scGraphETM: Graph-Based Deep Learning Approach for Unraveling Cell Type-Specific Gene Regulatory Networks from Single-Cell Multi-Omics Data

Wenqi Dong,[1,†,*] Manqi Zhou,[2,3,†] Boyu Han,[1] Fei Wang[3,4] and Yue Li[1,*]

[1]School of Computer Science, McGill University, 3480 University St.,Montréal, QC H3A 0G4, Québec, Canada, [2]Department of Computational Biology, Cornell University, 237 Tower Rd, 14853, New York, USA, [3]Institute of Artificial Intelligence for Digital Health, Weill Cornell Medicine, 575 Lexington Ave, 10021, New York, USA, [4]Division of Health Informatics, Department of Population Health Sciences, Weill Cornell Medicine, 402 E 67th St,, 10021, New York, USA and [†]These authors contributed equally.
[*]Corresponding author: vicky.dong@mail.mcgill.ca, yueli@cs.mcgill.ca

Abstract

In the forefront of single-cell multi-omics research, the challenge of elucidating intricate gene regulatory networks (GRNs) at a cellular level is paramount. This study introduces the Single Cell Graph Network Embedded Topic Model (scGraphETM), a novel computational approach aimed at unraveling the complexities of cell-specific GRNs from multi-omics single-cell sequencing data. Central to our investigation is the integration of single-cell RNA sequencing and single-cell ATAC sequencing data, leveraging the strengths of both to uncover the underpinnings of cellular regulation. The scGraphETM model innovatively combines a variational autoencoder framework with a graph neural network. By conceptualizing transcription factors (TFs), genes, and regulatory elements (RE) as nodes, and their regulatory interactions as edges, the model adeptly captures the dynamic regulatory interplay within cells. It uniquely incorporates both universal and cell-specific features, enabling the model to generalize across cell populations while also identifying unique regulatory dynamics within individual cells. Our results reveal that scGraphETM surpasses existing methodologies in accurately modeling cell-type clustering, cross-modality imputation and cell-type specific TF-RE relationships.

Key words: gene regulatory networks, single-cell multi-omics, deep learning

Introduction

The advent of single-cell technologies has revolutionized genomics by enabling the analysis of genetic material at the resolution of individual cells, offering a granular perspective essential for understanding cellular diversity and function in both health and disease (Stubbington et al., 2017; Poulin et al., 2016). While traditional bulk sequencing approaches often obscure crucial cellular differences by averaging signals across many cells, multi-omics single-cell sequencing techniques—such as single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq) provide a precise, cell type-specific view of genomic and epigenomic landscapes (Tsoucas et al., 2019).

These techniques are particularly valuable for studying gene regulatory networks (GRNs), which consist of transcription factors (TF), regulatory elements (RE) and target genes that control biological processes within cells (Bravo González-Blas et al., 2023). GRNs play a central role in understanding how cells interpret and respond to internal and external signals, providing insights into fundamental biological processes and disease mechanisms (Unger Avila et al., 2024; Singh et al., 2018). However, challenges remain in accurately modeling these networks due to the high-dimensional, noisy, and sparse nature of single-cell data, requiring sophisticated computational methods to extract meaningful insights (Fiers et al., 2018).

Recent advances in deep learning, particularly in embedded topic models (ETMs) (Dieng et al., 2020) and graph neural networks (GNNs) (Wu et al., 2020), offer promising approaches for analyzing complex single-cell multi-omics data (Zhao et al., 2021; Zhou et al., 2023; Wang et al., 2021; Cao and Gao, 2022). Existing methods, such as Seurat's integration tool, utilize the weighted nearest neighbor algorithm to merge multimodal single-cell data (Hao et al., 2021). MultiVI integrates scRNA and scATAC data via variational autoencoder (VAE) frameworks (Ashuach et al., 2023). BABEL pioneered cross-modality translation at single-cell resolution, enabling the prediction of one modality from another (Wu et al., 2021). moETM, a unified deep learning model, integrates single-cell multi-omics data into a common topic mixture representation and uses a linear decoder design to enhance interpretability while uncovering biologically significant patterns (Zhou et al., 2023). For regulatory network inference, LINGER demonstrated the power of incorporating atlas-scale external data to infer gene regulatory networks from single-cell multiome data, highlighting the importance of prior biological knowledge (Yuan and Duren, 2024). The GLUE framework introduced graph-linked embedding for multi-omics integration, leveraging the relationships between molecular measurements through graph representations (Cao and Gao, 2022). Although these methods show promising performance, they often require compromises in scalability, interpretability, and flexibility. Furthermore, many approaches focus on isolated tasks—such as integration, imputation, or GRN inference—and cannot address all of these tasks simultaneously. Additionally, they typically lack the ability to perform cell-type-specific inference, limiting their capacity to fully exploit the high resolution of single-cell data to capture cellular heterogeneity.

Here, we present the Single Cell Graph Network Embedding Topic Model (scGraphETM), which combines the strengths of Embedded Topic Models and Graph Neural Networks to unravel cell type-specific gene regulatory networks. scGraphETM employs a dual ETM architecture, utilizing modality-specific encoders and decoders data to preserve the distinct biological

information inherent in each modality. Additionally, the use of a linear decoder design enhances the interpretability of our model, ensuring that biologically significant patterns can be uncovered. To further enrich the GRN dynamics, we incorporate external biological knowledge from the cisTarget databases (Imrichová et al., 2015; Delgado et al., 2023), which helps model the intricate relationships between transcription factors, target genes, and regulatory elements across different cell types. Moreover, the framework's GNN-based neighborhood aggregation strategy ensures scalability, allowing the model to handle large datasets efficiently. scGraphETM is designed to perform three key tasks: (1) clustering cells into biologically meaningful groups to identify cell types, (2) imputing one omics modality using another, and (3) identifying cell type-specific transcription factor-regulatory element relationships. Through comprehensive experiments on three single-cell multimodal datasets, we demonstrate scGraphETM's superior performance in comparison to six state-of-the-art methods.

Material and Methods

scGraphETM model ooverview

scGraphETM integrates single-cell multimodal data across different experiments using interpretable latent embeddings and external GRN databases (**Fig.** 1). Building upon the widely used variational autoencoder, it combines multi-modal data integration with gene regulatory network to incorporate prior biological information. To tailor the framework for cell type-specific inference, we made three key contributions.

First, we adapted the xTrimoGene (Gong et al., 2023; Hao et al., 2024) approach to generate cell-specific embeddings. Initially, we used node2vec (Grover and Leskovec, 2016) to generate embeddings that reflect gene regulatory connectivity within the network, providing a macroscopic view based on random walk-derived patterns. These pre-trained node2vec embeddings are then integrated into the cell-specific embeddings computed previously (**Fig.** 1**a.**). The final gene embeddings combine both cell-specific embeddings and node2vec embeddings, enriching node features with both immediate data-driven insights and broader network context. Second, we adapted the neighborhood aggregation strategy from GraphSAGE (Hamilton et al., 2017) to reduce computational burden and efficiently scale to large datasets without sacrificing performance. Traditional graph attention mechanisms require pairwise attention calculations, which are computationally expensive, especially when dealing with single-cell data containing 10K+ genes and over 1 million peaks. GraphSAGE addresses this by sampling neighboring nodes, aggregating their feature information, and using this aggregation to predict graph context and labels. Third, we employed an embedding topic model as the linear encoder to enhance interpretability (Dieng et al., 2020). Building on our previous work (Zhao et al., 2021; Zhou et al., 2023), scGraphETM uses linear matrix factorization to reconstruct normalized count vectors from the cell embeddings. We hypothesize that by creating a linearly separable space, the encoder enables the decoder to achieve accurate reconstruction when the two networks are trained end-to-end.

Gene regulatory network construction

Our approach to constructing gene regulatory networks involves representing TFs, genes, and REs as nodes within a graph framework. Edges between these nodes represent regulatory interactions, which are initially identified based on transcription start site (TSS) proximity derived from the ENCODE database (Consortium et al., 2012). We employ a TSS distance threshold of 1 Mbp (Delgado et al., 2023) to determine relevant interactions, ensuring that only regulatory elements within this predefined proximity to a gene's TSS are considered for network construction. We incorporate motif information from the cisTarget database. CisTarget identifies transcription factors likely to bind specific motifs found within these regulatory elements, offering additional insights into potential regulatory mechanisms.

Specifically, TFs, genes, and REs (often referred to as peaks) are represented as nodes within a graph $G = (V, E)$, where V denotes the set of nodes and E represents the set of edges connecting these nodes. Each edge $e_{ij} \in E$ reflects a regulatory interaction between two nodes $v_i, v_j \in V$.

For a gene g and a peak p, an edge e_{gp} is added if the peak p lies within a predefined threshold distance 1 Mbp d_{TSS}, from the TSS of g, upstream or downstream. Mathematically, this condition is represented as:

$$e_{gp} \in E \quad \text{if} \quad \text{dist}(\text{TSS}_g, p) \leq d_{\text{TSS}},$$

where $\text{dist}(\text{TSS}_g, p)$ denotes the linear genomic distance between the TSS of gene g and the start of peak p.

For a transcription factor t and a peak p, an edge e_{tp} is established if cisTarget identifies binding motifs within p that are likely targets of t. This can be expressed as:

$$e_{tp} \in E \quad \text{if} \quad \text{MotifMatch}(t, p) = 1,$$

where we define $\text{MotifMatch}(t, p)$ to be a binary function that indicates whether cisTarget predicts a binding motif for t in p.

The resulting adjacency matrix A of the graph G has dimensions $(M \times M)$, where $M = \sum_{i=1}^{n} m_i$ represents the total number of feature across the n modalities. The off-diagonal elements of A capture the cross-modality interactions, while the diagonal blocks corresponding to gene-gene and peak-peak interactions remain unpopulated to exclude self-regulating relationships.

Cell-specific dynamic node features for the GRN

For the initial node feature embedding, we first train a node2vec model to capture the global structural properties of the GRN. Specifically, node2vec generates embeddings that reflect node connectivity within the network, providing a macroscopic view based on random walk-derived patterns. These pre-trained node2vec embeddings are integrated with the cell-specific embeddings following the implementation of xTrimoGene (Gong et al., 2023; Hao et al., 2024). When generating these gene and peak embeddings, an attention mechanism refines them based on current expression inputs, adjusting their importance dynamically. Importantly, xTrimoGene transforms each gene expression scalar value x_i into a latent vector $\mathbf{t}_i$ of dimension d, which effectively discretizes expression levels into distinct categories. Specifically, this is done as follows:

1. Project the scalar value $x_i^{\text{input}} \in \mathbb{R}$ onto score vector $\mathbf{h}_i \in \mathbb{R}^b$ with an FFN:

$$\mathbf{a}_i = \text{leakyReLU}(x_i^{\text{input}} \times \mathbf{w}_1), \quad \mathbf{z}_i = \mathbf{a}_i \mathbf{w}_2 + \alpha \mathbf{a}_i$$

where $\mathbf{w}_1 \in \mathbb{R}^{1 \times b}$, $\mathbf{w}_2 \in \mathbb{R}^{b \times b}$, and α are learnable parameters. The vector $\mathbf{a}_i \in \mathbb{R}^b$, and $\mathbf{z}_i \in \mathbb{R}^b$.

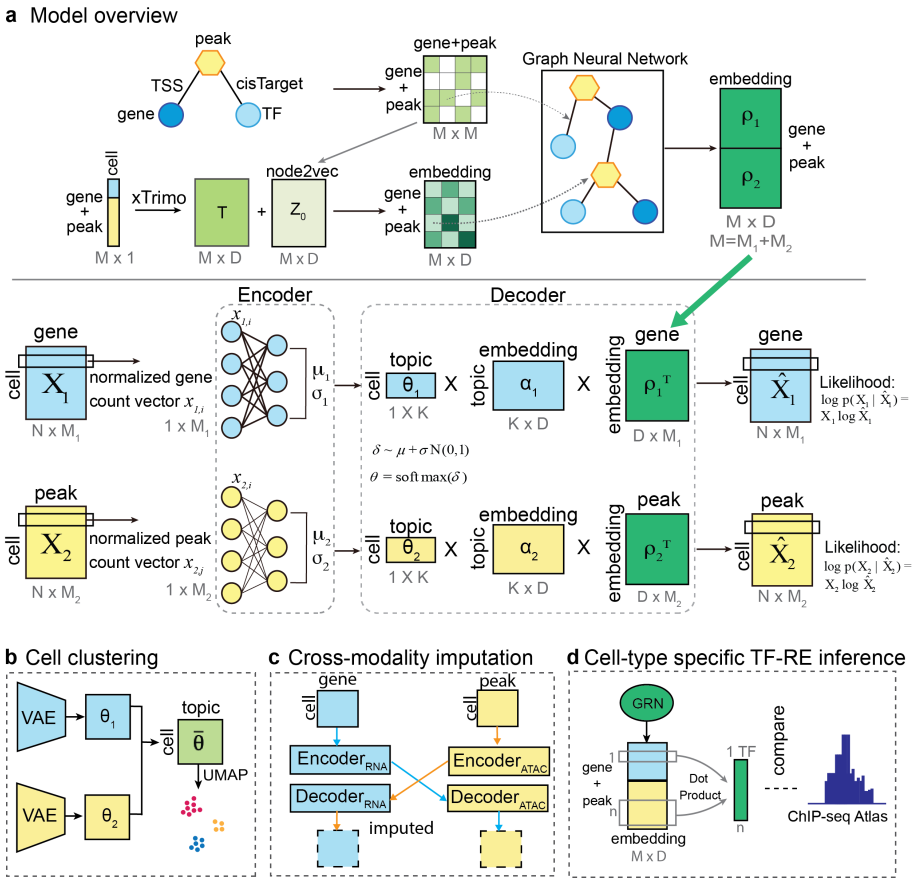

Fig. 1. scGraphETM overview a. The overall framework of scGraphETM. Model consists of a GNN, modality-specific encoder-decoders. Trained end to end. **b.** Cell clustering based on topic distributions. **c.** Cross-modality imputation. **d.** Cell type-specific TF-RE relationship inference using the feature embedding outputted by the GNN.

2. Normalize $\mathbf{z}_i$ with softmax to produce vector $\boldsymbol{\gamma}_i \in \mathbb{R}^b$:

$$\gamma_{i,j} = \frac{\exp(z_{i,j})}{\sum_{j=1}^{b} \exp(z_{i,j})}$$

3. The value embedding is a weighted sum of all b embeddings from the learnable embedding codebook $\mathbf{T} \in \mathbb{R}^{b \times d}$:

$$\mathbf{t}_i = \boldsymbol{\gamma}_i \mathbf{T}$$

   where $\mathbf{t}_i \in \mathbb{R}^d$.
   Zeros are encoded with $\mathbf{t}^0 \in \mathbb{R}^d$.
4. The final input embedding for gene i is:

$$\mathbf{V}_i^{\text{input}} = \mathbf{t}_i + \mathbf{z}_i^G$$

   where $\{\mathbf{z}_i^G\}_{i=1}^G = \mathbf{Z}^G \in \mathbb{R}^{G \times d}$ denotes node2vec embedding for all G genes, and $\mathbf{V}_i^{\text{input}} \in \mathbb{R}^d$.

In the context of token embedding in NLP, the gene-independent and value-dependent latent vector $\mathbf{t}_i$ can be considered as the "position embedding" and the gene-specific and value-independent node2vec embeddings $\mathbf{z}_i^G$ are the "word embeddings". The process culminates with the final gene embeddings being a combination of attention-adjusted features and node2vec embeddings, thus enriching the node features with both data-driven insights and broader network context.

Graph neural network component

We employed a neighborhood aggregation strategy inherited from GraphSAGE (Hamilton et al., 2017). GraphSAGE works by sampling and aggregating features from a node's immediate neighbors, thus reducing the computational load when handling large datasets. The core functionality of GraphSAGE is defined by its update rules, which specify how node representations are updated based on their neighborhoods:

$$\mathbf{h}_v^{(k)} = \sigma\left(\mathbf{W}^{(k)} \cdot \text{MEAN}\left(\{\mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v)\} \cup \{\mathbf{h}_v^{(k-1)}\}\right)\right)$$

Here, $\mathbf{h}_v^{(k)}$ represents the feature vector of node v at layer k, $\mathcal{N}(v)$ denotes the set of neighbors of v, and $\mathbf{W}^{(k)}$ are the trainable parameters at layer k. The function σ is a non-linear activation function such as the ReLU. This recursive update rule allows GraphSAGE to efficiently aggregate and update node features across multiple layers, enabling the learning of powerful representations that reflect both the local structure and node-specific information.

Embedded Topic Model component

The Embedded Topic Model (ETM) leverages an encoder-decoder framework where the encoder maps high-dimensional data to a lower-dimensional topic space, and the decoder reconstructs the original data from this topic representation, facilitating the discovery of underlying biological themes.

Encoder

The encoder of the scGraphETM model is structured as a two-layer fully connected neural network, tasked with inferring topic proportions from normalized count vectors of multi-omics data for individual cells. It processes each cell's multi-omics data through its layers to generate parameters that define a logistic normal distribution for each omics type. These distributions are assumed to capture the latent representation of the respective omics data, where each dimension corresponds to a potential topic or underlying biological factor. The primary goal of the encoder is to effectively synthesize these independent logistic normal distributions into a cohesive joint distribution that represents the comprehensive latent profile of the multi-omics data. This synthesis allows the model to capture the complex interdependencies and unique contributions of each omics layer.

The encoder is described by:

$$\boldsymbol{\mu}^*,\ \log \boldsymbol{\sigma}^* = \text{Encoder}(\mathbf{x}) \tag{1}$$

A reparameterization trick enables stochastic gradient descent through sampling:

$$\mathbf{z} = \boldsymbol{\mu}^* + \exp\left(\frac{\log \boldsymbol{\sigma}^*}{2}\right) \odot \boldsymbol{\epsilon} \tag{2}$$

where $\boldsymbol{\epsilon}$ is sampled from a standard normal distribution. Latent topic mixture are modeled as:

$$\theta_k = \text{softmax}(\mathbf{z})_k = \frac{\exp(z_k)}{\sum_k \exp(z_k)} \tag{3}$$

Decoder

In the scGraphETM model, the decoder is intricately designed to leverage the embeddings produced by a GNN to reconstruct the input data from the topic embeddings and the node embeddings.

$$\hat{\mathbf{x}} = \boldsymbol{\theta}\,\boldsymbol{\beta} \tag{4}$$

$$\log p(\mathbf{x}|\hat{\mathbf{x}}) = \sum_g x_g \log \hat{x}_g \tag{5}$$

where $\beta_{k,g} = \frac{\exp(\boldsymbol{\alpha}_k \boldsymbol{\rho}_g)}{\sum_{g'} \exp(\boldsymbol{\alpha}_k \boldsymbol{\rho}_{g'})}$ and $\boldsymbol{\rho}_g$ is the refined feature from the GNN component. The optimization with respect to the encoder and decoder maximizes an evidence lower bound (ELBO):

$$\mathbb{E}_{q(z|x)}[\log p(\mathbf{x}|\mathbf{z})] - KL[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})], \tag{6}$$

where $\mathbb{E}_{q(z|x)}[\log p(\mathbf{x}|\mathbf{z})] = \frac{1}{S}\log p(\mathbf{x}|\hat{\mathbf{z}}_s)$ are approximated by Monte Carlo sampling and $KL[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] = \mathbb{E}_{q(z)}[\log q(z|x)] - \mathbb{E}_{q(z)}[\log p(z)] = \sum_k -\frac{1}{2}\left[\log\{\sigma_k^*\}^2 - \{\mu_k^*\}^2 - \{\sigma_k^*\}^2 + 1\right]$ has closed-form expression.

Model training and inference

The scGraphETM model employs an end-to-end training approach where the encoder, decoder, and GNN are optimized simultaneously. The training is guided by a composite loss function that includes Average Negative Log-Likelihood (NLL), which assesses the fidelity of the reconstructed data; Kullback-Leibler (KL) divergence, which provides regularization by ensuring the latent variable distributions remain plausible; and Graph Reconstruction Loss, which ensures that the model accurately captures the biological interactions within gene regulatory networks. The overall loss function is formulated as $\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{NLL}} + \lambda_2\mathcal{L}_{\text{KL}} + \lambda_3\mathcal{L}_{\text{GR}}$. Specifically, $\mathcal{L}_{\text{NLL}} = -\sum_{i=1}^{m} \frac{1}{N_i M_i}\mathbf{x_i}\log(\hat{\mathbf{x}}_\mathbf{i})$, $\mathcal{L}_{\text{KL}} = \sum_k -\frac{1}{2}\left[\log\{\sigma_k^*\}^2 - \{\mu_k^*\}^2 - \{\sigma_k^*\}^2 + 1\right]$, $\mathcal{L}_{\text{GR}} = -\frac{1}{|E|}\sum_{(i,j)\in E}\left[\mathbf{A}_{ij}\log\sigma\left(\hat{\mathbf{A}}_{ij}\right) + (1-\mathbf{A}_{ij})\log\left(1-\sigma\left(\hat{\mathbf{A}}_{ij}\right)\right)\right]$, where $\mathbf{A} \in \mathbb{R}^{M\times M}$ is the adjacency matrix of the graph G, $\mathbf{V} \in \mathbb{R}^{M\times D}$ is the node embedding, $\hat{\mathbf{A}} = \mathbf{V}\mathbf{V}^T \in \mathbb{R}^{M\times M}$, $\sigma(\cdot)$ is the sigmoid function. λ_1, λ_2, and λ_3 are hyperparameters optimized during the validation phase to balance each component's impact effectively.

Single-cell multi-omic benchmark data

1. Peripheral Blood Mononuclear Cells (PBMC) from 10X Genomics, consisting of 9,631 cells, 29,095 genes, and 112,975 peaks across 19 cell types.
2. Multiome bone marrow mononuclear cells (BMMC) dataset from the 2021 NeurIPS challenge, consisting of 69,249 cells, 13,403 genes and 110,359 peaks with 22 cell types from 10 donors across 4 sites
3. Human Cortex dataset from BRAIN Initiative (Siletti et al., 2023; Li et al., 2023), consisting of 45,549 cells, 30,033 genes, 262,997 peaks with 13 cell types.

All datasets were processed into the format of samples-by-features matrices. Initially, the read count for each gene and peak were first normalized per cell by total counts within the same omic using scanpy.pp.normalize_total function in the scanpy. Next, log1p transformation was applied. Lastly, scanpy.pp.highly_variable_genes was used to select highly variable genes or peaks.

Evaluation metrics

For the cell type clustering task, we employed four evaluation metrics to assess performance:

1. Adjusted Rand Index (ARI) (Hubert and Arabie, 1985): quantifies the similarity between two clusterings while correcting for the chance that pairs of objects might be randomly assigned to the same clusters.
2. Normalized Mutual Information (NMI) (Danon et al., 2005): measures the amount of information shared between two clusterings, normalized by the average entropy of the clusterings.
3. kBET Büttner et al. (2019): tests whether batch labels are distributed differently across cells using Pearson's chi-square test.
4. GC: evaluates whether cells of the same type from different batches are close in the embedding space by constructing a k-nearest neighbor graph.

For the cross-modality imputation task, we compute the Pearson correlation coefficients (PCC) and Spearman correlation between the predicted and observed signals. For the cell-type-specific GRN inference task, we used the Area Under the Precision-Recall Curve (AUPRC) Spackman (1989) as the evaluation metric to assess accuracy, as the PR curve illustrates the trade-off between precision and recall at various decision thresholds.

Results

scGraphETM accurately integrates multimodal data for cell type clustering

We evaluated the integrated low-dimensional representation of scGraphETM by comparing it with three state-of-the-art multi-omics integration methods (moETM (Zhou et al., 2023), multiVI (Ashuach et al., 2023), and Seurat V4 (Hao et al., 2021)) across three published datasets. The performance of the multi-omics integration task was assessed using both biological conservation metrics—ARI and NMI—and batch removal metrics, including the kBET and GC. To ensure a comprehensive comparison, we performed 80/20 random split for training and testing and repeated 10 times.

Overall, scGraphETM outperformed the other methods on average across the three datasets, achieving the best overall performance in three out of four evaluation metrics (**Fig.** 2, **Table** S2). Specifically, scGraphETM achieved an ARI of 0.774, an NMI of 0.8215, and a GC of 0.9381. For kBET, scGraphETM ranked second with a score of 0.236. For individual datasets, scGraphETM was the top performer or second-best in all of the 3 datasets, demonstrating its robustness across different data sources. We hypothesize that scGraphETM's superior integration performance can be attributed to its GRN component and the use of highly variable genes (HVGs). To test this, we created two variants: scGraphETM_noGraph, which excluded the GRN component, and scGraphETM_allCodingGenes, which included all coding genes based on the BioMart database (Durinck et al., 2009). As expected, scGraphETM_noGraph consistently underperformed compared to scGraphETM across all datasets and evaluation metrics, further highlighting the importance of the GRN component. Additionally, scGraphETM_allCodingGenes performed better than scGraphETM_noGraph but slightly worse, which is expected since most non-HVGs are cell-type markers.

We also validated the clustering performance by visualizing the cell topic mixture embeddings using Uniform Manifold Approximation and Projection (UMAP) McInnes et al. (2018). For instance, 'CD16+ Mono' cells were tightly clustered together by scGraphETM, whereas MoETM and MultiVI split them into two smaller clusters, as highlighted by the red box in Fig.2c. In contrast, 'B1 B' cells were clearly separated from their neighboring cells using scGraphETM, while MoETM showed some overlap with 'Naive CD20+ B' cells. Additionally, MultiVI exhibited overlap between neighboring 'Naive CD20+ B' cells and 'Plasma celles', as highlighted by the orange box in **Fig.** 2c. These results indicate that scGraphETM improves cell clustering by incorporating the GRN component.

scGraphETM enhances interpretability through embedding topic model

By incorporating the embedding topic model as the linear encoder, we were able to interpret latent embedding by associating specific topics with cell types, thereby uncovering distinct cell-type signatures. Specifically, using the learned latent topic mixture cells-by-topics matrix $\boldsymbol{\theta}$, we linked each topic to the cell type exhibiting the highest average topic score across cells and explored top features from the topics-by-genes matrix $\boldsymbol{\beta}$.

For example, topic 28 and topic 88 were primarily associated with monocytes, while topic 93 was enriched for B cells (**Fig.** 4). These associations were clearly reflected in the topic mixture probabilities across individual cells. For each cell-type-enriched topic, we observed that some top genes played key roles in the biological processes of the corresponding cell types, as supported by the literature. For instance, *HECTD2*, associated with topic 28 monocytes, has been shown in previous studies to promote monocyte infiltration while inhibiting the infiltration of other cell types, such as activated mast cells (Gong et al., 2024). Additionally, *IPCEF1*, which is associated with topic 93 B cells, has been shown in prior research to exhibit a positive correlation between its expression level and B cells (Yin et al., 2024).

scGraphETM enables accurate cross-modality imputation

We evaluated the imputation accuracy of scGraphETM of gene expression from chromatin accessibility (ATAC2RNA) and vice versa (RNA2ATAC). To enhance its imputation ability, we fine-tuned scGraphETM with imputation loss, resulting in the scGraphETM_impute. We evaluated the imputed values by comparing them with three state-of-the-art imputation models (moETM, multiVI, BABEL) across three datasets. Performance was assessed using Pearson and Spearman correlations. Experiments were repeated 10 times each with 80/20 random training/test splits.

scGraphETM_impute outperformed the other methods across all the datasets and evaluation metrics. Specifically, when predicting gene expression from chromatin accessibility data, scGraphETM_impute achieved the best performance, with an average Pearson correlation of 0.73 (Spearman 0.40) (**Fig.** 3, **Table** S1). Predicting chromatin accessibility from gene expression presents a greater challenge, as it requires predicting from a lower-dimensional to a higher-dimensional space. Nonetheless, scGraphETM_impute performed best in this task, achieving a Pearson value of 0.57 and a Spearman value of 0.35. We hypothesize that our superior imputation performance arises from the incorporation of the imputation loss and the use of HVGs. To validate this, we conducted an ablation study and created three variants of the original scGraphETM model: scGraphETM_impute (with added imputation loss), scGraphETM_allCodingGene, and scGraphETM_allCodingGene_impute (with added imputation loss). Indeed, we observed lower correlation values when using all coding genes in the ATAC2RNA task, and comparable performances in the RNA2ATAC imputation. This is likely because non-variable genes do not provide additional useful information for high-to-low dimensional imputation tasks but may contribute more for low-to-high dimensional tasks. In our ablation study on imputation loss (comparing scGraphETM with scGraphETM_impute, and scGraphETM_allCodingGenes with scGraphETM_allCodingGenes_impute), we found that adding imputation loss consistently improved performance when using all coding genes. For HVGs, the addition of imputation loss enhanced performance in the ATAC2RNA task and yielded comparable results in the more challenging RNA2ATAC task. Qualitatively, the imputed gene expression values and chromatin accessibility peaks displayed consistent patterns and strong linear correlations with the observed data, further validating the model's accuracy (**Fig.** 3c).

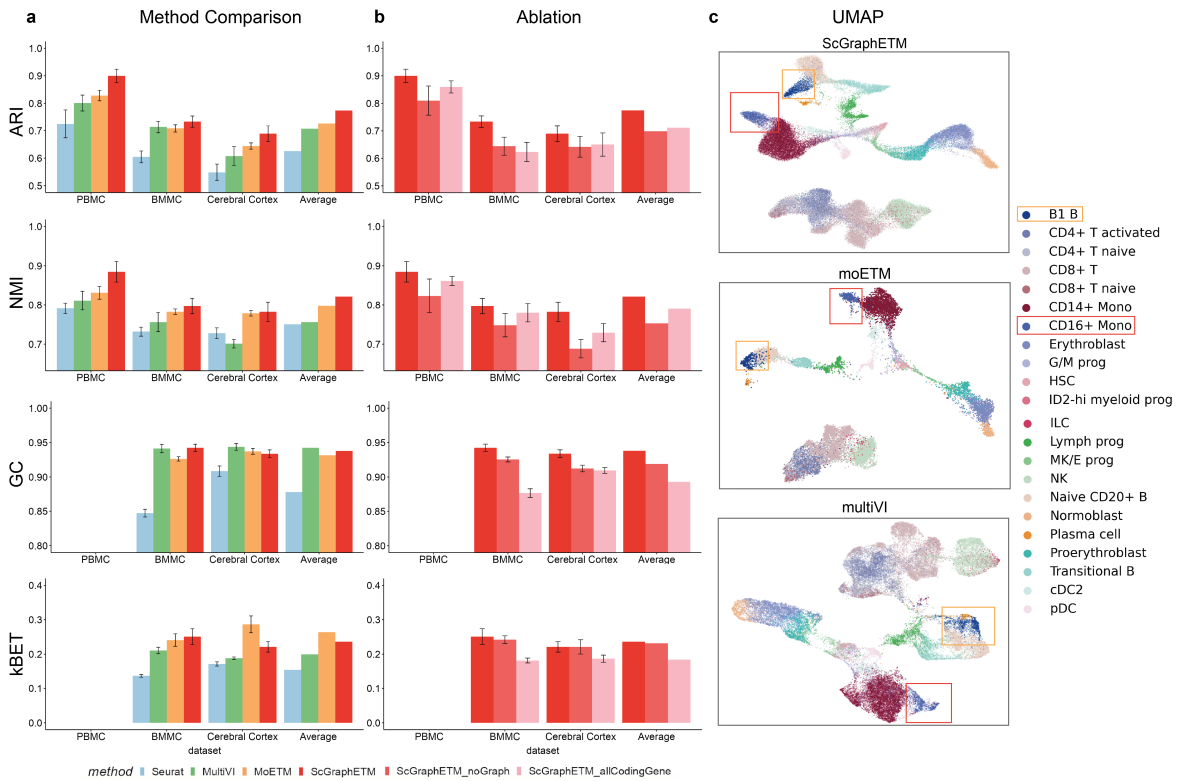

Fig. 2. Methods comparison based on cell clustering. a Individual performance of each method on each dataset, as well as the averaged values across all datasets. Each row corresponds to a different evaluation metric. Since the PBMC dataset consists of only one batch, the batch effect removal evaluation metrics, GC and kBET, were not applicable and are therefore left blank for the PBMC dataset. **b** Performance of scGraphETM with its ablated versions. **c** UMAP visualization on the BMMC dataset and distinguishable cell types clusterings.

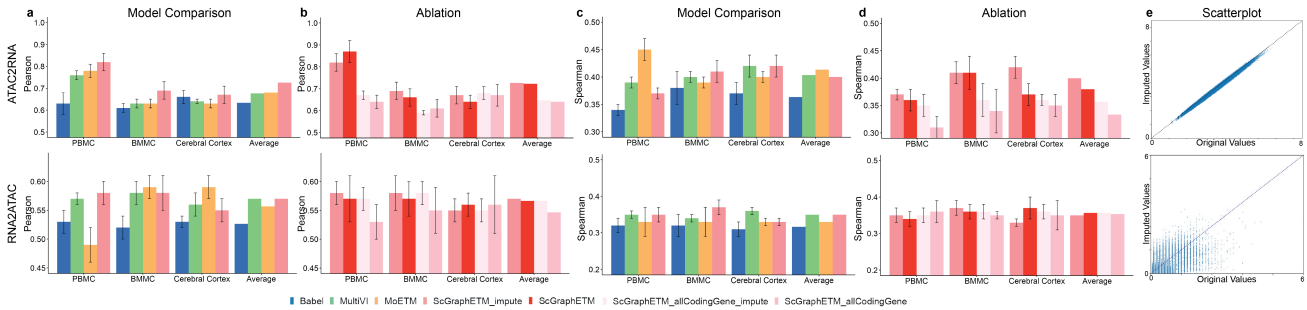

Fig. 3. Methods comparison based on cross-modality imputation. The upper panel displays performance on the ATAC2RNA imputation task, while the lower panel shows performance on the RNA2ATAC imputation task. **a,b** Pearson correlation for each method and scGraphETM's ablated versions on each dataset, along with the average Pearson correlation values across all datasets. **c,d** Spearman correlation for each method and scGraphETM's ablated versions on each dataset, along with the average values across all datasets. **e** Scatterplot of original versus imputed values, with the diagonal line shown in blue.

scGraphETM reveals cell-type-specific transcription factor-regulatory element relationships

To assess the cell-type-specific TF–RE binding potential, we first calculated the dot product of the cell-specific graph embeddings for the TF and the regulatory region: $r_{i,g,p} = (\boldsymbol{\rho}_{i,g}^{TF})^\top \cdot \boldsymbol{\rho}_{i,p}^{RE}$ for TF g and peak p in cell i; we then took the average across the cells for the same cell type: $r_{t,g,p} = \frac{1}{|\mathcal{S}_t|}\sum_{i\in\mathcal{S}_t} r_{i,g,p}$, where $\mathcal{S}_t$ denotes the set of cells of cell type t. We evaluated scGraphETM's cell-type-specific GRN inference using PBMC single-cell multiome data. Since most regions for each TF are negative examples, we used the AUPRC as the evaluation metric to focus on accurately identifying positive cases. An independent database, ChIP-seq Atlas (Zou et al., 2024), was used as the ground truth. We compared scGraphETM's performance with LINGER and GLUE.

scGraphETM achieved an average AUPRC of 0.613 across all 14 TFs and 3 cell types with ground truth data from the ChIP-seq Atlas database (**Fig.** 5, **Table** 1). In comparison, LINGER attained an average AUPRC of 0.599, and GLUE achieved 0.606. Specifically, for CD4 T cells, scGraphETM reached 0.608, while LINGER scored 0.598 and GLUE achieved 0.606. In B cells, scGraphETM outperformed both with a score of 0.609, compared to LINGER's 0.583 and GLUE's 0.605. On memory B cells, scGraphETM delivered the highest score of

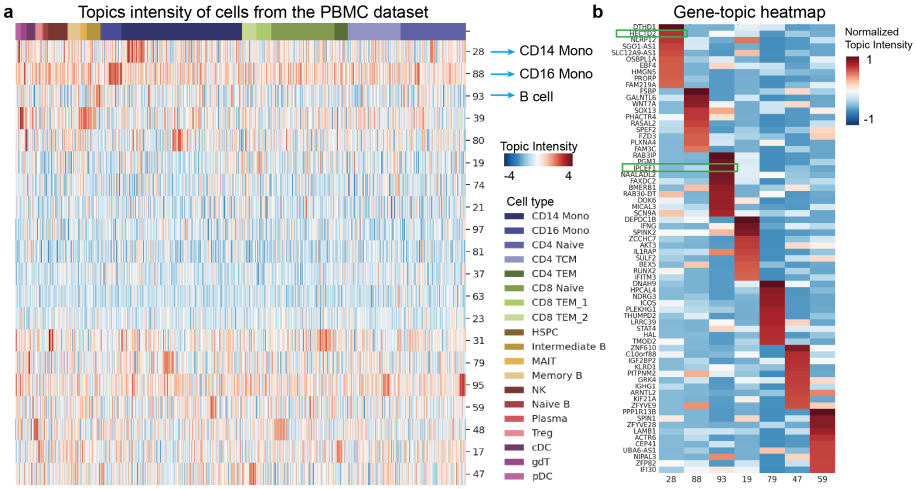

Fig. 4. Topic analysis on the PBMC dataset a Topic intensity of cells from the PBMC dataset. Each row represents a topic and each column represents a cell. **b** Top 10 genes per selected topics.

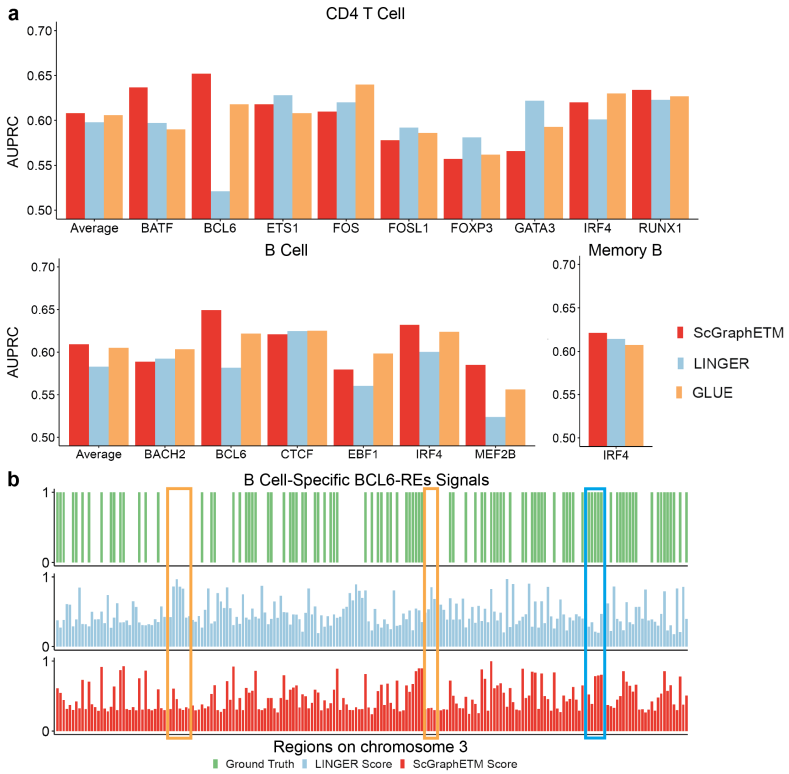

Fig. 5. Methods comparison based on cell-type-specific GRN. a. Comparison across different cell types, TFs, and methods on the task of TF-RE binding potential inference. The x-axis represents transcription factors, and the y-axis represents AUPRC values. **b.** B cell-specific BCL6-RE predicted signals, compared with ChIP-Seq Atlas ground truth. The x-axis represents peaks on chr3, and y-axis shows the TF-RE binding potential score.

0.6213, while LINGER averaged 0.6144 and GLUE achieved 0.6073. These results underscore scGraphETM's ability to consistently outperform other methods across multiple cell types and TFs. While LINGER requires bulk data and TF-specific training, limiting its ability to predict all TFs in a single model, GLUE is limited by its inability to handle large graph sizes, necessitating subsampling when graph features become too extensive. In contrast, scGraphETM excels by predicting TF–RE relationships within a single, unified framework. Additionally, scGraphETM boasts faster training speeds, completing an epoch within 30 seconds with a batch size of 32, outpacing GLUE, which takes 40 seconds, in terms of training efficiency in the same setting.

To further assess our predictions, we visualized the predicted potential scores and compared them to the ground truth signals. For instance, for the B cell type, scGraphETM achieved a score of 0.649 for the transcription factor BCL6, whereas LINGER scored 0.581. As shown in Fig. 5b, scGraphETM produced

Cell Type	TF	ScGraphETM	LINGER	GLUE
CD4 T cell	FOS	0.610	0.620	0.640
CD4 T cell	IRF4	0.620	0.601	0.630
CD4 T cell	BCL6	0.652	0.521	0.618
CD4 T cell	FOXP3	0.557	0.581	0.562
CD4 T cell	FOSL1	0.578	0.592	0.586
CD4 T cell	BATF	0.637	0.597	0.590
CD4 T cell	GATA3	0.566	0.622	0.593
CD4 T cell	RUNX1	0.634	0.623	0.627
CD4 T cell	ETS1	0.618	0.628	0.608
CD4 T cell	average	**0.608**	0.598	0.606
B cell	MEF2B	0.585	0.524	0.556
B cell	EBF1	0.580	0.561	0.598
B cell	BCL6	0.649	0.582	0.622
B cell	IRF4	0.632	0.600	0.624
B cell	BACH2	0.589	0.593	0.603
B cell	CTCF	0.621	0.625	0.625
B cell	average	**0.609**	0.581	0.605
Memory B cell	IRF4	**0.621**	0.614	0.607

Table 1. AUPRC for cell type-specifc TF-RE prediction on the PBMC dataset.For each cell type, the values are averaged across TFs, with the best performance highlighted in bold.

higher prediction scores in regions bound by the TF, as highlighted by the blue box; it also conferred higher specificity than LINGER in the true negative regions as highlighted by the orange boxes.

Conclusions and Discussion

Inferring cell-specific GRN becomes possible with the advent of single-cell multi-omic technologies. However, existing methods are limited in accurate GRN inference at the single-cell resolution. To tackle this challenge, the proposed scGraphETM in this study leverages graph-based representation learning to infer cell-specific GRNs from single-cell multiomic data by incorporating prior regulatory knowledge graph and cell-specific dynamic embedding technique. Compared to existing tools, scGraphETM achieves substantially better performance across multiple tasks, including cell-type clustering, cross-modality imputation, and GRN inference – all within a unified framework. scGraphETM has three key novel contributions: a GNN component based on xTrimoGene to incorporate cell-specific expression and chromatin accessibility into the graph, a neighborhood aggregation strategy via GraphSAGE to handle large-scale nodes and GRN relationships, and a linear encoder using embedding topic modeling to enhance interpretability.

As future work, we plan to explore several directions. First, we will explore transformer-based graph convolution to uncover de novo regulatory interactions not captured in existing GRN graph databases. Second, other omics data types such as single-cell proteomics could further improve downstream interpretability. This can be done in a mosaic data integration. Third, we will extend scGraphETM to model spatial transcriptomic data by identifying spatiotemporal-specific regulatory circuits and cell-cell communications via ligand-receptor interaction network Raghavan et al. (2023).

Data Availability

All datasets used in this study are publicly available and were obtained from established data repositories.The single-cell multimodal datasets used in this study are as follows: the PBMC dataset is available at 10X Genomics, the BMMC dataset and the Cerebral Cortex dataset can be found on Gene Expression Omnibus under series numbers GSE194122 and GSE204684, respectively. For the gene regulatory network databases, cisTarget is available at `https://resources.aertslab.org/cistarget/`, and the ChIP-seq Atlas is accessible at `https://chip-atlas.org/`.

Code Availability

All original code has been deposited at Github.

Acknowledgement

Y.L. is supported by Canada Research Chair (Tier 2) in Machine Learning for Genomics and Healthcare (CRC-2021-00547) and Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (RGPIN-2016-05174). We thank members of the Li lab for their feedback and comments on earlier iterations of this work.

References

Tal Ashuach, Mariano I Gabitto, Rohan V Koodli, Giuseppe-Antonio Saldi, Michael I Jordan, and Nir Yosef. Multivi: deep generative model for the integration of multimodal data. *Nature Methods*, 20(8):1222–1231, 2023.

Carmen Bravo González-Blas, Seppe De Winter, Gert Hulselmans, Nikolai Hecker, Irina Matetovici, Valerie Christiaens, Suresh Poovathingal, Jasper Wouters, Sara Aibar, and Stein Aerts. Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods*, 20(9):1355–1367, 2023.

Maren Büttner, Zhichao Miao, F Alexander Wolf, Sarah A Teichmann, and Fabian J Theis. A test metric for assessing single-cell rna-seq batch correction. *Nature methods*, 16(1): 43–49, 2019.

ZJ Cao and G Gao. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*, 40:1458–1466, 2022. doi: https://doi.org/10.1038/s41587-022-01284-4.

ENCODE Project Consortium et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57, 2012.

Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*, 2005(09):P09008, 2005.

Ricardo D Delgado, Ángela Chaves, Pierre Geurts, and Stein Aerts. Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods*, 20(8):1155–1166, 2023.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453, 2020.

Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic

datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184–1191, 2009.

Mark WEJ Fiers, Liesbeth Minnoye, Sara Aibar, Carmen Bravo González-Blas, Zeynep Kalender Atak, and Stein Aerts. Mapping gene regulatory networks from single-cell omics data. *Briefings in functional genomics*, 17(4):246–254, 2018.

Jing Gong, Minsheng Hao, Xin Zeng, Chiming Liu, Jianzhu Ma, Xingyi Cheng, Taifeng Wang, Xuegong Zhang, and Le Song. xtrimogene: An efficient and scalable representation learner for single-cell rna-seq data. biorxiv. 2023.

Libao Gong, Jiayi Huang, Xue Bai, Lin Song, Junjie Hang, and Jinfeng Guo. Expression of hectd2 predicts peritoneal metastasis of gastric cancer and reconstructs immune microenvironment. *Cancer Cell International*, 24 (1):380, 2024.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Xuegong Zhang, and Le Song. Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, pages 1–11, 2024.

Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13): 3573–3587, 2021.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2:193–218, 1985.

Hana Imrichová, Gert Hulselmans, Zeynep Kalender Atak, Delphine Potier, and Stein Aerts. i-cistarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic acids research*, 43(W1):W57–W64, 2015.

Yang Eric Li, Sebastian Preissl, Michael Miller, Nicholas D Johnson, Zihan Wang, Henry Jiao, Chenxu Zhu, Zhaoning Wang, Yang Xie, Olivier Poirion, et al. A comparative atlas of single-cell chromatin accessibility in the human brain. *Science*, 382(6667):eadf7044, 2023.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Jean-Francois Poulin, Bosiljka Tasic, Jens Hjerling-Leffler, Jeffrey M Trimarchi, and Rajeshwar Awatramani. Disentangling neural cell diversity using single-cell transcriptomics. *Nature neuroscience*, 19(9):1131–1141, 2016.

Vishvak Raghavan, Yue Li, and Jun Ding. Harnessing agent-based modeling in cellagentchat to unravel cell-cell interactions from single-cell data. *bioRxiv*, pages 2023–08, 2023.

Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, Ka Wai Lee, Song-Lin Ding, Lijuan Hu, Peter Lönnerberg, Trygve Bakken, Tamara Casper, Michael Clark, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046, 2023.

Arun J Singh, Stephen A Ramsey, Theresa M Filtz, and Chrissa Kioussi. Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, 75:1013–1025, 2018.

Kent A Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the sixth international workshop on Machine learning*, pages 160–163. Elsevier, 1989.

Michael JT Stubbington, Orit Rozenblatt-Rosen, Aviv Regev, and Sarah A Teichmann. Single-cell transcriptomics to explore the immune system in health and disease. *Science*, 358(6359):58–63, 2017.

Daphne Tsoucas, Rui Dong, Haide Chen, Qian Zhu, Guoji Guo, and Guo-Cheng Yuan. Accurate estimation of cell-type composition from gene expression data. *Nature communications*, 10(1):2975, 2019.

Paula Unger Avila, Tsimafei Padvitski, Ana Carolina Leote, He Chen, Julio Saez-Rodriguez, Martin Kann, and Andreas Beyer. Gene regulatory networks in disease and ageing. *Nature Reviews Nephrology*, pages 1–18, 2024.

Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scgnn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12 (1):1–11, 2021.

Kevin E Wu, Kathryn E Yost, Howard Y Chang, and James Zou. Babel enables cross-modality translation between multiomic profiles at single-cell resolution. *Proceedings of the National Academy of Sciences*, 118(48):e2023070118, 2021.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Dechao Yin, Kun Wang, Junyu Zhao, Jinming Yao, Xiaofang Han, Bo Yan, Jianjun Dong, and Lin Liao. Ipcef1: Expression patterns, clinical correlates and new target of papillary thyroid carcinoma. *Journal of Cancer*, 15(19):6434, 2024.

Qiuyue Yuan and Zhana Duren. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nature Biotechnology*, pages 1–11, 2024.

Yifan Zhao, Huiyu Cai, Zuobai Zhang, Jian Tang, and Yue Li. Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data. *Nature communications*, 12(1):5261, 2021.

Manqi Zhou, Hao Zhang, Zilong Bai, Dylan Mann-Krzisnik, Fei Wang, and Yue Li. Single-cell multi-omics topic embedding reveals cell-type-specific and covid-19 severity-related immune signatures. *Cell Reports Methods*, 3:100563, 2023. doi: 10.1016/j.crmeth.2023.100563.

Zhaonan Zou, Tazro Ohta, and Shinya Oki. Chip-atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Research*, page gkae358, 2024.