ST-LLM: Spatial Transcriptomics Embedding with Large Language Models

Zhetao Xu[1], Xiaohua Wan[1*], Le Li[1], Shuang Feng[1], Yiming Zhang[1], Fa Zhang[1*], Bin Hu[1*]

[1]School of Medical Technology, Beijing Institute of Technology, Beijing, China
3220242843@bit.edu.cn, wanxiaohua@bit.edu.cn, fengs@bit.edu.cn, zhangfa@bit.edu.cn, bh@bit.edu.cn

Abstract

Spatial transcriptomics provides unprecedented opportunities to analyze gene patterns while preserving spatial tissue architecture. However, traditional deep learning methods for spatial transcriptomics analysis face significant challenges in multi-modal data integration, spatial dependency modeling, and biological knowledge incorporation, while existing large language models lack explicit spatial modeling capabilities for transcriptomic data. So we first present a Spatial Transcriptomics Embedding with Large Language Models (ST-LLM), a novel simple and effective approach that transforms intricate spatial graph structures into structured textual representations suitable for large language models (LLMs). ST-LLM dynamically constructs graph adjacency construction using reinforcement learning paradigms to adaptively optimize spatial relationships, converts the resulting graphs into hierarchical textual descriptions with spatial context, and leverages pre-trained semantic understanding to generate high-dimensional spatial-aware representations. Comprehensive experiments on 14 datasets demonstrate that ST-LLM achieves comparable or better performance than traditional model. ST-LLM shows that LLMs embeddings provide a new simple and effective path to encoding spatial transcriptomics biological knowledge.

Introduction

Spatial transcriptomics technology, as a revolutionary biological analysis method, can measure both gene expression profiles and tissue spatial structural information. It provides unprecedented opportunities for understanding cellular functions within their native environments (Marx 2021). With the emergence of various platforms, spatial transcriptomics datasets are rapidly growing in scale and complexity, creating an urgent demand for advanced computational tools to interpret their biological significance (Stickels et al. 2021).

However, the analysis of spatial transcriptomics data faces severe challenges. First, spatial transcriptomics data exhibits multimodal characteristics, encompassing both high-dimensional gene expression information and two-dimensional spatial coordinate information, which increases the complexity of data modeling (Kleshchevnikov et al. 2022). Second, spatial proximity introduces complex spatial autocorrelation and spatial heterogeneity, rendering traditional independent and identically distributed assumptions inapplicable (Zhu, Sun, and Zhou 2021). Furthermore, biological knowledge incorporation remains difficult as existing methods lack effective mechanisms to integrate domain-specific biological prior knowledge, limiting their ability to produce meaningful interpretations (Maynard et al. 2021).

Prompt

Task: You are a spatial transcriptomics expert analyzing Visium spatial data. Please analyze the gene expression profile of the following spot.
Spatial Text:
-Spot Barcode:<spot_id>
-Expressed Genes(ordered by expression level):<genes>
-Neighbor Gene Information:<neighbor_info>
-Location_info:
--Tissue Status:<spatial_info['in_tissue']>
--Array Position:<spatial_info['array_row']>
--Image Coordinates:<spatial_info['pxl_row_in_fullres']>
Graph Text:
C[barcode_i] (gene_1, gene_2, ..., gene_K) [coords: (x,y)]
├── North[barcode_j1] (gene_1, gene_2, gene_3) [spatial_weight: α_1]
├── Northeast[barcode_j2] (gene_1, gene_2, gene_3) [spatial_weight: α_2]
└── East[barcode_jn] (gene_1, gene_2, gene_3) [spatial_weight: α_n]
├── North[barcode_k1] (gene_1, gene_2) [2nd_order, spatial_weight: α'_1]
└── East[barcode_k2] (gene_1, gene_2) [2nd_order, spatial_weight: α'_2]
Biological context: The gene expression profile includes genes associated with signaling pathways such as mitosis (COX6C), tumor suppressor functions (MIF), and energy metabolism (SLC39A6). These suggest pathways related to cell cycle regulation, growth, and survival.

Figure 1: Structured prompt template for spatial transcriptomics analysis in ST-LLM.

Traditional deep learning methods have achieved certain progress in spatial transcriptomics analysis, primarily including graph neural network-based approaches such as SpaGCN (Hu et al. 2021), STAGATE (Dong and Zhang 2022), and GraphST (Long et al. 2023), variational autoencoder frameworks like SpaceFlow (Ren et al. 2022) and SEDR (Xu et al. 2024) and diffusion model-based approaches such as stDiff (Li et al. 2024) and diffuST (Jiao et al. 2024). However, these methods face significant limitations: they heavily rely on manually designed feature engineering and lack effective integration mechanisms for biological prior knowledge. To address these limitations, general-purpose LLMs like Deepseek (Liu et al. 2024), Llama (Grattafiori et al. 2024) and Qwen (Yang et al. 2025)

*Corresponding authors

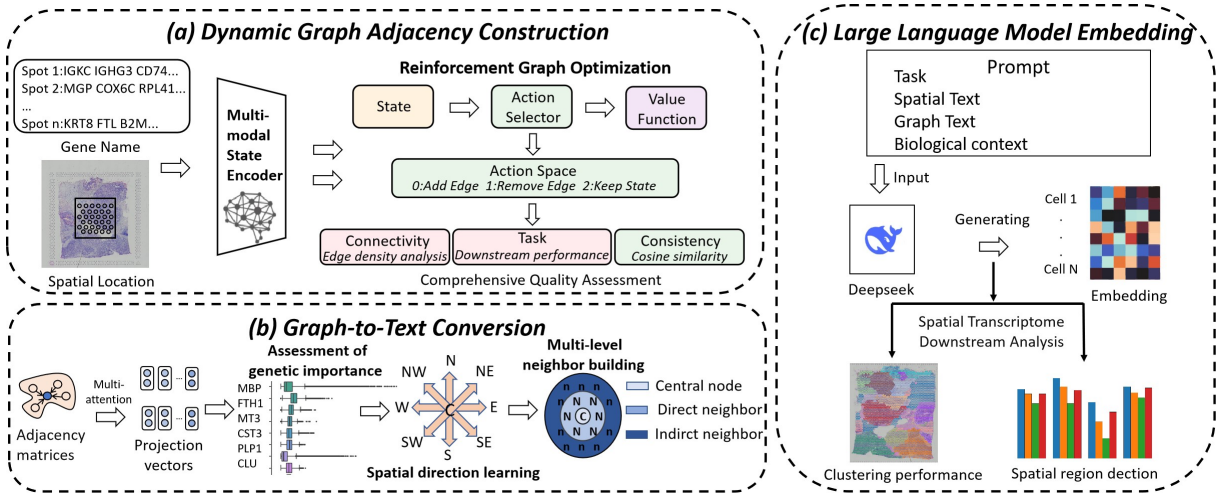

Figure 2: Framework of the Spatial Transcriptomics Embedding with Large Language Models (ST-LLM). (a) Dynamic Graph Adjacency Construction Module employs a multi-modal encoder and a reinforcement graph optimization to adaptively change edges in the spatial graph by balancing different rewards. (b) Graph-to-Text Conversion Module transforms the optimized graph into a structured prompt by Assessment of genetic importance, Spatial direction learning and Multi-level neighbor building. (c) Large Language Model Embedding Module feeds the prompt into Deepseek model to obtain spatially aware embeddings that drive downstream analyses such as spot clustering and spatial region annotation.

have emerged as a promising solution (Lin et al. 2023). In single-cell analysis, LLMs such as scGPT (Cui et al. 2024), scBERT (Yang et al. 2022) and CellLM (Zhao, Zhang, and Nie 2023) have successfully utilized pre-trained language models for cell type identification and trajectory inference. However, training a foundational large-scale model requires extensive computational resources and time. Prompt Engineering such as C2S-Scale (Rizvi et al. 2025), scMPT (Palayew, Wang, and Bader 2025) and BiomedRag (Li et al. 2025) offers a lightweight alternative to full model retraining by eliciting new behaviours from an already-trained model through carefully crafted input text. However, how to design appropriate prompt engineering strategies to convert complex spatial transcriptomics data into text representations that LLMs can understand and process has become a core challenge in this field (Liu et al. 2023).

To address the aforementioned challenges, this paper proposes a spatial transcriptomics analysis framework based on large language model embeddings, which converts complex spatial transcriptomics graph structure data into structured text descriptions understandable by LLMs. The specific contributions are as follows:

- We propose ST-LLM, the first attempt to encode spatial transcriptomics biological knowledge through LLMs embedding in our knowledge.
- We present a reinforcement graph optimization, solving the traditional problems of modeling of spatial dependencies and incorporation of biological knowledge .
- We present a graph-to-text conversion that effectively addresses the challenge of multi-modal data integration in spatial transcriptomics systems.
- Extensive downstream experiments demonstrate the superiority of ST-LLM on 14 datasets across different evaluation protocols. We present prompts comparison, LLMs generality analysis, ablation study and complexity analysis to validate the effectiveness and general of ST-LLM.

Related Work

Traditional Deep Learning Models in Spatial Transcriptomics. In recent years, deep learning methods have been extensively applied to spatial transcriptomics data analysis. Graph neural network-based approaches, such as SpaGCN (Hu et al. 2021), STAGATE (Dong and Zhang 2022), GraphST (Long et al. 2023), and stLearn (Pham et al. 2020), model spatial relationships between cells by constructing spatial adjacency graphs. Another category of methods based on deep generative models includes SpaceFlow (Ren et al. 2022), SEDR (Xu et al. 2024), SpatialDE (Svensson, Teichmann, and Stegle 2018), and SPARK-X (Zhu, Sun, and Zhou 2021), which employ variational autoencoders, generative adversarial networks, and other frameworks to learn low-dimensional representations of spatial transcriptomics data. Additionally, diffusion model-based approaches such as stDiff (Li et al. 2024) and diffuST (Jiao et al. 2024) model the complex distributions of spatial transcriptomics data through progressive denoising processes. Although these methods have achieved satisfactory performance on specific datasets, they rely on complex model architecture and lack effective integration mechanisms for biological prior knowledge, limited model representational capacity and integration of multimodal biological information. (Li et al. 2022).

LLMs in Bioinformatics. LLMs have demonstrated revolutionary application potential in bioinformatics, particularly in single-cell transcriptomics analysis (Lin et al. 2023; Rives et al. 2021; Madani et al. 2023). Recent advances,like scGPT (Cui et al. 2024), scBERT (Yang et al. 2022), CellLM

(Zhao, Zhang, and Nie 2023) and Geneformer (Theodoris et al. 2023) further explore Transformer-based architectures for gene expression modeling, gene regulatory network inference, and cell fate prediction. However, these methods are primarily designed for traditional single-cell RNA sequencing data, lacking explicit modeling of spatial information, and cannot effectively capture topological structural information. (Gayoso et al. 2022; Lopez et al. 2018). At the same time, prompt engineering, a critical technique linking pre-trained language models with downstream tasks, has recently begun to gain attention within bioinformatics (Liu et al. 2023; Qin and Eisner 2021). Models such as C2S-Scale (Rizvi et al. 2025), scMPT (Palayew, Wang, and Bader 2025) and BiomedRag (Li et al. 2025) utilize structured prompts for biomedical text understanding. Nevertheless, effective prompt strategies for spatial transcriptomics, which encode spatial transcriptomics information into textual formats comprehensible by language models, remain an open research challenge (Shin et al. 2020; Li and Liang 2021; Lester, Al-Rfou, and Constant 2021).

Method

In this section, We propose the spatial Transcriptomics Embedding with Large Language Models (ST-LLM). The core innovation lies in the incorporation of reinforcement learning paradigms into adjacency matrix, the design of learnable graph-to-text conversion and the comprehensive utilization of Deepseek's semantic understanding capabilities.The structed prompt template is shown in Figure 1 and the framework of our method is shown in Figure 2.

Dynamic Graph Adjacency Construction

The Dynamic Graph Adjacency Construction abandons traditional fixed adjacency graph construction strategies in favor of a reinforcement learning paradigm to model spatial dependencies. (Moses and Pachter 2022).

Multi-modal State Encoder. Spatial transcriptomics data encompass spatial location and gene names as two heterogeneous modalities (Chen et al. 2022; Medina-Ortiz et al. 2024). Our module processes these two types of information separately through specialized encoding branches, then employs attention mechanisms for adaptive fusion. Finally, we obtain the comprehensive feature representation for each node, which contains both spatial and gene information.

Reinforcement Graph Optimization. After obtaining the comprehensive feature representation, we reformulate the adjacency matrix construction problem as a Markov Decision Process. For each pair of nodes (i, j), the module constructs a state representation $\mathbf{z}_{ij}$ that incorporates interaction information, fusing the feature vectors of both nodes, their element-wise product, and the absolute value of feature differences. This multi-dimensional state representation comprehensively captures similarity and dissimilarity information between nodes. Subsequently, the action selector f_{action} generates an action probability distribution $\mathbf{p}_{ij}$ based on the state representation, corresponding to three possible adjacency operations: Add Edge, Remove Edge, and Keep State. During training, the module employs an ϵ-greedy strategy to balance exploration of new adjacency relationships and exploitation of known high-quality connections.

$$\mathbf{z}_{ij} = [\mathbf{h}_i; \mathbf{h}_j; \mathbf{h}_i \odot \mathbf{h}_j; |\mathbf{h}_i - \mathbf{h}_j|]$$

$$\mathbf{p}_{ij} = \text{Softmax}(f_{\text{action}}(\mathbf{z}_{ij}; \boldsymbol{\theta}_a))$$

where $\odot$ denotes element-wise multiplication, action space $\mathcal{A} = \{\text{add}, \text{remove}, \text{keep}\}$ corresponds to probability distribution $\mathbf{p}_{ij}$, and training employs ϵ-greedy strategy to balance exploration and exploitation.

To accurately assess the quality of adjacency matrix $\mathbf{A}$, the module designs a comprehensive assessment function $Q(\mathbf{A})$ that simultaneously considers edge density (Connectivity), cosine similiarity (Consistency) and downstream performance (Task). The topological quality $Q_{\text{conn}}(\mathbf{A})$ evaluates global connectivity and local density balance, where $C(\mathbf{A})$ represents the number of connected components, $\text{degree}(i)$ denotes the degree of node i, and k_{avg} is the average degree. Feature consistency $Q_{\text{consist}}(\mathbf{A})$ measures feature similarity between adjacent nodes through cosine similarity, ensuring that connected nodes are indeed close in the feature space. Task performance $Q_{\text{task}}(\mathbf{A})$ directly evaluates the constructed adjacency matrix's performance on clustering and spatial correlation tasks, where $\mathcal{L}_{\text{cluster}}$ and $\mathcal{L}_{\text{spatial}}$ represent clustering loss and spatial loss respectively.

$$Q(\mathbf{A}) = \alpha \cdot Q_{\text{conn}}(\mathbf{A}) + \beta \cdot Q_{\text{consist}}(\mathbf{A}) + \gamma \cdot Q_{\text{task}}(\mathbf{A})$$

$$Q_{\text{conn}}(\mathbf{A}) = 1 - \frac{C(\mathbf{A})}{N} + \frac{1}{N}\sum_{i=1}^{N} \min(1, \frac{\text{degree}(i)}{k_{\text{avg}}})$$

$$Q_{\text{consist}}(\mathbf{A}) = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} \frac{\mathbf{h}_i^T \mathbf{h}_j}{\|\mathbf{h}_i\|_2 \|\mathbf{h}_j\|_2} \cdot A_{ij}$$

$$Q_{\text{task}}(\mathbf{A}) = \frac{1}{1 + \mathcal{L}_{\text{cluster}}(\mathbf{A}) + \mathcal{L}_{\text{spatial}}(\mathbf{A})}$$

where Q_{conn} evaluates global connectivity and local density, Q_{consist} measures adjacent nodes' feature similarity, and Q_{task} ensures practical applicability. The dynamic graph adjacency construction loss function is:

$$\mathcal{L}_{\text{graph}} = -Q(\mathbf{A}) + \gamma \|\mathbf{A}\|_F^2$$

Graph-to-Text Conversion

While general-purpose LLMs are crucial for understanding complex biological systems, current LLMs cannot directly capture spatial transcriptomic features through textual inputs. Therefore, the Graph-to-Text Conversion transforms complex graph topological structures and high-dimensional gene information into structured textual representations comprehensible to LLMs.

The module first utilizes multi-head attention mechanisms to process graph structural information and generate projection vectors to obtain graph semantic representations.

Assessment of genetic importance. To assess the importance of gene subsets $s_g^{(i)}$, we utilize gene importance weights $w_g^{(i)}$ to evaluate and select the most discriminative genes for spatial patterns. For each gene g and node i, the

Dataset	Metric	Method								
		ST-LLM	Graph Network Approaches			Variational autoencoder			Diffusion model	
			SpaGCN	STAGATE	GraphST	SpaceFlow	SEDR	Stplus	Stdiff	DiffuST
DLPFC 151507	ARI	**0.567**	0.528	0.502	0.396	0.246	0.527	0.195	0.416	0.224
	AMI	**0.715**	0.688	0.674	0.618	0.448	0.689	0.376	0.592	0.428
	NMI	**0.719**	0.682	0.674	0.618	0.447	0.680	0.559	0.626	0.418
DLPFC 151508	ARI	**0.529**	0.488	0.528	0.491	0.289	0.466	0.265	0.477	0.423
	AMI	**0.689**	0.648	0.649	0.649	0.488	0.626	0.539	0.604	0.427
	NMI	0.646	**0.680**	0.646	0.647	0.486	0.629	0.447	0.583	0.422
DLPFC 151509	ARI	**0.589**	0.416	0.510	0.418	0.255	0.511	0.315	0.396	0.337
	AMI	**0.690**	0.623	0.641	0.607	0.470	0.663	0.447	0.594	0.438
	NMI	**0.693**	0.628	0.648	0.601	0.478	0.668	0.493	0.584	0.436
DLPFC 151510	ARI	0.513	**0.519**	0.502	0.508	0.408	0.479	0.428	0.496	0.338
	AMI	0.639	0.648	0.642	**0.653**	0.529	0.630	0.516	0.607	0.436
	NMI	**0.655**	0.633	0.647	0.634	0.529	0.638	0.516	0.615	0.436
DLPFC 151669	ARI	**0.603**	0.548	0.508	0.487	0.359	0.358	0.279	0.427	0.319
	AMI	**0.663**	0.583	0.638	0.593	0.487	0.553	0.483	0.583	0.396
	NMI	**0.662**	0.597	0.636	0.593	0.487	0.559	0.482	0.583	0.397
DLPFC 151670	ARI	0.438	0.426	0.418	**0.473**	0.288	0.373	0.385	0.408	0.419
	AMI	0.552	**0.589**	0.555	0.586	0.438	0.544	0.448	0.518	0.382
	NMI	0.558	0.572	0.569	**0.582**	0.446	0.549	0.467	0.539	0.388
DLPFC 151671	ARI	0.614	0.597	0.598	**0.616**	0.184	0.613	0.285	0.523	0.501
	AMI	**0.729**	0.698	0.624	0.727	0.418	0.703	0.483	0.593	0.519
	NMI	0.719	0.698	0.723	0.726	0.413	**0.793**	0.482	0.616	0.518
DLPFC 151672	ARI	**0.639**	0.592	0.626	0.628	0.371	0.486	0.355	0.583	0.327
	AMI	0.669	0.687	0.724	**0.729**	0.483	0.650	0.513	0.618	0.448
	NMI	0.723	0.683	**0.729**	0.668	0.482	0.659	0.473	0.588	0.482
DLPFC 151673	ARI	**0.655**	0.592	0.617	0.633	0.293	0.603	0.237	0.518	0.279
	AMI	**0.743**	0.712	0.730	0.736	0.482	0.725	0.473	0.673	0.447
	NMI	**0.743**	0.717	0.736	0.736	0.486	0.727	0.469	0.673	0.448
Mean	ARI	**0.572**	0.523	0.534	0.517	0.299	0.491	0.305	0.472	0.352
	AMI	**0.677**	0.653	0.653	0.655	0.471	0.643	0.475	0.598	0.436
	NMI	**0.680**	0.654	0.668	0.645	0.473	0.656	0.488	0.601	0.438

Table 1: Clustering performance comparison across multiple DLPFC spatial transcriptomics datasets. Performance metrics are shown for ST-LLM and baseline methods. The best score for each dataset is bolded, and the second-best score is underline.

module computes gene importance weight $w_g^{(i)}$, which combines the gene's learnable embedding vector $\mathbf{e}_g$, node features $\mathbf{h}_i$, and context features $\mathbf{c}_i$ based on neighbor information. Context features $\mathbf{c}_i$ are obtained by weighted averaging of all neighbor features of node i, with weights being the corresponding connection strengths A_{ij} in the adjacency matrix. Gene selection score $s_g^{(i)}$ comprehensively considers gene importance weights, expression levels, and spatial relevance, ultimately selecting the top K_{select} genes with highest scores to form gene subset $\mathcal{G}_i$.

$$w_g^{(i)} = \sigma\left(\mathbf{W}_g^T \text{ReLU}(\mathbf{W}_{\text{gene}}[\mathbf{e}_g; \mathbf{h}_i; \mathbf{c}_i] + \mathbf{b}_{\text{gene}}) + b_g\right)$$

$$\mathbf{c}_i = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} A_{ij} \mathbf{h}_j$$

$$s_g^{(i)} = w_g^{(i)} \cdot \log(1 + x_{ig}) \cdot \text{Spatial-Relevance}(g, i)$$

$$\mathcal{G}_i = \arg \max_{|\mathcal{G}|=K_{\text{select}}} \sum_{g \in \mathcal{G}} s_g^{(i)}$$

where $\mathbf{e}_g$ is the learnable embedding vector for gene g, $\mathbf{W}_g$, $\mathbf{W}_{\text{gene}}$ are weight matrices, $\mathbf{b}_{\text{gene}}$, b_g are bias parameters, σ is the sigmoid activation function, $\mathcal{N}(i)$ is the neighbor set of node i, and x_{ig} is the expression level of gene g at node i.

Spatial Direction Learning. To accurately capture directional patterns of spatial proximity relationships, we partition two-dimensional space into nine semantic regions, computing directions d_{ij} through angle quantization strategies. The learning of directional weights $\boldsymbol{\omega}_i$ combines global directional preference parameters $\boldsymbol{\psi}$, node features $\mathbf{h}_i$, and spatial coordinates $\mathbf{s}_i$, enabling the model to adjust directional preferences according to different biological scenarios (Ståhl et al. 2016). This design allows the module to understand and express complex spatial relationship patterns.

$$d_{ij} = \text{Quantize}\left(\arctan 2(\mathbf{s}_{j,y} - \mathbf{s}_{i,y}, \mathbf{s}_{j,x} - \mathbf{s}_{i,x})\right)$$

$$\boldsymbol{\omega}_i = \text{Softmax}(\boldsymbol{\psi} + \mathbf{W}_{\text{dir}} \mathbf{h}_i + \mathbf{W}_{\text{spatial}} \mathbf{s}_i)$$

where $\mathbf{s}_{j,y}$, $\mathbf{s}_{j,x}$ are the y and x coordinates of node j respectively, $\boldsymbol{\psi}$ is the global directional preference parameter, and $\mathbf{W}_{\text{dir}}$, $\mathbf{W}_{\text{spatial}}$ are weight matrices for directional and spatial features.

Multi-level neighbor building. In constructing neighbor information, we implement multi-level information organization strategies, constructing first-order and second-order neighbor sets while introducing adaptive quality threshold control mechanisms to ensure information quality:

$$\mathcal{N}_1(i) = \{j | A_{ij} > \tau_1, j \neq i\}$$

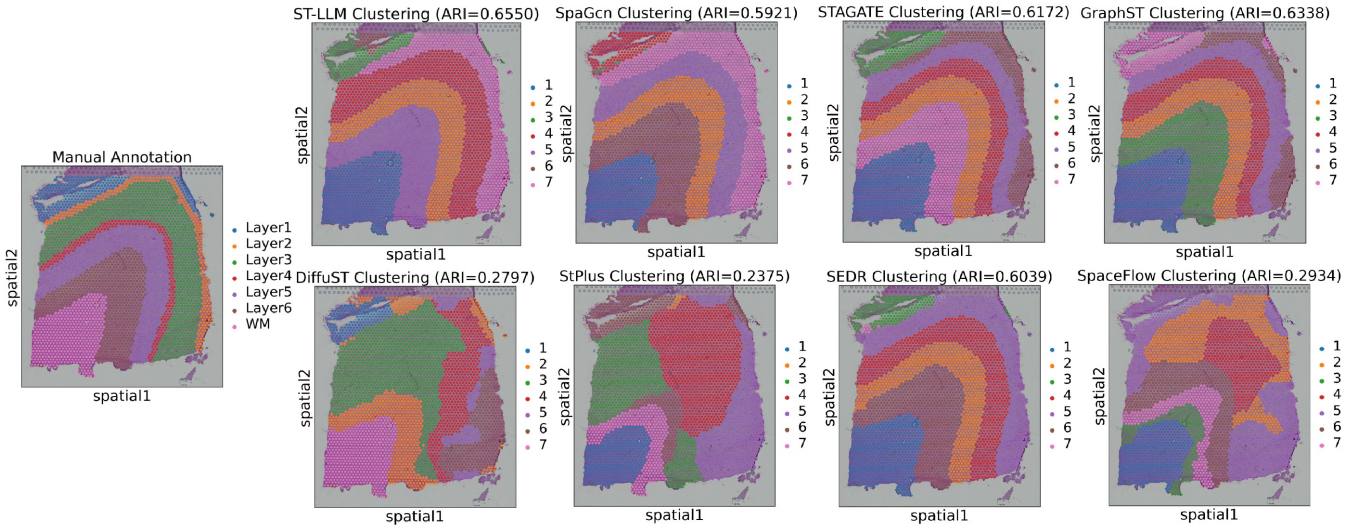

Figure 3: UMAP embeddings of spatial domain clustering results on DLPFC 151673 datasets. The visualizations illustrate the clustering performance of ST-LLM compared to baseline methods, highlighting clearer biological boundaries and reduced over-clustering phenomena achieved by ST-LLM.

Dataset	Embeddings	Accuracy	Precision	Recall
Human Breast Cancer	ST-LLM	**0.882**	**0.894**	**0.885**
	GraphST	0.843	0.856	0.841
	Stdiff	0.762	0.753	0.764
	SpaceFlow	0.841	0.834	0.842
Human Lymph Node	ST-LLM	**0.973**	**0.976**	**0.971**
	GraphST	0.905	0.914	0.903
	Stdiff	0.763	0.872	0.761
	SpaceFlow	0.874	0.902	0.873
Mouse Brain Anterior	ST-LLM	**0.774**	**0.803**	**0.772**
	GraphST	0.612	0.604	0.614
	Stdiff	0.473	0.552	0.463
	SpaceFlow	0.693	0.742	0.691
Mouse Brain Posterior	ST-LLM	0.814	0.813	0.812
	GraphST	**0.823**	**0.825**	**0.821**
	Stdiff	0.772	0.784	0.773
	SpaceFlow	0.783	0.801	0.782
Mouse Olfactory	ST-LLM	**0.804**	**0.803**	**0.802**
	GraphST	0.782	0.764	0.781
	Stdiff	0.593	0.583	0.584
	SpaceFlow	0.692	0.674	0.693

Table 2: Region Detection metrics on the test data for different datasets and embeddings.The best score for each dataset is bolded, and the second-best score is underline.

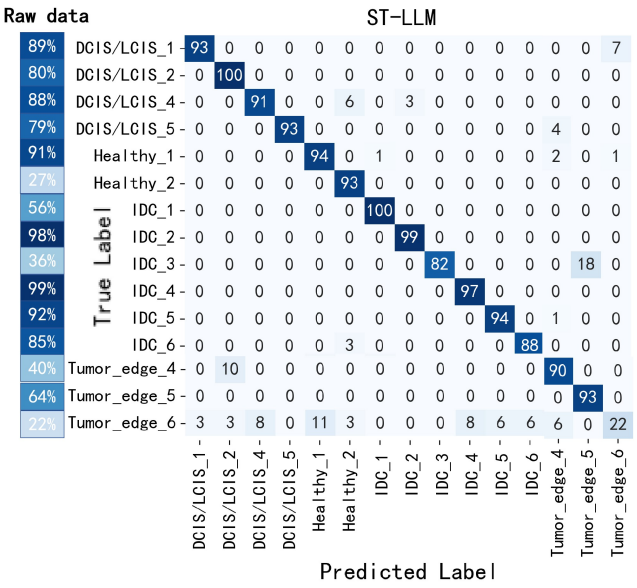

Figure 4: Confusion matrices comparing classification performance between raw data and ST-LLM on spatial transcriptomics region detection.

$$\mathcal{N}_2(i) = \bigcup_{j \in \mathcal{N}_1(i)} \{k | A_{jk} > \tau_2, k \neq i, k \notin \mathcal{N}_1(i) \cup \{j\}\}$$

$$\tau_{\text{quality}}^{(i)} = \sigma(\phi + \mathbf{w}_{\text{thresh}}^T \mathbf{h}_i + \mathbf{w}_{\text{degree}}^T \text{degree-features}(i))$$

$$\mathcal{N}_{\text{display}}(i) = \{j \in \mathcal{N}_1(i) | \omega_{i,d_{ij}} \cdot A_{ij} > \tau_{\text{quality}}^{(i)}\}$$

where τ_1, τ_2 are connection strength thresholds, ϕ is the global parameter for quality threshold, and $\mathbf{w}_{\text{thresh}}$, $\mathbf{w}_{\text{degree}}$ are weight vectors for quality threshold learning.

To optimize gene selection and directional weight learning, we design specialized loss functions. Gene selection loss functions combine L1 regularization with distributional constraints, encouraging sparse yet effective gene selection strategies. Directional consistency loss ensures spatial directional weight learning remains consistent with actual spatial patterns:

$$\mathcal{L}_{\text{gene}} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{g=1}^{G}|w_g^{(i)}|_1 + \text{KL}(\mathbf{w}^{(i)}\|\mathbf{u})\right)$$

$$\mathcal{L}_{\text{direction}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)}\|\boldsymbol{\omega}_i - \boldsymbol{\omega}_j\|_2^2 \cdot \text{similarity}(d_{ij})$$

Large Language Model Embedding

The Large Language Model Embedding transforms structured graph-text descriptions into high-dimensional vector representations with rich semantic information (Stuart and Satija 2019).

The module designs modular prompt templates incorporating four core components as shown in Figure 1. The prompt template $\text{Prompt}(i)$ integrates task description Task, spatial basic information $\text{Spatial}(i)$, graph structural information carrying hierarchical neighbor, direction weight and subordinates ordered genes $\text{Graph}(i)$, and biological context $\text{Bio}(i)$ into complete textual input through concatenation operation $\oplus$. The task description module provides clear explanations of analytical tasks, helping language models understand current analytical objectives. This design enables language models to understand complex semantics of spatial transcriptomics data from multiple perspectives.

$$\text{Prompt}(i) = \text{Task} \oplus \text{Spatial}(i) \oplus \text{Graph}(i) \oplus \text{Bio}(i)$$

where $\oplus$ denotes text concatenation operation.

Through designed prompt templates, the module guides Deepseek to generate semantically rich vector representations. The language model's embedding layer $\text{LLM}_{\text{embed}}$ maps textual input to high-dimensional vectors $\mathbf{v}_i$ through model parameters $\boldsymbol{\Theta}$, including spatial relationships, gene patterns, and biological contexts across multiple aspects. The advantage of pre-trained language models lies in their learned rich semantic representations, enabling understanding of complex textual structures and semantic associations.

$$\mathbf{v}_i = \text{LLM}_{\text{embed}}(\text{Prompt}(i); \boldsymbol{\Theta})$$

where $\boldsymbol{\Theta}$ represents the parameters of the LLM.

Experiments

Experimental Setup

- **Datasets**. We evaluate ST-LLM on 14 spatial transcriptomics datasets from 10x Genomics Visium platform including Human Dorsolateral Prefrontal Cortex (DLPFC) samples 151507-151510 and 151669-151673 (Maynard et al. 2021), Human Breast Cancer (Wu et al. 2021), Mouse Olfactory tissues (Wang et al. 2022), Human Lymph Node (Ji et al. 2020) and Mouse Brain Anterior and Posterior (Stickels et al. 2021). All datasets undergo standardized preprocessing including quality control, normalization, and highly variable gene selection.
- **Baselines**. We select three categories of state-of-the-art spatial transcriptomics analysis methods for baseline comparison such as SpaGCN, STAGATE, GraphST, SpaceFlow, SEDR, Stplus, StDiff and DiffuST. In addition to conducting primary experiments on DeepSeek, we also evaluate our model's performance across 5 general-purpose LLMs such as Granite, Llama, Sailor, Smollm and Qwen to demonstrate the generalizability of our model.
- **Evaluation Protocols**. We employ multiple evaluation metrics to assess model performance. For downstream clustering analysis, we use Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI), and Normalized Mutual Information (NMI) to quantify the consistency between clustering quality and true spatial domain annotations. For tissue region detection tasks, we adopt Accuracy, Precision and Recall to evaluate the precision of spatial domain boundary identification.

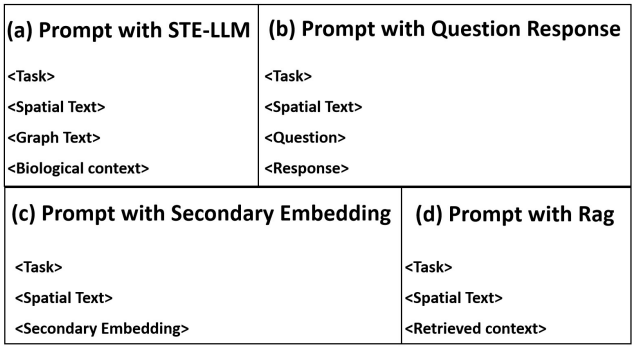

Figure 5: The schematic diagram of the four prompt methods.

Comparative Analysis of Spatial Transcriptomics Clustering Performance

In this experiment, we evaluate the clustering performance of ST-LLM and several state-of-the-art spatial transcriptomics methods across DLPFC datasets. ST-LLM demonstrates superior clustering accuracy as shown in Table 1, achieving the highest ARI and AMI scores on the majority of DLPFC slices. Specifically, on the representative section 151673, ST-LLM attains ARI, AMI, and NMI scores of 0.655, 0.743, and 0.743, respectively, significantly surpassing graph-based methods (e.g., SpaGCN, STAGATE, GraphST), as well as variational autoencoder (StPlus) and diffusion-based methods (SpaceFlow, DiffuST). On average, ST-LLM yields substantial improvements over the second-best performing baseline, with an increase of 3.7% in mean ARI scores. Further inspection of clustering results using UMAP embeddings on DLPFC 151673 are shown in Figure 3. It confirms that spatial domains identified by ST-LLM align closely with known cortical layer structures, reflecting clear and biologically meaningful boundaries. These findings collectively highlight the effectiveness of integrating LLMs embeddings with adaptive spatial graph modeling in accurately delineating complex tissue structures.

Spatial Transcriptomics Region Detection

In this section, we systematically assess the accuracy of our proposed method in annotating region detection across various spatial transcriptomics datasets, including Human

Metrics	ST_LLM	QR	SE	RAG
AMI	0.635	0.312	0.157	0.485
ACC	0.850	0.633	0.587	0.771

Table 3: Prompts Comparison results comparing evaluation of different prompt strategies on the Human Breast Cancer dataset.

Metric	Deepseek	Granite	Llama	Sailor	Smollom	Qwen
AMI	0.636	0.587	0.613	0.556	0.520	0.623
ACC	0.844	0.769	0.831	0.778	0.748	0.839

Table 4: LLMs generality analysis comparing the contributions of individual embedding components across different LLMs on the Human Breast Cancer dataset.

Breast Cancer, Mouse Olfactory, Human Lymph Node, Mouse Brain Anterior and Mouse Brain Posterior, compared to three state-of-the-art baseline methods (GraphST, Stdiff, and SpaceFlow). Quantitative results are summarized in Table 2. Our model consistently achieves superior performance across all datasets. Specifically, on the Human lymph node dataset, our model attains an accuracy of 0.973, representing a 7.2% improvement over the second-best baseline. Similarly, our method demonstrates clear advantages on the Mouse brain anterior dataset, achieving 0.77 accuracy compared to GraphST's 0.612. Figure 4 presents confusion matrices comparing raw data embeddings with our ST-LLM approach on the Human Breast Cancer dataset. Notably, our method achieves enhanced classification for DCIS/LCIS cell types (93% vs 89%) and remarkable improvement for IDC subtypes, with IDC_2 improving from 60% to 99% accuracy. These results underscore our model's capability in distinguishing complex cellular populations, laying a solid foundation for subsequent biological interpretation tasks.

Prompts Comparison

As contrasted in Figure 3, the three alternative prompts-Question Response (QR), Secondary Embedding (SE) and Rag—with our ST-LLM template; their exact token layouts are listed in Supplementary F.1. We show the main metrics AMI and ACC on the Human Breast Cancer in Table 3. ST-LLM leads the field with AMI 0.635 and ACC 0.850, demonstrating the value of encoding both spatial topology and biological context. RAG, which injects retrieved knowledge, follows at AMI 0.485 and ACC 0.771 but still lacks explicit spatial guidance. QR's label-style prompt reaches AMI 0.312 and ACC 0.633, while SE, relying solely on secondary embeddings, trails at AMI 0.157 and ACC 0.587. Other datasets and more specific metrics are provided in supplementary material T.1.

LLMs generality analysis

To evaluate the generalizability of our model across different LLMs, we conducted comprehensive experiments using 6 LLMs. As shown in Table 4 on the Human Breast Cancer, our ST-LLM framework demonstrates robust performance across all tested models, with consistent improvements. Deepseek achieves the highest performance with AMI of 0.636 and ACC of 0.844, followed closely by Qwen and Llama. Notably, even smaller models like Smollm maintain reasonable performance, indicating that our framework can effectively leverage various LLM architectures regardless of their size or training paradigms. Other datasets are provided in supplementary material T.2.

Metric	Task only	Spatial only	Graph only	Bio only	Full
AMI	0.033	0.275	0.473	0.029	0.626
ACC	0.196	0.356	0.643	0.132	0.846

Table 5: Ablation study results comparing the contributions of individual embedding components on the Human Breast Cancer dataset.

Ablation Study

We conducted an ablation study to investigate contributions of each embedding component (Task, Spatial, Graph, Bio) individually. Quantitative results indicate the full model outperforms partial variants on Human Breast Cancer as shown in Table 5. The 'Graph only' variant achieves second-best results in AMI 0.473 and ACC 0.643, highlighting the critical role of spatial graph structures. In contrast, models relying solely on 'Task' or 'Bio' embeddings exhibit poorer performance of 0.033 and 0.196, reflecting limited single-modality capacity. The moderate performance of the 'Spatial only' variant of 0.275 and 0.356 further underscores the necessity of combining multi-modal embeddings. Other datasets are provided in supplementary material T.3.

Complexity Analysis

We analyze the model complexity of ST-LLM across its three stages by evaluating evaluating the number of parameters, computational performance (MFlops), time complexity, training time, and token length. We also compare these metrics with eight traditional model and three different prompts. Detailed results are provided in supplementary material T.4. ST-LLM significantly outperforms traditional methods in computational efficiency while achieving superior accuracy. In summary, ST-LLM not only demonstrates good efficiency in our experiments but also is a simple method to encoding spatial transcriptomics biological knowledge.

Conclusion

In this work, we introduced a novel Spatial Transcriptomics Embedding with Large Language Models (ST-LLM), which addresses critical challenges in spatial transcriptomics analysis through dynamic graph adjacency construction, graph-to-text conversion, and large language model embedding. Experimental evaluations demonstrated substantial performance improvements over existing methods. Prompts Comparison, LLMs generality analysis, Ablation study and Complexity Analysis highlighted the importance of integrating graph structural, spatial and biological context within our embedding strategy. ST-LLM not only delivers robust analytical performance but also uses short training time and simple method.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos. W2511070, 32241027, 62472034, 62227807), and in part by the National Key R&D Program of China (No. 2019YFA0706200).

References

Chen, A.; Liao, S.; Cheng, M.; Ma, K.; Wu, L.; Lai, Y.; Qiu, X.; Yang, J.; Xu, J.; Hao, S.; et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 185(10): 1777–1792.

Cui, H.; Wang, C.; Maan, H.; Pang, K.; Luo, F.; Duan, N.; and Wang, B. 2024. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*, 21(8): 1470–1480.

Dong, K.; and Zhang, S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, 13(1): 1739.

Gayoso, A.; Lopez, R.; Xing, G.; Boyeau, P.; Valiollah Pour Amiri, V.; Hong, J.; Wu, K.; Jayasuriya, M.; Mehlman, E.; Langevin, M.; et al. 2022. A Python library for probabilistic analysis of single-cell omics data. *Nature biotechnology*, 40(2): 163–166.

Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Hu, J.; Li, X.; Coleman, K.; Schroeder, A.; Ma, N.; Irwin, D. J.; Lee, E. B.; Shinohara, R. T.; and Li, M. 2021. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods*, 18(11): 1342–1351.

Ji, A. L.; Rubin, A. J.; Thrane, K.; Jiang, S.; Reynolds, D. L.; Meyers, R. M.; Guo, M. G.; George, B. M.; Mollbrink, A.; Bergenstråhle, J.; et al. 2020. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *cell*, 182(2): 497–514.

Jiao, S.; Lu, D.; Zeng, X.; Wang, T.; Wang, Y.; Dong, Y.; and Peng, J. 2024. DiffuST: a latent diffusion model for spatial transcriptomics denoising. *bioRxiv*, 2024–06.

Kleshchevnikov, V.; Shmatko, A.; Dann, E.; Aivazidis, A.; King, H. W.; Li, T.; Elmentaite, R.; Lomakin, A.; Kedlian, V.; Gayoso, A.; et al. 2022. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 40(5): 661–671.

Lester, B.; Al-Rfou, R.; and Constant, N. 2021. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*.

Li, B.; Zhang, W.; Guo, C.; Xu, H.; Li, L.; Fang, M.; Hu, Y.; Zhang, X.; Yao, X.; Tang, M.; et al. 2022. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature methods*, 19(6): 662–670.

Li, K.; Li, J.; Tao, Y.; and Wang, F. 2024. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Briefings in Bioinformatics*, 25(3): bbae171.

Li, M.; Kilicoglu, H.; Xu, H.; and Zhang, R. 2025. Biomedrag: A retrieval augmented large language model for biomedicine. *Journal of Biomedical Informatics*, 162: 104769.

Li, X. L.; and Liang, P. 2021. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.

Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.

Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; and Neubig, G. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM computing surveys*, 55(9): 1–35.

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.

Lopez, R.; Regier, J.; Cole, M. B.; Jordan, M. I.; and Yosef, N. 2018. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12): 1053–1058.

Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos Jr, J. L.; Xiong, C.; Sun, Z. Z.; Socher, R.; et al. 2023. Large language models generate functional protein sequences across diverse families. *Nature biotechnology*, 41(8): 1099–1106.

Marx, V. 2021. Method of the Year: spatially resolved transcriptomics. *Nature methods*, 18(1): 9–14.

Maynard, K. R.; Collado-Torres, L.; Weber, L. M.; Uytingco, C.; Barry, B. K.; Williams, S. R.; Catallini, J. L.; Tran, M. N.; Besich, Z.; Tippani, M.; et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24(3): 425–436.

Medina-Ortiz, D.; Contreras, S.; Fernández, D.; Soto-García, N.; Moya, I.; Cabas-Mora, G.; and Olivera-Nappa, Á. 2024. Protein language models and machine learning facilitate the identification of antimicrobial peptides. *International Journal of Molecular Sciences*, 25(16): 8851.

Moses, L.; and Pachter, L. 2022. Museum of spatial transcriptomics. *Nature methods*, 19(5): 534–546.

Palayew, S.; Wang, B.; and Bader, G. 2025. Towards Applying Large Language Models to Complement Single-Cell Foundation Models. *arXiv preprint arXiv:2507.10039*.

Pham, D.; Tan, X.; Xu, J.; Grice, L. F.; Lam, P. Y.; Raghubar, A.; Vukovic, J.; Ruitenberg, M. J.; and Nguyen, Q. 2020. stLearn: integrating spatial location, tissue morphology and

gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *biorxiv*, 2020–05.

Qin, G.; and Eisner, J. 2021. Learning how to ask: Querying LMs with mixtures of soft prompts. *arXiv preprint arXiv:2104.06599*.

Ren, H.; Walker, B. L.; Cang, Z.; and Nie, Q. 2022. Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nature communications*, 13(1): 4076.

Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; et al. 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15): e2016239118.

Rizvi, S. A.; Levine, D.; Patel, A.; Zhang, S.; Wang, E.; He, S.; Zhang, D.; Tang, C.; Lyu, Z.; Darji, R.; et al. 2025. Scaling Large Language Models for Next-Generation Single-Cell Analysis. *bioRxiv*, 2025–04.

Shin, T.; Razeghi, Y.; Logan IV, R. L.; Wallace, E.; and Singh, S. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. *arXiv preprint arXiv:2010.15980*.

Ståhl, P. L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J. F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J. O.; Huss, M.; et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294): 78–82.

Stickels, R. R.; Murray, E.; Kumar, P.; Li, J.; Marshall, J. L.; Di Bella, D. J.; Arlotta, P.; Macosko, E. Z.; and Chen, F. 2021. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature biotechnology*, 39(3): 313–319.

Stuart, T.; and Satija, R. 2019. Integrative single-cell analysis. *Nature reviews genetics*, 20(5): 257–272.

Svensson, V.; Teichmann, S. A.; and Stegle, O. 2018. SpatialDE: identification of spatially variable genes. *Nature methods*, 15(5): 343–346.

Theodoris, C. V.; Xiao, L.; Chopra, A.; Chaffin, M. D.; Al Sayed, Z. R.; Hill, M. C.; Mantineo, H.; Brydon, E. M.; Zeng, Z.; Liu, X. S.; et al. 2023. Transfer learning enables predictions in network biology. *Nature*, 618(7965): 616–624.

Wang, I.-H.; Murray, E.; Andrews, G.; Jiang, H.-C.; Park, S. J.; Donnard, E.; Durán-Laforet, V.; Bear, D. M.; Faust, T. E.; Garber, M.; et al. 2022. Spatial transcriptomic reconstruction of the mouse olfactory glomerular map suggests principles of odor processing. *Nature neuroscience*, 25(4): 484–492.

Wu, S. Z.; Al-Eryani, G.; Roden, D. L.; Junankar, S.; Harvey, K.; Andersson, A.; Thennavan, A.; Wang, C.; Torpy, J. R.; Bartonicek, N.; et al. 2021. A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics*, 53(9): 1334–1347.

Xu, H.; Fu, H.; Long, Y.; Ang, K. S.; Sethi, R.; Chong, K.; Li, M.; Uddamvathanak, R.; Lee, H. K.; Ling, J.; et al. 2024. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1): 12.

Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Yang, F.; Wang, W.; Wang, F.; Fang, Y.; Tang, D.; Huang, J.; Lu, H.; and Yao, J. 2022. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nature Machine Intelligence*, 4(10): 852–866.

Zhao, S.; Zhang, J.; and Nie, Z. 2023. Large-scale cell representation learning via divide-and-conquer contrastive learning. *arXiv preprint arXiv:2306.04371*.

Zhu, J.; Sun, S.; and Zhou, X. 2021. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome biology*, 22(1): 184.