two baselines, scattered noise is still evident when compared with TLAGC. The pseudo-spatiotemporal maps in Figure 4C further demonstrate that TLAGC traces a developmental trajectory that closely matches the biological ground truth. To assess cross-dataset generalizability, we directly transferred the trained models to a Mouse brain section; TLAGC successfully delineates deep structures—such as the thalamus and hippocampus—whereas the baselines lose fine-grained details. Taken together, these results indicate that TLAGC mitigates the long-range information-propagation limitations of graph neural networks, effectively handles the sparsity intrinsic to spatial multi-omics data, and yields a stable, interpretable latent representation Z.

**Ablation Studies Quantifying Module Contributions** To assess the contribution of each key component in TLAGC, we conducted ablation studies by individually removing the LocalGCN, TLA, the spatial KNN graph ($G_S$), the feature KNN graph ($G_F$), and the two reconstruction losses. As illustrated in Table 2, the results indicate that spatial information is indispensable for spatial-omics data: eliminating $G_S$ or the LocalGCN leads to a marked per-

| Methods | ARI | AMI | F1 score | Homogeneity |
|---|---|---|---|---|
| TLAGC* | **29.49** | **40.85** | **43.52** | **45.36** |
| w/o TLA | 27.52 | 37.37 | 41.43 | 42.99 |
| w/o GCN | 17.43 | 28.17 | 30.59 | 33.34 |
| w/o $L_{KL}$ | 25.24 | 33.27 | 36.75 | 38.27 |
| w/o $L_{recon}$ | 28.22 | 38.23 | 40.42 | 44.73 |
| w/o $G_S$ | 13.83 | 20.89 | 29.87 | 30.54 |
| w/o $G_F$ | 24.94 | 31.04 | 34.73 | 36.10 |

Table 2: Ablation Study of TLAGC on External Clustering Metrics for the Human Lymph Node A1 Section. The best results are highlighted in bold.

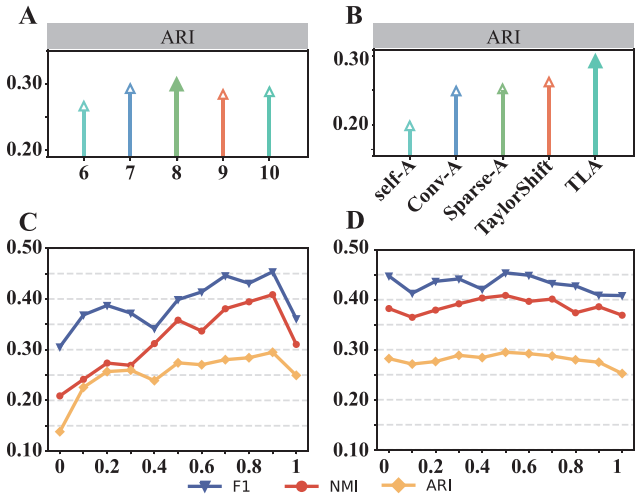


Figure 5: Parameter Sensitivity and Module Replacement Experiments. A. Sensitivity analysis of $K_S$ in the spatial graph. B.Comparison of TLA with Self-Attention, Convolutional Attention, Sparse Attention, and TaylorShift. C. Sensitivity analysis of the fusion coefficient $\alpha$ in Hybrid Spatial–Feature Graph construction. D. Evaluation of model performance under different reconstruction loss weights.

formance drop, confirming that spatial neighborhoods and local message passing are crucial for coherent boundaries and high inter-cluster separation. Removing TLA likewise degrades performance, demonstrating that long-range dependencies effectively compensate for the limited propagation of distant, same-domain signals in GNNs, while $G_F$ provides a similar complementary effect. Although the reconstruction losses exert only a modest influence on external clustering metrics, they markedly improve training convergence and local consistency. Collectively, these findings show that the synergy among all modules enables TLAGC to deliver robust and interpretable clustering even on sparse and noisy spatial multi-omics datasets.

**Comparison of TLA with Alternative Attention Mechanisms** To evaluate the advantage of TLA, we performed a set of replacement experiments in which TLA was substituted by Self-Attention, Convolutional Attention, Sparse Attention, and TaylorShift, and we compared their impact on model performance. As illustrated in Figure 5B, Convolutional Attention incurs the highest computational overhead because its 1-D convolutions inject local bias, whereas the lack of such locality explains the poorest performance of Self-Attention. In addition, we observed that replacing Self-Attention with TLA reduces GPU memory usage by 58% and shortens training time by 44%. Sparse Attention, TaylorShift, and TLA all seek to reduce time complexity in different ways, yet they are executed independently on each slice, so runtime efficiency is not our main concern; the primary focus is the validity of the clustering results. Because data sparsity can amplify noise during training, TLA outperforms TaylorShift by discarding negligible terms in its Taylor expansion, thereby suppressing noise more effectively and achieving the best overall performance. Finally, we conducted additional masking experiments on the HLN A1 dataset, where TLA exhibited a substantially smaller performance drop than the other attention mechanisms, demonstrating its stronger robustness.

**Hyper-Parameter Tuning Assessing Model Robustness** We investigated the impact of key hyper-parameters on TLAGC. Building on our ablation results that highlighted the importance of spatial neighborhoods, we tuned the number of spatial neighbors $K_S$ (Figure 5A), the fusion coefficient $\alpha$ in the Hybrid Spatial–Feature graph (Figure 5C), and the weight of the reconstruction loss $\lambda$ (Figure 5D). When $K_S$ varied between 7 and 9, performance fluctuated only slightly, indicating that the model is stable within a reasonable range and does not rely on precise fine-tuning; we therefore fixed $K_S = 8$. While molecular features compensate for sparsity caused by spatial distance, maintaining coordinate constraints remains essential for domain continuity; accordingly, $\alpha = 0.9$ yields optimal F1, NMI, and ARI, underscoring the primacy of spatial information. The reconstruction-loss weight has a modest effect on these metrics, but an excessively large $\lambda$ can over-emphasize reconstruction and weaken the discriminative power of the latent embeddings. Overall, TLAGC demonstrates strong robustness and computational reliability across a wide spectrum of parameter settings.

## Conclusion

In this paper, we propose Taylor Linear Attention-guided Graph Convolution, a unified framework for spatial multi-omics integration. TLA adopts a normalized dot-product, and our theoretical analysis shows its effectiveness in alleviating spatial omics sparsity. Through residual coupling, TLAGC captures long-range spatial interactions while preserving local neighborhood structure. We further employ a Hybrid Spatial–Feature Graph Construction strategy to integrate spatial topology with modality-specific features. Experiments across multiple tissue sections demonstrate that TLAGC yields more accurate spatial domain segmentation and biologically meaningful pseudo-spatiotemporal gradients, providing a coherent and interpretable latent representation for downstream spatial multi-omics analysis.

## Acknowledgments

The work is supported by the National Natural Science Foundation of China (No. 62261018), the Hainan Province Science and Technology Special Fund (ZDYF2024GXJS018), the Hainan Provincial Natural Science Foundation of China (324MS009) and the Graduate Research and Innovation Projects of Hainan Province (Qhys2024-232).

## References

Ben-Chetrit, N.; Niu, X.; Swett, A. D.; Sotelo, J.; Jiao, M. S.; Stewart, C. M.; Potenski, C.; Mielinis, P.; Roelli, P.; Stoeckius, M.; et al. 2023. Integration of whole transcriptome spatial profiling with protein markers. *Nature biotechnology*, 41(6): 788–793.

Chen, S.; Lake, B. B.; and Zhang, K. 2019. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature biotechnology*, 37(12): 1452–1457.

Choromanski, K.; Likhosherstov, V.; Dohan, D.; Song, X.; Gane, A.; Sarlos, T.; Hawkins, P.; Davis, J.; Mohiuddin, A.; Kaiser, L.; et al. 2020. Rethinking attention with performers. *arXiv preprint arXiv:2009.14794*.

Coleman, K.; Schroeder, A.; Loth, M.; Zhang, D.; Park, J. H.; Sung, J.-Y.; Blank, N.; Cowan, A. J.; Qian, X.; Chen, J.; et al. 2025. Resolving tissue complexity by multimodal spatial omics modeling with MISO. *Nature methods*, 22(3): 530–538.

Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q. V.; and Salakhutdinov, R. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*.

Dass, J.; Wu, S.; Shi, H.; Li, C.; Ye, Z.; Wang, Z.; and Lin, Y. 2023. Vitality: Unifying low-rank and sparse approximation for vision transformer acceleration with a linear taylor attention. In *2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 415–428. IEEE.

Davies, D. L.; and Bouldin, D. W. 2009. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2): 224–227.

Duan, H.; Zhang, Q.; Cui, F.; Zou, Q.; and Zhang, Z. 2024a. MVST: identifying spatial domains of spatial transcriptomes from multiple views using multi-view graph convolutional networks. *PLOS Computational Biology*, 20(9): e1012409.

Duan, H.; Zhang, Y.; Qiu, H.; Fu, X.; Liu, C.; Zang, X.; Xu, A.; Wu, Z.; Li, X.; Zhang, Q.; et al. 2024b. Machine learning-based prediction model for distant metastasis of breast cancer. *Computers in Biology and Medicine*, 169: 107943.

Huang, X.; Ma, Z.; Meng, D.; Liu, Y.; Ruan, S.; Sun, Q.; Zheng, X.; and Qiao, Z. 2025. PRAGA: prototype-aware graph adaptive aggregation for spatial multi-modal omics analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 326–333.

Jiang, F.; Zhou, X.; Qian, Y.; Zhu, M.; Wang, L.; Li, Z.; Shen, Q.; Wang, M.; Qu, F.; Cui, G.; et al. 2023. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nature Methods*, 20(7): 1048–1057.

Li, F.; Zhang, H.; Liu, S.; Guo, J.; Ni, L. M.; and Zhang, L. 2022. Dn-detr: Accelerate detr training by introducing query denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13619–13627.

Liu, Y.; DiStasio, M.; Su, G.; Asashima, H.; Enninful, A.; Qin, X.; Deng, Y.; Nam, J.; Gao, F.; Bordignon, P.; et al. 2023. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nature Biotechnology*, 41(10): 1405–1409.

Liu, Y.; Yang, M.; Deng, Y.; Su, G.; Enninful, A.; Guo, C. C.; Tebaldi, T.; Zhang, D.; Kim, D.; Bai, Z.; et al. 2020. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell*, 183(6): 1665–1681.

Long, Y.; Ang, K. S.; Li, M.; Chong, K. L. K.; Sethi, R.; Zhong, C.; Xu, H.; Ong, Z.; Sachaphibulkij, K.; Chen, A.; et al. 2023. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications*, 14(1): 1155.

Long, Y.; Ang, K. S.; Sethi, R.; Liao, S.; Heng, Y.; van Olst, L.; Ye, S.; Zhong, C.; Xu, H.; Zhang, D.; et al. 2024. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nature Methods*, 21(9): 1658–1667.

Lv, T.; Zhang, Y.; Liu, J.; Kang, Q.; and Liu, L. 2024. Multi-omics integration for both single-cell and spatially resolved data based on dual-path graph attention auto-encoder. *Briefings in Bioinformatics*, 25(5).

Ma, A.; Wang, X.; Li, J.; Wang, C.; Xiao, T.; Liu, Y.; Cheng, H.; Wang, J.; Li, Y.; Chang, Y.; et al. 2023. Single-cell biological network inference using a heterogeneous graph transformer. *Nature Communications*, 14(1): 964.

Ma, S.; Zhang, B.; LaFave, L. M.; Earl, A. S.; Chiang, Z.; Hu, Y.; Ding, J.; Brack, A.; Kartha, V. K.; Tay, T.; et al. 2020. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell*, 183(4): 1103–1116.

Nauen, T. C.; Palacio, S.; and Dengel, A. 2025. Taylorshift: Shifting the complexity of self-attention from squared to linear (and back) using taylor-softmax. In *International Conference on Pattern Recognition*, 1–16. Springer.

Peterson, V. M.; Zhang, K. X.; Kumar, N.; Wong, J.; Li, L.; Wilson, D. C.; Moore, R.; McClanahan, T. K.; Sadekova, S.; and Klappenbach, J. A. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nature biotechnology*, 35(10): 936–939.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20: 53–65.

Stoeckius, M.; Hafemeister, C.; Stephenson, W.; Houck-Loomis, B.; Chattopadhyay, P. K.; Swerdlow, H.; Satija, R.; and Smibert, P. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9): 865–868.

Sun, Y.; Pan, Z.; Wang, Z.; Wang, H.; Wei, L.; Cui, F.; Zou, Q.; and Zhang, Z. 2024. Single-cell transcriptome analysis reveals immune microenvironment changes and insights into

the transition from DCIS to IDC with associated prognostic genes. *Journal of translational medicine*, 22(1): 894.

Tang, Z.; Li, Z.; Hou, T.; Zhang, T.; Yang, B.; Su, J.; and Song, Q. 2023. SiGra: single-cell spatial elucidation through an image-augmented graph transformer. *Nature Communications*, 14(1): 5618.

Tian, T.; Zhang, J.; Lin, X.; Wei, Z.; and Hakonarson, H. 2024. Dependency-aware deep generative models for multitasking analysis of spatial omics data. *Nature Methods*, 21(8): 1501–1513.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, B.; Luo, J.; Liu, Y.; Shi, W.; Xiong, Z.; Shen, C.; and Long, Y. 2023. Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. *Briefings in Bioinformatics*, 24(5): bbad262.

Wang, S.; Li, B. Z.; Khabsa, M.; Fang, H.; and Ma, H. 2020. Linformer: Self-attention with linear complexity. *arXiv preprint arXiv:2006.04768*.

Wang, T.; Shu, H.; Hu, J.; Wang, Y.; Chen, J.; Peng, J.; and Shang, X. 2024a. Accurately deciphering spatial domains for spatially resolved transcriptomics with stCluster. *Briefings in Bioinformatics*, 25(4): bbae329.

Wang, T.; Zhu, H.; Zhou, Y.; Ding, W.; Ding, W.; Han, L.; and Zhang, X. 2024b. Graph attention automatic encoder based on contrastive learning for domain recognition of spatial transcriptomics. *Communications Biology*, 7(1): 1351.

Wang, Y.; Liu, Z.; and Ma, X. 2024. MNMST: topology of cell networks leverages identification of spatial domains from spatial transcriptomics data. *Genome Biology*, 25(1): 133.

Xiong, Y.; Zeng, Z.; Chakraborty, R.; Tan, M.; Fung, G.; Li, Y.; and Singh, V. 2021. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 14138–14148.

Xu, K.; Lu, Y.; Hou, S.; Liu, K.; Du, Y.; Huang, M.; Feng, H.; Wu, H.; and Sun, X. 2024. Detecting anomalous anatomic regions in spatial transcriptomics with STANDS. *Nature Communications*, 15(1): 8223.

Yang, C.; Fu, X.; Luo, Z.; Wei, L.; Li, J.; Cui, F.; Zou, Q.; Zhang, Q.; and Zhang, Z. 2025. GCNLA: Inferring Cell-Cell Interactions From Spatial Transcriptomics With Long Short-Term Memory and Graph Convolutional Networks. *IEEE Journal of Biomedical and Health Informatics*.

Yu, Z.; Yang, Y.; Chen, X.; Wong, K.-C.; Zhang, Z.; Zhao, Y.; and Li, X. 2025. Accurate Spatial Heterogeneity Dissection and Gene Regulation Interpretation for Spatial Transcriptomics using Dual Graph Contrastive Learning. *Advanced Science*, 12(3): 2410081.

Zhou, Y.; Xiao, X.; Dong, L.; Tang, C.; Xiao, G.; and Xu, L. 2025. Cooperative integration of spatially resolved multiomics data with COSMOS. *Nature communications*, 16(1): 27.