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

Vicky Dong

School of Computer Science

McGill University, Montreal

March, 2025

A thesis submitted to McGill University in partial fulfillment

of the requirements of the degree of

Master of Computer Science

Abstract

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. Single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq) provide a precise, cell-specific view of genomic and epigenomic landscapes. These techniques are particularly valuable for studying gene regulatory networks (GRNs), which consist of transcription factors (TFs), regulatory elements (REs), and target genes that control biological processes within cells. At the forefront of single-cell multi-omics research, the challenge of elucidating intricate GRNs at a cellular level remains paramount due to single-cell data's high-dimensional, noisy, and sparse nature. To address this challenge, we present Single Cell Graph Network Embedded Topic Model (scGraphETM), building upon the previously published single-cell Embedded Topic Model (scETM). The approach leverages the advantages of topic modeling, graph neural networks, and multi-modal data integration techniques to unravel the complexities of cell-specific GRNs from multi-omics single-cell sequencing data. This structure adeptly captures the dynamic regulatory interplay within cells while uniquely incorporating both universal and cell-specific features. This dual approach enables the model to generalize across cell populations while also identifying unique regulatory dynamics within individual cells. Our comprehensive evaluation demonstrates that scGraphETM surpasses existing methodologies in accurately modeling cell-type clustering, cross-modality imputation, and cell-specific TF-RE relationship prediction.

Abrégé

L'avènement des technologies à cellule unique a révolutionné la génomique en permettant l'analyse du matériel génétique à la résolution de cellules individuelles, offrant une perspective granulaire essentielle pour comprendre la diversité et la fonction cellulaires, tant en bonne santé qu'en maladie. Le séquençage de l'ARN unicellulaire (scRNA-seq) et le séquençage ATAC unicellulaire (scATAC-seq) fournissent une vue précise et spécifique au type cellulaire des paysages génomiques et épigénomiques. Ces techniques sont particulièrement précieuses pour étudier les réseaux de régulation génique (GRNs), qui se composent de facteurs de transcription (TFs), d'éléments régulateurs (REs) et de gènes cibles qui contrôlent les processus biologiques au sein des cellules. À l'avant-garde de la recherche multi-omique unicellulaire, le défi d'élucider les GRNs complexes au niveau cellulaire reste primordial en raison de la nature hautement dimensionnelle, bruitée et éparse des données unicellulaires. Pour relever ce défi, nous présentons scGraphETM, une nouvelle approche computationnelle visant à démêler les complexités des GRNs spécifiques aux cellules à partir de données de séquençage multi-omiques unicellulaires.

Le modèle combine de manière innovante un framework d'auto-encodeur variationnel avec un réseau neuronal de graphe, conceptualisant les TFs, les gènes et les REs comme des nœuds, et leurs interactions régulatrices comme des arêtes. Cette structure capture adroitement l'interaction régulatrice dynamique au sein des cellules tout en incorporant de manière unique des caractéristiques universelles et spécifiques aux cellules. Cette double approche permet au modèle de généraliser à travers les populations cellulaires tout en identifiant également les dynamiques régulatrices uniques au sein de

cellules individuelles. Notre évaluation complète démontre que scGraphETM surpasse les méthodologies existantes dans la modélisation précise du clustering de types cellulaires, l'imputation inter-modalité et la prédiction des relations TF-RE spécifiques aux cellulaires.

Acknowledgements

I extend my deepest gratitude to Dr. Yue Li, my supervisor, whose exceptional guidance and unwavering support have been the cornerstone of my master's journey. Your keen insights, scholarly expertise, and remarkable patience transformed challenging moments into opportunities for growth.

To my invaluable collaborators, Manqi Zhou and Boyu Han, I owe profound thanks for your valuable contributions and stimulating discussions. Toast to the countless hours we spent brainstorming, troubleshooting, and refining our approaches.

To my family, who has always believed in and motivated me through every challenge, setback, and triumph. Your support has been my anchor.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Chapter 1

Introduction

The inference of gene regulatory networks (GRNs) represents one of the most significant challenges in computational biology. These networks model the complex interplay between transcription factors (TFs), chromatin accessibility, and target genes that collectively establish, maintain, and potentially disrupt cellular identity. Understanding GRNs has critical implications for engineering cell fate and disease prevention.

Historically, GRN reconstruction has relied on experimentally validated regulation events compiled in databases (Garcia-Alonso et al., 2019; Han et al., 2018) or inferred de novo from gene co-expression patterns in bulk transcriptomics data (Huynh-Thu et al., 2010; Langfelder and Horvath, 2008) However, these approaches face significant limitations. Transcriptomics data alone fail to capture many underlying regulatory mechanisms, such as TF protein abundance, DNA binding events, cooperation between TFs and cofactors, alternative transcript splicing, post-translational modifications, and the accessibility and structure of the genome. Whereas bulk profiling provides mixed measurements across different cell types within a tissue sample, preventing the disentanglement of regulatory programs specific to particular cell types (Cha and Lee, 2020; Fiers et al., 2018). This limitation has been addressed by the advent of single-cell technologies (Klein et al., 2015), which allow for more refined GRN inference across different cell types.

The introduction of multimodal profiling technologies, which simultaneously measure different molecular modalities such as gene expression and chromatin accessibility from the same cell (Chen et al., 2019; Liu et al., 2019; Ma et al., 2020), has led to the development of novel computational methods that can leverage multi-modal data to infer GRNs at unprecedented resolution. These advances provide new opportunities to model gene regulation more accurately, but also present new challenges in data integration, computational scalability, and biological interpretation.

The goal of this thesis is to explore a scalable and interpretable model that integrates established genomic interaction data and known biological relationships. We propose a graph-augmented single-cell embedded topic model that extends the previously published single-cell Embedded Topic Model (Zhao et al., 2021). This enhanced framework offers three interconnected modules that collectively improve GRN inference. Cell-type annotation through multi-omics data integration, which provides the cellular context necessary for identifying cell-type-specific regulatory relationships; missing modalities imputation, which addresses the technical challenge of sparse multi-modal measurements and enables more complete regulatory network reconstruction; and inference of cell-specific gene regulatory networks, which is the ultimate goal of our approach.

1.1 Contribution of Authors

Dr. Yue Li conceived the project concept and provided comprehensive supervision throughout its development, offering critical feedback on model design and conducting weekly progress meetings. Manqi Zhou contributed to downstream analysis approaches and manuscript preparation. Vicky Dong and Boyu Han collaborated on the implementation of the scGraphETM model. Vicky Dong was responsible for performing the experiments and conducting the analyses presented in this thesis.

Chapter 2

Background and Related Works

2.1 Transcriptomics Data

Transcriptomics data explores the geneomc RNA transcripts and reveals how gene expression patterns change in response to developmental processes, environmental stimuli, or disease states (Wang et al., 2009). Bulk RNA-seq typically involves extracting total RNA from a tissue sample or cell population. The sequencing reads are then mapped to a reference genome or transcriptome, allowing quantification of known transcripts and discovery of novel ones (Conesa et al., 2016). Despite its advantages, bulk RNA-seq still faces several challenges: it obscures cell-to-cell variability and potentially misses rare but functionally important cell populations as it measures the average expression across thousands or millions of cells (Stegle et al., 2015). Additionally, technical variations between library preparation batches or sequencing runs can introduce confounding factors that complicate data interpretation (Leek et al., 2010).

Increasingly granular analyses can now capture expression profiles at the level of individual cells. This evolution has been driven by technological advancements that have increased throughput, reduced costs, and improved resolution, enabling researchers to address increasingly complex biological questions (Stark et al., 2019).

2.1.1 Single Cell Sequencing

The recognition of bulk RNA-seq limitations drove the development of single-cell RNA sequencing (**scRNA-seq**), which characterizes transcriptomes at individual cell resolution. Following Tang et al. (2009)'s pioneering work, the field expanded dramatically with droplet-based methods such as Drop-seq (Macosko et al., 2015) and commercial platforms such as 10x Genomics (10x Genomics, 2019), enabling simultaneous profiling of thousands to millions of cells. Despite these advances, scRNA-seq data remains challenging to analyze due to its sparsity, high dimensionality, and susceptibility to technical variations (Haghverdi et al., 2018; Luecken and Theis, 2019).

Complementing transcriptomic analysis, single-cell Assay for Transposase-Accessible Chromatin sequencing (**scATAC-seq**) profiles chromatin accessibility patterns by leveraging Tn5 transposase to insert sequencing adapters into open chromatin regions (Buenrostro et al., 2015). This technique maps the regulatory landscape across diverse cell types, identifying regulatory regions where transcription factors can bind to influence gene expression (Cusanovich et al., 2015).

Multi-omic approaches now enable simultaneous profiling of transcriptomes alongside other molecular features such as chromatin accessibility in the same cells (Zhu et al., 2020). Current multi-omic methods typically profile fewer cells than single-modality approaches, often restricting their application to smaller-scale studies.

The explosive growth of single-cell technologies has led to the establishment of valuable data repositories and resources. The Human Cell Atlas (Regev et al., 2017) aims to create comprehensive reference maps of all human cells, while databases like the Single Cell Portal (Broad Institute), Gene Expression Omnibus (GEO), and cellxgene (Megill et al., 2021) facilitate data sharing and reuse. Furthermore, specialized databases like scATLAS (Franzén and Björkegren, 2019) and CellMarker (Zhang et al., 2019) provide curated collections of cell type-specific markers. These resources, alongside consortium-driven efforts like the NIH Human Biomolecular Atlas Program (HuBMAP) (Consortium, 2019), have dramatically accelerated our understanding of cellular heterogeneity in health

and disease, enabling the identification of previously unknown cell types and states, reconstruction of developmental trajectories, and characterization of cellular responses to perturbations at unprecedented resolution.

2.2 Graph Neural Networks

Graph Neural Networks (**GNN**) represent a widely used class of deep learning architectures specifically designed to work with graph-structured data. Unlike traditional neural networks that operate on sequences, GNNs' ability to model complex relationships makes them particularly well-suited for domains where interactions between entities are central to the problem.

GNNs have demonstrated remarkable effectiveness in diverse genomic applications. In gene regulatory network inference, GNNs model transcription factors as nodes connected to target genes via existing databases, enabling the reconstruction of complex regulatory circuits that control gene expression patterns (Chen et al., 2021; Dutil et al., 2018). Single-cell genomics benefits from GNNs by modeling cells as nodes in a high-dimensional feature space connected by similarity-based edges, facilitating cell type identification, trajectory inference, and gene-gene correlation analysis that captures developmental processes with unprecedented resolution (Wagner et al., 2020; Zeisel et al., 2022).

Protein structure prediction has been revolutionized by Graph Neural Networks (GNNs), which represent amino acids as nodes connected by chemical and spatial proximity edges, modeling complex folding patterns and predicting structural features from sequence data alone (Gainza et al., 2022; Jumper et al., 2021). Cancer genomics applications utilize GNNs to model tumor heterogeneity by constructing patient similarity networks based on multi-omics data, stratifying patients into clinically relevant subtypes and identifying molecular signatures associated with treatment response (Chaudhary et al., 2022; Wang et al., 2021b)

Early work by Scarselli et al. (2009) proposed the original Graph Neural Network framework, which learned node representations through recursive neighborhood aggre-

gation. The field experienced a renaissance with the development of Graph Convolutional Networks (**GCN**) by Kipf and Welling (2017), which simplified earlier approaches and enabled efficient training on large-scale graph data. Their spectral-based approach provided a tractable approximation that balanced expressivity with computational efficiency. Subsequent innovations included GraphSAGE by Hamilton et al. (2017b), which introduced inductive learning capabilities for previously unseen nodes, and Graph Attention Networks (**GAT**) by Veličković et al. (2018), which incorporated attention mechanisms to weight neighbor contributions according to their connections.

GNNs operate on the principle of message passing between nodes in a graph. The key intuition is that a node's representation should incorporate information from its neighborhood. This process typically involves neighborhood aggregation, where each node collects feature information from its neighbors and updates its representation based on its current features and the aggregated neighborhood information. This process repeats across multiple layers, allowing information to propagate through the graph and allowing nodes to capture increasingly broader contextual information.

The general update equation for a node v at layer l can be expressed as:

$$h_v^{(l+1)} = \text{UPDATE}(h_v^{(l)}, \text{AGGREGATE}(\{h_u^{(l)} : u \in \mathcal{N}(v)\}))$$

where $h_v^{(l)}$ is the node representation at layer l, $N(v)$ represents the neighbors of node v, and AGGREGATE and UPDATE are differentiable functions. Different GNN variants implement these functions in various ways, leading to architectures with distinct inductive powers and computational properties.

2.2.1 Graph Convolution Network

Graph Convolutional Networks (GCNs) provide an efficient approximation of spectral graph convolutions. The key innovation of GCNs is their layer-wise propagation rule.

For a given graph with adjacency matrix $\mathbf{A}$ and node feature matrix $\mathbf{X}$, the layer-wise propagation rule of GCN can be written as:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{2.1}$$

where $\mathbf{H}^{(l)}$ is the matrix of node features at layer l (with $\mathbf{H}^{(0)} = \mathbf{X}$). $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix with added self-connections. $\tilde{\mathbf{D}}$ is the degree matrix of $\tilde{\mathbf{A}}$, $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$. $\mathbf{W}^{(l)}$ is the layer-specific trainable weight matrix. σ is a non-linear activation function, such as Rectified Linear Unit (**ReLU**)(Glorot et al., 2011).

The term $\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ performs symmetric normalization of the adjacency matrix, preventing numerical instabilities and exploding or vanishing gradients. The normalization ensures the convolution operation averages features from a node's neighborhood, weighted by the inverse square root of the node degrees.

Per node updates of GCN can be expressed as:

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\mathbf{W}^{(l)} \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}} \mathbf{h}_u^{(l)}\right) \tag{2.2}$$

where $\mathcal{N}(v)$ represents the neighbors of node v, and $\tilde{d}_v$ is the degree of node v in the graph with self-connections.

2.2.2 GraphSAGE

Graph SAmple and aggreGatE (**GraphSAGE**) introduces an inductive framework that enables generating embeddings for previously unseen nodes (Hamilton et al., 2017c). Unlike GCN, which operates on the entire graph simultaneously, GraphSAGE employs a neighborhood sampling strategy that makes it more scalable for large graphs. It learns a function that generates embeddings by sampling and aggregating features from the local neighborhood of the current node. The general update rule for GraphSAGE can be described as such:

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\mathbf{W}^{(l)} \cdot \text{CONCAT}\left(\mathbf{h}_v^{(l)}, \text{AGGREGATE}^{(l)}\left(\{\mathbf{h}_u^{(l)}, \forall u \in \mathcal{N}(v)\}\right)\right)\right) \tag{2.3}$$

where $\mathbf{h}_v^{(l+1)}$ is the feature vector of node v at layer $l+1$. $\mathbf{W}^{(l)}$ is the trainable weight matrix for layer l that transforms the concatenated vector into the output embedding space. $\mathbf{h}_v^{(l)}$ is the current feature vector of node v at layer l. $\{\mathbf{h}_u^{(l)}, \forall u \in \mathcal{N}(v)\}$ is the set of feature vectors from all neighbors u of node v. $\mathcal{N}(v)$ represents the neighborhood of node v. σ is the non-linear activation function, such as ReLU. $\text{AGGREGATE}^{(l)}$ is a differentiable, learnable function that combines the feature vectors from the node's neighbors into a single vector.

Mean Aggregator The mean aggregator takes the element-wise mean of the feature vectors of all neighboring nodes.

$$\text{AGGREGATE}_{\text{mean}}^{(l)} = \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} \mathbf{h}_u^{(l)} \tag{2.4}$$

LSTM Aggregator The LSTM aggregator applies a Long Short-Term Memory neural network to the neighbors' feature vectors (Hochreiter and Schmidhuber, 1997). Compared to the mean aggregator, this approach can capture more complex neighborhood patterns and dependencies.

$$\text{AGGREGATE}_{\text{LSTM}}^{(l)} = \text{LSTM}\left(\{\mathbf{h}_u^{(l)}, \forall u \in \pi(\mathcal{N}(v))\}\right) \tag{2.5}$$

where π denotes a random permutation of the neighbors.

Pooling Aggregator The pooling aggregator applies an element-wise max-pooling operation to the neighborhood features after a non-linear transformation.

$$\text{AGGREGATE}_{\text{pool}}^{(l)} = \max\left(\{\sigma(\mathbf{W}_{\text{pool}}\mathbf{h}_u^{(l)} + \mathbf{b}), \forall u \in \mathcal{N}(v)\}\right) \tag{2.6}$$

where max is applied element-wise across all transformed neighbor features.

After each aggregation step, GraphSAGE normalizes the resulting embeddings:

$$\mathbf{h}_v^{(l+1)} \leftarrow \frac{\mathbf{h}_v^{(l+1)}}{\|\mathbf{h}_v^{(l+1)}\|_2} \tag{2.7}$$

L2 normalization applied helps to prevent numerical instabilities during training.

2.2.3 Graph Attention Network

Graph Attention Networks (GATs), introduced by Veličković et al. (2018), incorporate attention mechanisms into GNNs, allowing the model to focus on the most relevant parts of the neighborhood when aggregating information. Unlike GCN, which weights neighbors based on graph structure alone, GAT learns to assign different importances to different neighbors, while they may have the same structural relationship to the target node. The attention mechanism framework popularized by Transformer architectures (Vaswani et al., 2017) is applied in GAT:

$$\mathbf{Q}_v = \mathbf{W}_Q \mathbf{h}_v \quad \text{(transformed features of the target node)} \tag{2.8}$$

$$\mathbf{K}_u = \mathbf{W}_K \mathbf{h}_u \quad \text{(transformed features of the neighbor node)} \tag{2.9}$$

$$\mathbf{V}_u = \mathbf{W}_V \mathbf{h}_u \quad \text{(transformed features to be aggregated)} \tag{2.10}$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are learnable parameter matrices.

The attention score between a target node v and its neighbor u is computed as:

$$e_{vu} = \frac{\mathbf{Q}_v^T \mathbf{K}_u}{\sqrt{d}} \tag{2.11}$$

where d is the dimension of the key vectors and serves as a scaling factor. This dot product measures how well the query matches the key.

The attention coefficient is obtained by normalizing across all neighbors using the softmax function:

$$\alpha_{vu} = \text{softmax}_u(e_{vu}) = \frac{\exp(e_{vu})}{\sum_{k \in \mathcal{N}(v)} \exp(e_{vk})} \tag{2.12}$$

The final output feature vector for node v is calculated as a weighted sum of the value vectors:

$$\mathbf{h}'_v = \sum_{u \in \mathcal{N}(v)} \alpha_{vu} \mathbf{V}_u \tag{2.13}$$

In practice, GATs typically employ multi-head attention where K independent attention mechanisms are computed in parallel:

$$\mathbf{h}'_v = \|_{k=1}^{K} \sigma \left(\sum_{u \in \mathcal{N}(v)} \alpha_{vu}^k \mathbf{W}^k \mathbf{h}_u \right) \tag{2.14}$$

where $\|$ represents concatenation and α_{vu}^k is the attention coefficient from the k-th attention head.

Despite their success, GNNs face several challenges. Scalability remains a significant concern when processing large-scale graphs with billions of nodes and edges, requiring efficient sampling and partitioning strategies that balance computational efficiency with representational accuracy (Hu et al., 2020a). Heterogeneity presents another challenge, as real-world graphs often contain different types of nodes and edges with varying properties and semantics, necessitating model architectures that can accommodate this diversity without losing structural information (Hu et al., 2020b).

2.3 Topic Models

Topic models have emerged as unsupervised techniques for discovering hidden thematic structures in large document collections. These models operate on the principle that doc-

uments are mixtures of topics, where each topic is characterized by a probability distribution over words. The primary aim of topic modeling is to discover these latent semantic patterns automatically. Topics serve as underlying hidden variables that generate the observed word distributions in documents, creating a probabilistic framework for understanding document content. They provide great interpretability and demonstrate remarkable versatility in successful applications across diverse domains, including single-cell genomics (Zhao et al., 2021).

2.3.1 Latent Dirichlet Allocation (LDA)

Introduced by Blei et al. (2003), Latent Dirichlet Allocation (**LDA**) is a generative probabilistic model that represents documents as random mixtures over latent topics, where each topic is characterized by a distribution over words.

Algorithm 1: Latent Dirichlet Allocation (LDA) Pseudocode

Input: Document corpus D, number of topics K, hyperparameters α and β

Output: Topic-word distributions ϕ, document-topic distributions θ

LDA Generative Process: for *each topic* $k = 1$ *to* K **do**

  Sample topic-word distribution $\phi_k \sim \text{Dirichlet}(\beta)$;

end

for *each document* d *in corpus* **do**

  Sample document-topic distribution $\theta_d \sim \text{Dirichlet}(\alpha)$;

  for *each word position* i *in document* d **do**

    Sample topic $z_{di} \sim \text{Multinomial}(\theta_d)$;

    Sample word $w_{di} \sim \text{Multinomial}(\phi_{z_{di}})$;

  end

end

In Latent Dirichlet Allocation, model's behavior is governed by Dirichlet and Multinomial distribution:

The topic-word distribution ϕ_k represents the probability of each word in the vocabulary occurring in topic k. Mathematically, each ϕ_k is a multinomial distribution over the vocabulary, where $\phi_{k,w}$ represents the probability of word w appearing in topic k. These distributions characterize what each topic is "about".

In the generative process, each topic-word distribution is drawn from a Dirichlet prior with hyperparameter β:

$$\phi_k \sim \text{Dirichlet}(\beta) \tag{2.15}$$

The Dirichlet prior encourages sparsity in the distribution when $\beta < 1$, which aligns with the intuition that each topic should focus on a subset of the vocabulary rather than uniformly distributing probability across all words. Typically, β is set to a small value to encourage sparsity.

The document-topic distribution θ_d represents the mixture of topics present in document d. Similarly, each θ_d is a multinomial distribution over the K topics, where $\theta_{d,k}$ represents the proportion of document d that belongs to topic k which captures the intuition that documents typically cover multiple topics in varying proportions.

In the generative process, each document-topic distribution is drawn from a Dirichlet prior with hyperparameter α:

$$\theta_d \sim \text{Dirichlet}(\alpha) \tag{2.16}$$

A small hyperparameter α encourages documents to focus on a few dominant topics, while a larger α allows for more uniform topic mixtures. The choice of α reflects assumptions about how topics are distributed within documents in the corpus.

2.3.2 Collapsed Gibbs Sampling for LDA

Collapsed Gibbs sampling is an efficient Markov Chain Monte Carlo (MCMC) method for inferring the latent topic assignments in LDA (Porteous et al., 2008). The algorithm begins by randomly assigning topics to each word occurrence in the corpus and iteratively calculates the conditional probability of assigning each possible topic to the word and then

updating count statistics. After many iterations, the Markov chain approaches its stationary distribution, which approximates the true posterior distribution of topic assignments given the observed words.

The power of Gibbs sampling lies in its ability to explore the complex posterior distribution of topic assignments without having to directly compute this distribution.

Algorithm 2: Collapsed Gibbs Sampling

LDA Inference via Collapsed Gibbs Sampling:

Initialize topic assignments z_{di} randomly for all words in all documents;

repeat

  for *each document* d *in corpus* **do**

    for *each word position* i *in document* d **do**

      Remove current topic assignment z_{di} from counts;

      Calculate $P(z_{di} = k|\text{all other } z, w) \propto \frac{n_{d,k}^{-di}+\alpha}{n_d^{-di}+K\alpha} \cdot \frac{n_{k,w_{di}}^{-di}+\beta}{n_k^{-di}+V\beta}$;

      Sample new topic $z_{di} \sim P(z_{di} = k|\text{all other } z, w)$;

      Update counts based on new z_{di};

    end

  end

until *convergence or maximum iterations reached*;

for *each topic* $k = 1$ *to* K **do**

  for *each word* w *in vocabulary* **do**

    $\phi_{k,w} = \frac{n_{k,w}+\beta}{n_k+V\beta}$;

  end

end

for *each document* d *in corpus* **do**

  for *each topic* $k = 1$ *to* K **do**

    $\theta_{d,k} = \frac{n_{d,k}+\alpha}{n_d+K\alpha}$;

  end

end

return ϕ, θ;

2.3.3 Embedded Topic Model (ETM)

The Embedded Topic Model (**ETM**), proposed by Dieng et al. (2020), addresses several limitations of traditional topic models by incorporating word embeddings. ETM rep-

resents each topic as topic embeddings, leveraging the semantic relationships captured by pre-trained word embeddings. ETM demonstrates improved performance on short documents by utilizing semantic information beyond simple co-occurrence. Its use of amortized variational inference enables faster computation than traditional sampling approaches.

Algorithm 3: Embedded Topic Model (ETM)

Input: Document corpus D, number of topics K, hyperparameter α, embedding dimension L, word embeddings $\rho \in \mathbb{R}^{V \times L}$

Output: Topic embeddings α, document-topic distributions θ

for *each topic* $k = 1$ *to* K **do**

  Define topic embedding $\alpha_k \in \mathbb{R}^L$;

end

for *each document* d *in corpus* **do**

  Sample document-topic distribution $\theta_d \sim \text{Dirichlet}(\alpha)$;

  for *each word position* i *in document* d **do**

    Sample topic $z_{di} \sim \text{Multinomial}(\theta_d)$;

    Sample word w_{di} with probability $\propto \exp(\rho_w^T \alpha_{z_{di}})$;

  end

end

Algorithm 4: ETM Inference via Variational Method

Initialize neural network parameters ϕ_{enc} for encoder network;

Initialize topic embeddings $\alpha_k \in \mathbb{R}^L$ for each topic k;

for *each iteration* **do**

  Sample a batch of documents D_{batch};

  for *each document* $d \in D_{batch}$ **do**

    $(\mu_d, \log \sigma_d^2) = \text{Encoder}_{\phi_{\text{enc}}}(d)$;

    $\epsilon \sim \mathcal{N}(0, \mathbf{I})$;

    $\eta_d = \mu_d + \sigma_d \odot \epsilon$;

    $\theta_d = \text{softmax}(\eta_d)$;

    for *each topic* $k = 1$ *to* K **do**

      for *each word* w *in vocabulary* **do**

        $\beta_{kw} = \frac{\exp(\rho_w^T \alpha_k)}{\sum_{w'} \exp(\rho_{w'}^T \alpha_k)}$;

      end

    end

    $\mathcal{L}_d = \mathbb{E}_{q(\theta_d)}[\log p(w_d|\theta_d, \alpha, \rho)] - \text{KL}(q(\theta_d)||p(\theta_d))$;

  end

end

for *each document* d **do**

  Compute $(\mu_d, \log \sigma_d^2) = \text{Encoder}_{\phi_{\text{enc}}}(d)$;

  Set document-topic distribution $\theta_d = \text{softmax}(\mu_d)$;

end

return α, θ;

In ETM, each topic k is represented as an embedding vector α_k. The probability of word w under topic k is defined as:

$$\beta_{kw} = \frac{\exp(\rho_w^T \alpha_k)}{\sum_{w'} \exp(\rho_{w'}^T \alpha_k)} \tag{2.17}$$

where ρ_w is the embedding of word w. The document-topic distributions are inferred using an encoder network that maps a document to the parameters of a variational distribution over the document-topic proportions. The goal of inference is to estimate the posterior distribution $p(\theta|w)$, the distribution of topic proportions given the observed words. Since this posterior is intractable, ETM approximates it with a variational distribution $q(\theta|w)$.

In ETM, this variational distribution is parameterized by a neural network (the encoder), which takes a document as input and outputs the parameters of a logistic normal distribution μ_d and $\log \sigma_d^2$. These parameters define a normal distribution in a transformed space and the topic proportions θ_d are obtained by applying the softmax function. The training objective in ETM is to maximize the Evidence Lower Bound (**ELBO**):

$$\text{ELBO} = \mathbb{E}[\log p(w|\theta, \alpha, \rho)] - \text{KL}(q(\theta) \parallel p(\theta)) \tag{2.18}$$

The Expected log-likelihood, $\mathbb{E}[\log p(w|\theta, \alpha, \rho)]$ measures how well the model reconstructs the observed words given the inferred topics. For each word, the probability is computed as:

$$p(w|\theta) = \sum_k p(w|\text{topic } k) \times p(\text{topic } k|\theta) = \sum_k \beta_{kw} \times \theta_k \tag{2.19}$$

Where the topic-word distribution is defined using word embeddings (ρ) and topic embeddings (α):

$$\beta_{kw} = \frac{\exp(\rho_w^\top \alpha_k)}{\sum_{w'} \exp(\rho_{w'}^\top \alpha_k)} \tag{2.20}$$

The KL divergence, $\text{KL}(q(\theta) \parallel p(\theta))$ is a regularization term that penalizes the variational distribution for being too far from the prior, which is typically a Dirichlet distribution, and prevents the model from overfitting.

2.4 Related Works

2.4.1 Embedded Topic Models in Single Cell Application

The Single-cell Embedded Topic Model (scETM) method introduces a neural-network-based embedded topic modeling approach specifically designed for single-cell RNA sequencing data analysis (Zhao et al., 2021). The model assumes that each cell possesses a topic mixture following a logistic-normal distribution. scETM employs an encoder neural network that maps normalized gene expression to parameters of a Gaussian distribution in latent space, which is then transformed via softmax to obtain the cell's topic mixture. Prior to modeling, highly variable genes are selected to improve computational efficiency and focus the model on biologically informative signals.

2.4.2 Muti-Omics Integration Methods

The emergence of single-cell multimodal assays has created a powerful means of examining multiple facets of cellular states. A major challenge in analyzing these data is devising effective strategies to integrate information from different modalities (Li et al., 2025). Data integration encompasses three primary approaches: horizontal, vertical, and diagonal integration. Horizontal integration uses genomic features as the anchor when merging the same data modality from different cells, for instance, combining scRNA-seq datasets across multiple experimental batches. In vertical integration, cells serve as the anchor when multiple modalities are measured from the same individual cells, such as simultaneously profiling RNA expression and chromatin accessibility within a single cellular sample. Diagonal integration bridges individual scRNA-seq and scATAC-seq experiments where both cells and features exhibit significant variations between datasets, presenting the most complex integration challenge (Argelaguet et al., 2020b). The choice of integration strategy significantly impacts downstream analyses and the biological insights that can be derived from the data.

Mutual Nearest Neighbors (**MNN**), introduced by Haghverdi et al. (2018), represented a significant advancement in addressing batch effects in single-cell RNA sequencing data. MNN performs horizontal integration via identifying pairs of cells that are mutual nearest neighbors in a low-dimensional space by principal components. The algorithm then computes batch correction vectors for each MNN pair and applies them locally, preserving biological heterogeneity while removing technical variation. This local correction approach maintains the underlying biological structure and is particularly useful when cell type proportions vary between batches. A key drawback of MNN is its susceptibility to overcorrection when datasets lack common biological states, potentially forcing alignment where none should exist.

Seurat v3 is an anchor-based integration method that leverages canonical correlation analysis (**CCA**) to align datasets across batches (Stuart et al., 2019). Seurat v3 identifies anchors as mutual nearest neighbors in CCA space, representing pairs of cells in different datasets that likely originate from the same biological state. These anchors then guide a weighted transformation process that corrects each cell's expression values based on multiple relevant anchors. The method incorporates anchor scoring to filter potentially incorrect pairs and provides corrected expression values for downstream analysis. In comparative benchmarking by Luecken et al. (2022), Seurat v3 consistently ranks among the top performers for scRNA-seq integration.

Seurat v4 extended the framework to address multimodal single-cell data integration (Hao et al., 2021b). The core innovation was the Weighted Nearest Neighbor (WNN) analysis, which provides a strategy for vertical integration. Unlike previous approaches that created a single integrated representation, Seurat v4 preserves modality-specific information while enabling joint analysis. WNN maintains separate low-dimensional embeddings for each data modality and constructs a weighted cell-cell similarity graph that combines information across modalities. The method dynamically learns modality weights. A key drawback of WNN is its requirement for matched measurements across modalities, limiting its application to vertical integration tasks.

Lopez et al. (2018a) introduced scVI to address horizontal integration through a probabilistic formulation that explicitly accounts for batch-specific variation. The model utilizes a VAE architecture to learn a nonlinear embedding of the data where batch effects are removed while preserving biological signal. scVI employs a zero-inflated negative binomial (**ZINB**) distribution for its underlying generative process to model the sparsity nature of single-cell data.

Harmony is an iterative and linear clustering method operating in principal component analysis (**PCA**) space for single-cell data integration (Korsunsky et al., 2019). Harmony employs a mixture model clustering approach with a diversity penalty that encourages clusters to contain cells from all batches, followed by a linear correction of PC coordinates within each cluster. The method alternates between these clustering and correction steps until convergence to a harmonized embedding. Harmony demonstrates excellent computational efficiency, scaling linearly with cell number and enabling integration of millions of cells within minutes. A key drawback of Harmony is its inability to correct the full gene expression matrix directly, limiting certain downstream analyses that require corrected expression values.

Multi-Omics Factor Analysis+ (**MOFA+**) is an unsupervised Bayesian framework for integrating multimodal single-cell data (Argelaguet et al., 2020a). MOFA+ employs a matrix factorization approach that decomposes multiple data matrices into a set of latent factors and modality-specific weights, capturing both shared and modality-specific sources of variation. The method utilizes structured sparsity priors to identify which factors are active in which data modalities, providing interpretable outputs. MOFA+ handles missing data naively and accommodates different likelihood models appropriate for diverse data types. In applications to multimodal data from mouse gastrulation, MOFA+ successfully identified coordinated epigenetic and transcriptional changes associated with lineage commitment. A key drawback of MOFA+ is its assumption of linear relationships between latent factors and observed features.

Multi-Omics Embedded Topic Model (moETM) is a probabilistic framework for integrating multiple single-cell modalities that builds upon topic modeling concepts (Zhou et al., 2023). moETM uses a hierarchical Bayesian approach where cells are represented as mixtures of topics, and each topic represents a distribution over features in each modality. The method employs a negative binomial distribution for RNA counts and a logistic normal distribution for chromatin accessibility, linked through a shared latent space. Information from different modalities is combined through a product-of-experts formulation.

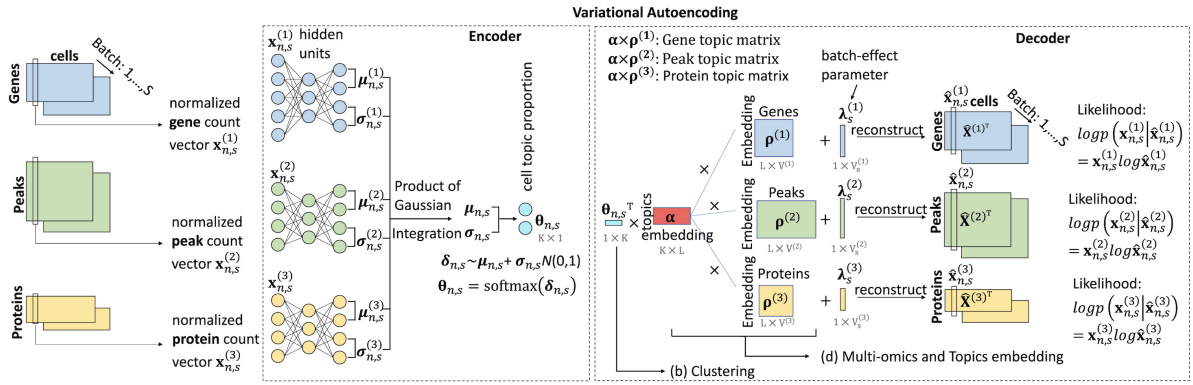

Figure 2.1: MoETM Model Overview (Zhou et al., 2023)

2.4.3 Cross-Modality Imputation Methods

Simultaneous profiling of multiple molecular modalities within a single cell represents one of the most formidable challenges in modern genomics (Wu et al., 2021). This approach faces several fundamental obstacles such as molecular incompatibility, where measurement techniques for different modalities may be mutually exclusive or compounded technical noise and dropout effects that compromise data quality. Single-cell cross-modality imputation methods have emerged as a critical solution to the challenge of simultaneous multimodal profiling. By leveraging the inherent relationships between different omics, imputation methods can computationally predict unmeasured modalities from measured ones (Li et al., 2025).

BABEL is a deep learning framework designed to enable cross-modality translation between scRNA-seq and scATAC-seq data (Wu et al., 2021). The method employs a variational autoencode (**VAE**) architecture. BABEL consists of four modular neural networks, an RNA encoder that projects RNA expression profiles into a shared latent space, an ATAC encoder that projects ATAC accessibility profiles into the same shared latent space, an RNA decoder that infers RNA expression from the input latent representation and an ATAC decoder that infers ATAC accessibility from the input latent representation. For the RNA-to-ATAC translation, the encoder maps RNA expression data x_{RNA} to a latent representation z through a probabilistic encoding:

$$q_\phi(z|x_{\text{RNA}}) = \mathcal{N}(z|\mu_\phi(x_{\text{RNA}}), \sigma^2_\phi(x_{\text{RNA}})) \tag{2.21}$$

Where μ_ϕ and σ^2_ϕ are neural networks parameterized by ϕ that predict the mean and variance of the latent distribution.

The decoder then reconstructs ATAC accessibility data from this latent representation:

$$p_\theta(x_{\text{ATAC}}|z) = \text{Bern}(x_{\text{ATAC}}|\pi_\theta(z)) \tag{2.22}$$

Where π_θ is a neural network parameterized by θ that outputs the probability of each genomic region being accessible.

Similarly, for ATAC-to-RNA translation, the encoding process is:

$$q_\psi(z|x_{\text{ATAC}}) = \mathcal{N}(z|\mu_\psi(x_{\text{ATAC}}), \sigma^2_\psi(x_{\text{ATAC}})) \tag{2.23}$$

And the decoding to RNA expression is modeled as:

$$p_\omega(x_{\text{RNA}}|z) = \text{NB}(x_{\text{RNA}}|\mu_\omega(z), \theta_\omega(z)) \tag{2.24}$$

Where NB represents a negative binomial distribution with mean μ_ω and dispersion θ_ω, both functions of the latent representation z.

BABEL's training objective comprises four components, encompassing all possible encoder-decoder combinations:

$$L = L_{NB}(r, r_{RNA}) + \beta L_{BCE}(a, a_{ATAC}) + \beta L_{BCE}(a, a_{RNA}) + L_{NB}(r, r_{ATAC}) \tag{2.25}$$

Where L_{NB} is the negative binomial loss for RNA predictions. L_{BCE} is the binary cross-entropy loss for ATAC predictions. β is a balancing constant.

The negative binomial loss, previously shown useful in Lopez et al. (2018b) is defined as:

$$\begin{aligned} L_{NB}(y; \hat{y}, \theta) = -\theta(\log(\theta + \epsilon) - \log(\theta + \hat{y})) - y(\log(\hat{y} + \epsilon) - \log(\theta + \hat{y})) \\ - \log \Gamma(y + \theta) + \log \Gamma(y + 1) + \log \Gamma(\theta + \epsilon) \end{aligned} \tag{2.26}$$

Where ϵ is a small constant for numerical stability.

The binary cross-entropy loss for ATAC predictions is (Xiong et al., 2019):

$$L_{BCE}(x; \hat{x}) = -(x \log \hat{x} + (1 - x) \log(1 - \hat{x})) \tag{2.27}$$

Where x represents the measured ATAC signal and $\hat{x}$ is the model's prediction.

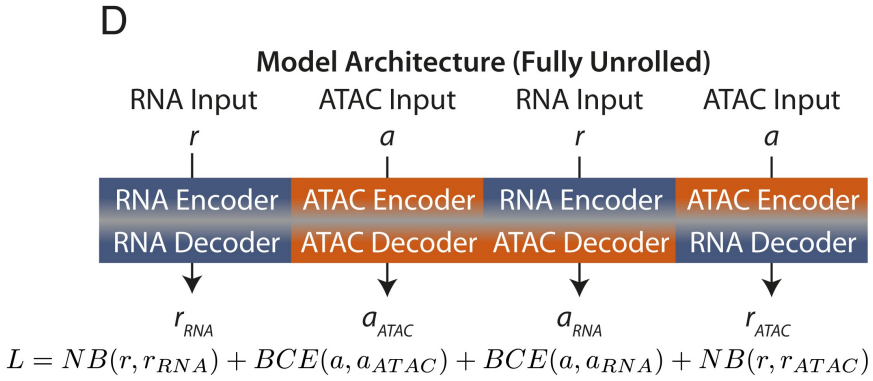

Figure 2.2: Babel Loss Overview

BABEL was trained on cells from peripheral blood mononuclear cells (PBMCs), colon adenocarcinoma cells, and colorectal adenocarcinoma cells (10x Genomics, 2019). While it can generalize to new contexts once trained, BABEL requires substantial amounts of paired multimodal data for training.

MultiVI also builds on (VAE) architectures to integrate different single-cell modalities (Ashuach et al., 2023). Similar to BABEL, RNA-seq data is modeled using a negative binomial distribution, ATAC-seq data is modeled with a Bernoulli distribution, but additionally protein expression is modeled as a mixture of negative binomial distributions. Initially, each modality is assigned its own latent representation, an isotropic multivariate normal distribution. These are combined to create a unified representation via the default approach by taking the average or learnable cell-specific weights across modalities.

Using the variational approximation, the evidence lower bound is computed and optimized. To enforce the similarity between chromatin and transcription latent representations, the model penalized the distance between representations using a symmetric Jeffrey's divergence between distributions:

$$d(Z_c^{\mathrm{A}}, Z_c^{\mathrm{R}}) = \mathrm{symmKL}(q(z_c^{\mathrm{A}}), q(z_c^{\mathrm{R}})) = \mathrm{KL}(q(z_c^{\mathrm{A}}), q(z_c^{\mathrm{R}})) + \mathrm{KL}(q(z_c^{\mathrm{R}}), q(z_c^{\mathrm{A}})). \tag{2.28}$$

In the case of three or more distributions:

$$\begin{aligned} d(Z_c^{\mathrm{A}}, Z_c^{\mathrm{R}}, Z_c^{\mathrm{P}}) =& \mathrm{symmKL}(q(z_c^{\mathrm{R}}), q(z_c^{\mathrm{A}})) + \mathrm{symmKL}(q(z_c^{\mathrm{R}}), q(z_c^{\mathrm{P}})) + \\ & \mathrm{symmKL}(q(z_c^{\mathrm{A}}), q(z_c^{\mathrm{P}})). \end{aligned} \tag{2.29}$$

2.4.4 Gene Regulatory Network Inference Methods

The complex interactions among chromatin structure, transcription factors, and genetic elements create intricate regulatory systems named Gene Regulatory Network (**GRN**). These networks provide valuable insights into the mechanisms that establish, maintain, and potentially disrupt cellular identity during disease. The recent advancement of single-

cell multi-omics technologies has revolutionized GRN inference methodologies which were traditionally constructed using either published literature or bulk omics experimental data (Badia-i Mompel et al., 2023). These novel computational approaches simultaneously analyze genomic sequences, transcriptional activity, and chromatin accessibility patterns.

Typical GRN inference workflow involves rigorous pre-processing of expression data to construct comprehensive interaction matrices, identifying known transcription factors (**TF**), and employing advanced predictive modeling techniques to elucidate potential TF-gene interactions to synthesize a comprehensive network representation that captures the complex regulatory relationships governing gene expression. For chromatin accessibility data, the approach involves pre-processing to create accessibility matrices, associating cis-regulatory elements (CREs) with nearby genes, predicting TF binding to CREs using motif databases, generating TF-CRE-gene triplets. When utilizing multi-omics data, transcriptomics and chromatin accessibility are first preprocessed separately, and then integrated to simultaneously build a more complete GRN as shown in fig.2.3.

The GEne Network Inference with Ensemble of trees (**GENIE3**) algorithm has emerged as a powerful approach for GRN inference, winning the DREAM4 In Silico Multifactorial network inference challenge (Huynh-Thu et al., 2010). GENIE3 employs random forest models to predict the expression of each gene in the dataset using the expression of transcription factors as input. For each gene j, GENIE3 attempts to identify which transcription factor genes are the most important for predicting its expression with the the ensemble of trees aiming to minimize the following error:

$$\sum_{k=1}^{N}(x_j^k - f_j(x_{-j}^k))^2 \tag{2.30}$$

For a single tree, the importance of an input variable gene i is computed as:

$$I(N) = \#S \cdot \text{Var}(S) - \#S_t \cdot \text{Var}(S_t) - \#S_f \cdot \text{Var}(S_f) \tag{2.31}$$

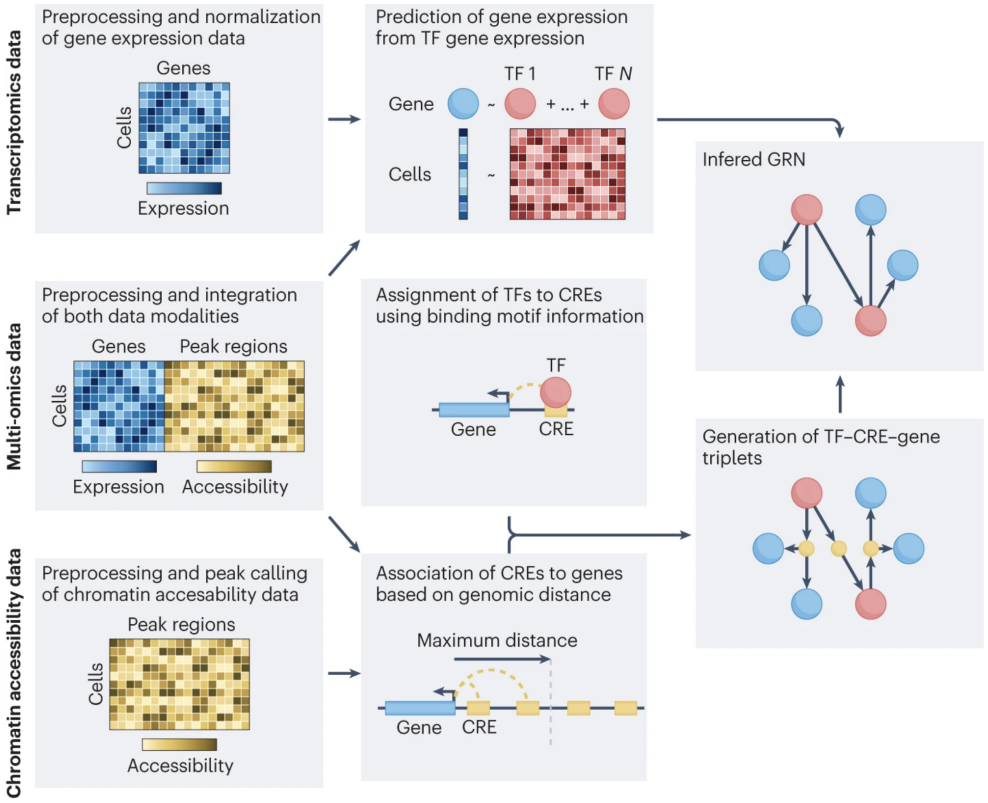

Figure 2.3: Flow chart of methods for GRN inference (Badia-i Mompel et al., 2023)

where N is a node in the tree where gene i is used for splitting. S is the set of samples that reach node N. S_t and S_f are the subsets of samples for which the test at node N is true or false, respectively.# denotes the cardinality of a set, and Var$(\cdot)$ is the variance of the output values in a set. The overall importance of gene i in predicting the expression of gene j is computed by summing the I values over all tree nodes where gene i is used for splitting, and averaging over all trees in the ensemble. Regulatory links are ranked according to these importance scores.

Single-Cell Regulatory Network Inference and Clustering (**SCENIC**) aims to solve the gene regulatory network reconstruction problem along with cell state identification through a three-stage workflow (Aibar et al., 2017). The first step uses GENIE3 to identify potential transcription factor (TF) targets based on co-expression patterns. For larger datasets, SCENIC implements GRNBoost as a scalable alternative to GENIE3. GRNBoost uses gradient boosting machines instead of random forests. In the second step, RcisTar-

get performs motif enrichment analysis on previously identified co-expression modules to distinguish direct binding targets from indirect relationships (Herrmann et al., 2012). SCENIC finally calculates cell level regulon activatity via AUCell. The output of the AUCell step is a matrix containing the AUC score for each regulon in each cell representing the proportion of expressed genes in the regulon and their relative expression compared to other genes within the cell. This activity matrix served as the input for cell clustering based on shared regulatory network activity rather than raw gene expression.

SCENIC+ integrates chromatin accessibility data alongside gene expression measurements to construct more comprehensive enhancer-driven gene regulatory networks (eGRNs) (Bravo González-Blas et al., 2023). The SCENIC+ workflow also follows a three step process. First, it identifies candidate enhancers by analyzing cell-specific chromatin accessibility patterns using pycisTopic, which detects both differentially accessible regions and sets of co-accessible regions. Second, pycisTarget predicts transcription factor binding sites within these accessible regions using an extensive collection of over 30,000 motifs that have been carefully curated and clustered (Imrichová et al., 2015). Finally, SCENIC+ employs GRNBoost2 to quantify the importance of both transcription factors and enhancer regions for target genes combined with motif enrichment analysis to determine the optimal transcription factor for each set of motifs (Moerman et al., 2019).

Graph-Linked Unified Embedding (**GLUE**) is a knowledge- guided Bayesian framework designed for integrating unpaired single-cell multi-omics data and inferring gene regulatory networks (Cao and Gao, 2022). GLUE addresses the challenge of diagonal integration by learning feature embeddings refined to reconstruct both a guidance graph and the single-cell multi-omics data simultaneously. The cosine similarities between these feature embeddings are then used as regulatory scores that reflect the strength of regulatory relationships.

Assuming that there are K different omics layers to be integrated, each with a distinct feature set V_k, $k = 1, 2, \ldots, K$. For instance, in scRNA-seq, V_k is the set of genes, while in scATAC-seq, V_k is the set of accessible chromatin regions.

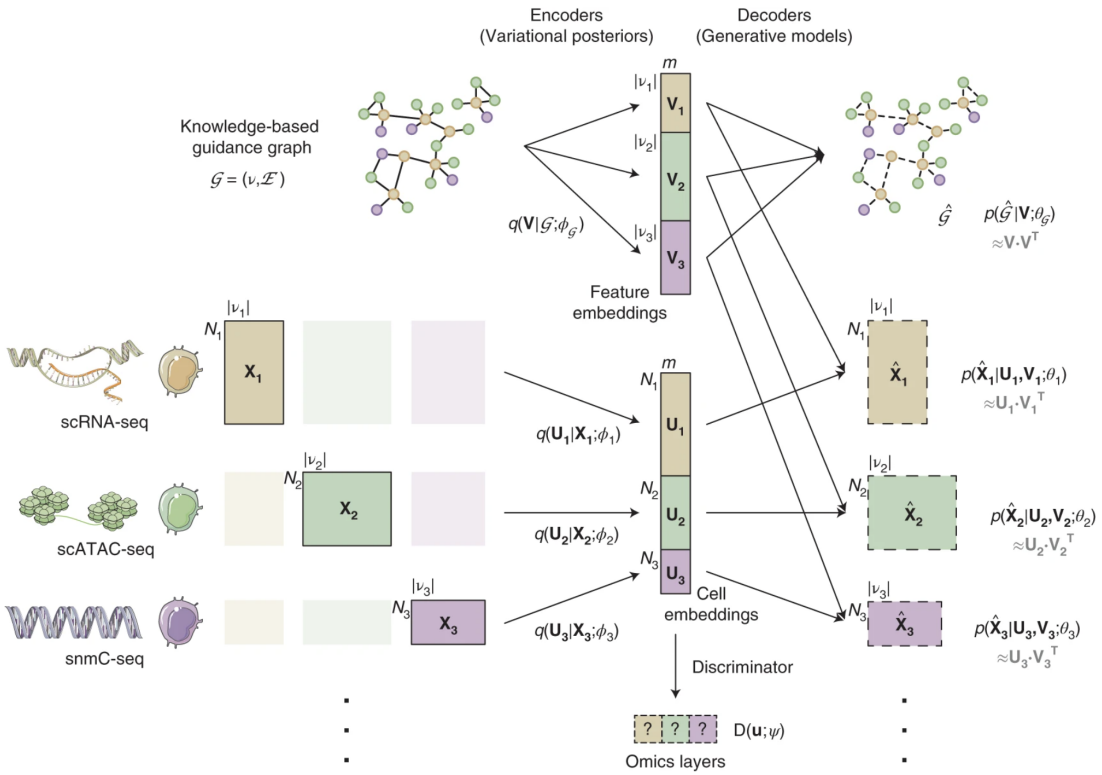

Figure 2.4: GLUE Model Overview (Cao and Gao, 2022)

The data spaces of different omics layers are denoted as $X_k \subseteq \mathbb{R}^{|V_k|}$ with varying dimensionalities. We use $x_k^{(n)} \in X_k$, $n = 1, 2, \ldots, N_k$ to denote cells from the k-th omics layer, and $x_{ki}^{(n)}$, $i \in V_k$ to denote the observed value of feature i of the k-th layer in the n-th cell. Notably, the cells from different omics layers are unpaired and can have different sample sizes. GLUE modeled the observed data from different omics layers as generated by a low-dimensional latent variable, cell embedding $u \in \mathbb{R}^m$:

$$p(x_k; \theta_k) = \int p(x_k|u; \theta_k)p(u)du \tag{2.32}$$

where $p(u)$ is the prior Gaussian distribution of the latent variable. $p(x_k|u; \theta_k)$ are learnable generative distributions in the data decoders. θ_k denotes learnable parameters in the decoders. The cell latent variable u is shared across different omics layers, representing the common cell states.

GLUE uses a guidance graph $G = (V, E)$, which incorporates prior knowledge about regulatory interactions between features across the omics to link different feature spaces. Each edge is associated with signs and weights, denoted as s_{ij} and w_{ij}, respectively, $w_{ij} \in (0, 1]$ represents interaction credibility and $s_{ij} \in \{-1, 1\}$ specifies the sign of the regulatory interaction. The guidance graph is treated as an observed variable and is modeled as generated by low-dimensional feature latent variables $v_i \in \mathbb{R}^m$, $i \in V$. The combined feature embeddings matrix is denoted as $V \in \mathbb{R}^{m \times |V|}$. The GLUE framework employs factorized variational inference, where the joint posterior of cell and feature embeddings is approximated as the product of a data-specific encoder and a GCN encoder.

To properly align the various omics layers, GLUE uses adversarial alignment strategy. A discriminator D with a K-dimensional softmax output predicts the omics layers of cells based on their embeddings u with cell-specific weights $w^{(n)}$ incorporated to handle imbalanced cell type compositions.

$$\mathcal{L}_D(\phi, \psi) = -\frac{1}{K} \sum_{k=1}^{K} \frac{1}{W_k} \sum_{n=1}^{N_k} w^{(n)} \cdot \mathbb{E}_{u \sim q(u|x_k^{(n)}; \phi_k)} \log D_k(u; \psi) \tag{2.33}$$

where D_k represents the k-th dimension of the discriminator output, and ψ represents learnable parameters in the discriminator.

At last, the overall training objective of GLUE consists of:

$$\min_{\psi} \lambda_D \cdot \mathcal{L}_D(\phi, \psi) \tag{2.34}$$

$$\max_{\theta, \phi} \lambda_D \cdot \mathcal{L}_D(\phi, \psi) + \lambda_G K \cdot \mathcal{L}_G(\theta_G, \phi_G) + \sum_{k=1}^{K} \mathcal{L}_{X_k}(\theta_k, \phi_k, \phi_G) \tag{2.35}$$

where λ_D scales the contribution of adversarial alignment and λ_G scales the contribution of graph-based feature embedding. Training is performed using stochastic gradient descent with the discriminator being updated according to Equation 2.34. Then the encoders and decoders are updated according to Equation 2.35.

Lifelong neural network for gene regulation (**LINGER**) is a deep learning framework that utilizes atlas-scale external bulk data across diverse cellular contexts and prior knowledge of TF motifs to improve GRN inference accuracy (Yuan and Duren, 2024). LINGER first pretrains on paired 201 samples of bulk RNA-seq and ATAC-seq data obtained from the ENCODE project (Consortium et al., 2012). For each gene, a neural network predicts the gene expression based on TF expression and chromatin accessibility within 1 Mb of the transcription starter site (**TSS**). The method undergoes comprehensive validation to assess different aspects of its regulatory predictions using groundtruth data including Chromatin Immunoprecipitation sequencing (**ChIP-seq**) data for trans-regulatory interactions and expression Quantitative Trait Loci (**eQTL**) data for cis-regulatory predictions.

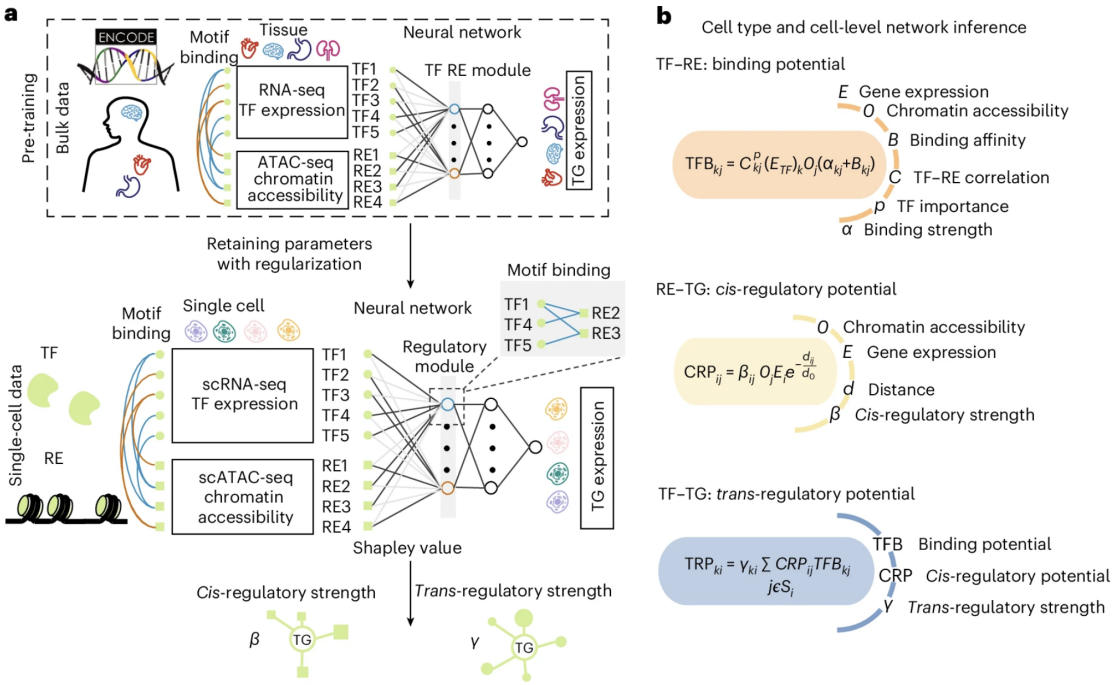

Figure 2.5: LINGER Model Overview (Yuan and Duren, 2024)

The complete LINGER loss function consists of four components:

$$\mathcal{L}_{\text{LINGER}} = \lambda_1 \mathcal{L}_{\text{MSE}} + \lambda_2 \mathcal{L}_{\text{L1}} + \lambda_3 \mathcal{L}_{\text{Laplace}} + \lambda_4 \mathcal{L}_{\text{EWC}} \tag{2.36}$$

The mean squared error (**MSE**) quantifies how accurately the model predicts target gene expression. L1 regularization helps identify the most relevant transcription factors and regulatory elements influencing the target gene expression, while eliminating noise from less important features. The Laplacian loss (Manifold Regularization) incorporates TF-RE motif matching knowledge by encouraging parameters in the first hidden layer to form regulatory modules that align with biological knowledge.

$$\mathcal{L}_{\text{Laplace}}(E_{\text{TF}}, O, E_{i,\cdot}, \theta_l) = \text{tr}((\theta_l^{(1)})^T L_{\text{Norm}} \theta_l^{(1)}) \tag{2.37}$$

where L_{Norm} is the normalized Laplacian matrix based on TF-RE binding affinity, and $\theta_l^{(1)}$ are the parameters of the first hidden layer.

The elastic weight consolidation (**EWC**) Loss penalizes large deviations from previously learned bulk data, preserving lifelong learning. The Fisher information matrix assigns higher importance to influential parameters in the bulk model.

$$\mathcal{L}_{\text{EWC}}(E_{\text{TF}}, O, E_{i,\cdot}, \theta_l) = \frac{1}{(N_{\text{TF}} + N_{\text{RE}}) \times 64} \sum_{i=1}^{N_{\text{TF}}+N_{\text{RE}}} \sum_j F_{ij} (\theta_l^{(1)})_{i,j} (\theta_b^{(1)})_{i,j} \tag{2.38}$$

where F is the Fisher information matrix measuring parameter importance.

At population level, LINGER determines the overall regulatory dynamics across the cell population by identifying cis-regulatory strengths, the average of absolute Shapley values and trans-regulatory strengths, the Pearson Correlation Coefficient (**PCC**) between the TF and RE embeddings.

For each cell type, the TF-RE regulatory potential is computed by summing TF binding affinities and cis-regulatory potential:

$$TRP_{ki} = \gamma_{ki} \sum_{j \in S_i} \text{TFB}_{kj} \text{CRP}_{ij} \tag{2.39}$$

$$TFB_{kj} = C_{kj} s_k (E_{\text{TF}})_k O_j (\alpha_{kj} + B_{kj}) \tag{2.40}$$

Where C_{kj} is the correlation, s_k the importance score, O_j the chromatin accessibility, and B_{kj} the binding affinity.

$$CRP_{ij} = \beta_{ij} O_j E_i e^{-\frac{d_{ij}}{d_0}} \tag{2.41}$$

The cell type-specific cis-regulatory (CRP_{ij}) and trans-regulatory potentials (TRP_{ki}) are calculated to assess the influence of REs and TFs on TGs based on their proximities and interactions.

Multimodal integration is essential for GRN inference as it provides a more comprehensive view of regulatory mechanisms by combining complementary information from different molecular layers. Methods such as SCENIC+ and GLUE explicitly address this integration challenge, with SCENIC+ incorporating chromatin accessibility alongside gene expression to construct enhancer-driven regulatory networks, and GLUE employing a graph-linked unified embedding approach to integrate unpaired multi-omics data. Similarly, the challenge of missing data in single-cell measurements has been addressed by approaches like GLUE, which can handle unpaired data through its latent space alignment strategy, implicitly performing cross-modality imputation. LINGER further demonstrates the power of integration by leveraging atlas-scale external data to enhance prediction accuracy. This synergy between integration and imputation creates a more robust foundation for GRN inference, allowing for the detection of subtle regulatory interactions that might be overlooked when analyzing incomplete or single-modality datasets

Despite these advances, significant challenges remain in GRN inference. Current methods rely heavily on TF binding motif databases and genomic distance cutoffs, which may not accurately capture the complex three-dimensional organization of the genome or cell type-specific binding preferences.

Chapter 3

Manuscript

3.1 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.

3.2 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 (Poulin et al., 2016; Stubbington et al., 2017). 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 (Singh et al., 2018; Unger Avila et al., 2024). 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 (Cao and Gao, 2022; Wang et al., 2021; Zhao et al., 2021; Zhou et al., 2023). 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 (Cha and Lee, 2020; Fiers et al., 2018).

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 (Delgado et al., 2023; Imrichová et al., 2015), 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 tran-

scription 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.

3.3 Material and Methods

3.3.1 scGraphETM model overview

scGraphETM integrates single-cell multimodal data across different experiments using interpretable latent embeddings and external GRN databases (**Fig.** 3.1). 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.** 3.1**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;

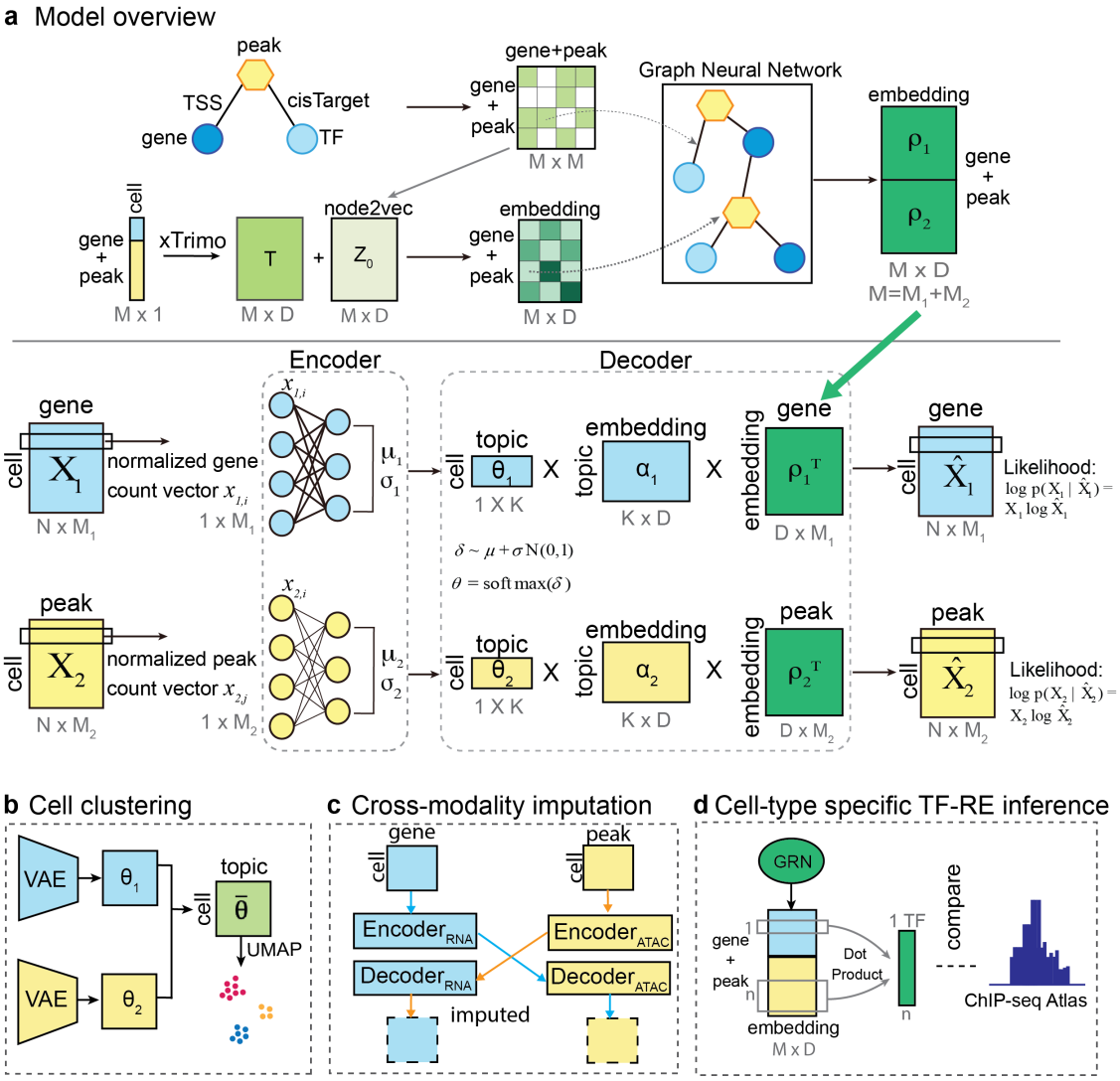

Figure 3.1: scGraphETM overview

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.

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.

3.3.2 Gene regulatory network construction

Multimodal GRN inference methods use an extended framework to that used by single-modality methods to reconstruct GRNs. For example, they may predict gene expression from TF gene expression, they assign TFs to accessible CREs using binding motif information (Yuan and Duren, 2024), and they associate CREs with target genes constrained by genomic distance. For the prediction of TF binding events, different methods use different, highly heterogeneous TF binding motif databases and prediction algorithms. Our approach to constructing gene regulatory networks involves representing transcription factor genes (TFs), non-transcription factor 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 the TSS of a gene are considered for network construction. While proximity-based approaches may have limitations in capturing long-range interactions, in future work, we plan to integrate chromosome conformation capture data such as Hi-C (Lieberman-Aiden et al., 2009; Rao et al., 2014) to more accurately map regulatory element-gene interactions based on their actual three-dimensional proximity rather than linear genomic distance (Fulco et al., 2019; Gasperini et al., 2020). 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 as these accessible regions show significantly higher read density than surrounding closed chromatin regions. We treat each ATAC-seq measured region as a RE. These entities 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. We focus on interactions between distinct genomic entities since self-regulation can be difficult to distinguish, and the inclusion of self-loops could diminish the impact of other regulatory relationships (Pratapa et al., 2020).

3.3.3 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 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:

$$\mathbf{a}_i = \text{leakyReLU}(x_i^{\text{input}} \times \mathbf{w}_1), \quad \mathbf{z}_i = \mathbf{a}_i \mathbf{w}_2 + \alpha \mathbf{a}_i \tag{3.1}$$

   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 $\mathbf{z}_i \in \mathbb{R}^b$.

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

$$\gamma_{i,j} = \frac{\exp(z_{i,j})}{\sum_{j=1}^{b} \exp(z_{i,j})} \tag{3.2}$$

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

$$\mathbf{t}_i = \boldsymbol{\gamma}_i \mathbf{T} \tag{3.3}$$

   where $\mathbf{t}_i \in \mathbb{R}^d$.

4. Zeros are encoded with $\mathbf{t}^0 \in \mathbb{R}^d$.

5. The final input embedding for gene i is:

$$\mathbf{V}_i^{\text{input}} = \mathbf{t}_i + \mathbf{z}_i^G \tag{3.4}$$

   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 t_i can be considered as the "position embedding" and the gene-specific and value-independent node2vec embeddings 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.

3.3.4 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:

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

Here, $h_v^{(k)}$ represents the feature vector of node v at layer k, $\mathcal{N}(v)$ denotes the set of neighbors of v, and $W^{(k)}$ 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.

3.3.5 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:

$$\boldsymbol{\mu}^*,\ \log \boldsymbol{\sigma}^* = \text{Encoder}(\mathbf{x}) \tag{3.5}$$

A reparameterization trick enables stochastic gradient descent through sampling:

$$\mathbf{z} = \boldsymbol{\mu}^* + \exp\left(\frac{\log \boldsymbol{\sigma}^*}{2}\right) \odot \boldsymbol{\epsilon} \tag{3.6}$$

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

$$\theta_k = \text{softmax}(\mathbf{z})_k = \frac{\exp(z_k)}{\sum_k \exp(z_k)} \tag{3.7}$$

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.

$$\hat{\mathbf{x}} = \boldsymbol{\theta}\boldsymbol{\beta} \tag{3.8}$$

$$\log p(\mathbf{x}|\hat{\mathbf{x}}) = \sum_g x_g \log \hat{x}_g \tag{3.9}$$

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):

$$\mathbb{E}_{q(z|x)}[\log p(\mathbf{x}|\mathbf{z})] - \mathrm{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})], \tag{3.10}$$

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 $\mathrm{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] = \mathbb{E}_{q(z)}[\log q(z|x)] - \mathbb{E}_{q(z)}[\log p(z)] = \sum_k -\frac{1}{2}\left[\log\{\sigma_k^*\}^2 - \{\mu_k^*\}^2 - \{\sigma_k^*\}^2 + 1\right]$ 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:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{NLL}} + \lambda_2 \mathcal{L}_{\text{KL}} + \lambda_3 \mathcal{L}_{\text{GR}}, \tag{3.11}$$

$$\mathcal{L}_{\text{NLL}} = -\sum_{i=1}^{m} \frac{1}{N_i M_i} \mathbf{x}_i \log(\hat{\mathbf{x}}_i), \tag{3.12}$$

$$\mathcal{L}_{\text{KL}} = \sum_k -\frac{1}{2}\left[\log(\sigma_k^*)^2 - (\mu_k^*)^2 - (\sigma_k^*)^2 + 1\right], \tag{3.13}$$

$$\mathcal{L}_{\text{GR}} = -\frac{1}{|E|} \sum_{(i,j)\in E} \left[A_{ij} \log \sigma(\hat{A}_{ij}) + (1 - A_{ij}) \log(1 - \sigma(\hat{A}_{ij})) \right]. \tag{3.14}$$

where $A \in \mathbb{R}^{M\times M}$ is the adjacency matrix of the graph G, $\mathbf{V} \in \mathbb{R}^{M\times D}$ is the node embedding, and $\hat{A} = \mathbf{V}\mathbf{V}^T \in \mathbb{R}^{M\times M}$, and $\sigma(\cdot)$ is the sigmoid function.

λ_1, λ_2, and λ_3 are dynamic weight values calculated during training based on the current epoch to balance the importance of different loss components over time.

3.3.6 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 (Li et al., 2023; Siletti 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 of 1e4 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.We selected top 3000 highly variable genes for the HVG scenarios.

3.3.7 Evaluation metrics

Using the cell topic distribution θ as input, the Louvain algorithm assigns cells to clusters (Blondel et al., 2008).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.

For the first two tasks, experiments were conducted on both highly variable genes and coding genes, while for gene regulatory network inference, highly variable genes were used.

3.4 Results

3.4.1 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.** 3.2, **Table** S3). 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.**??**c. 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

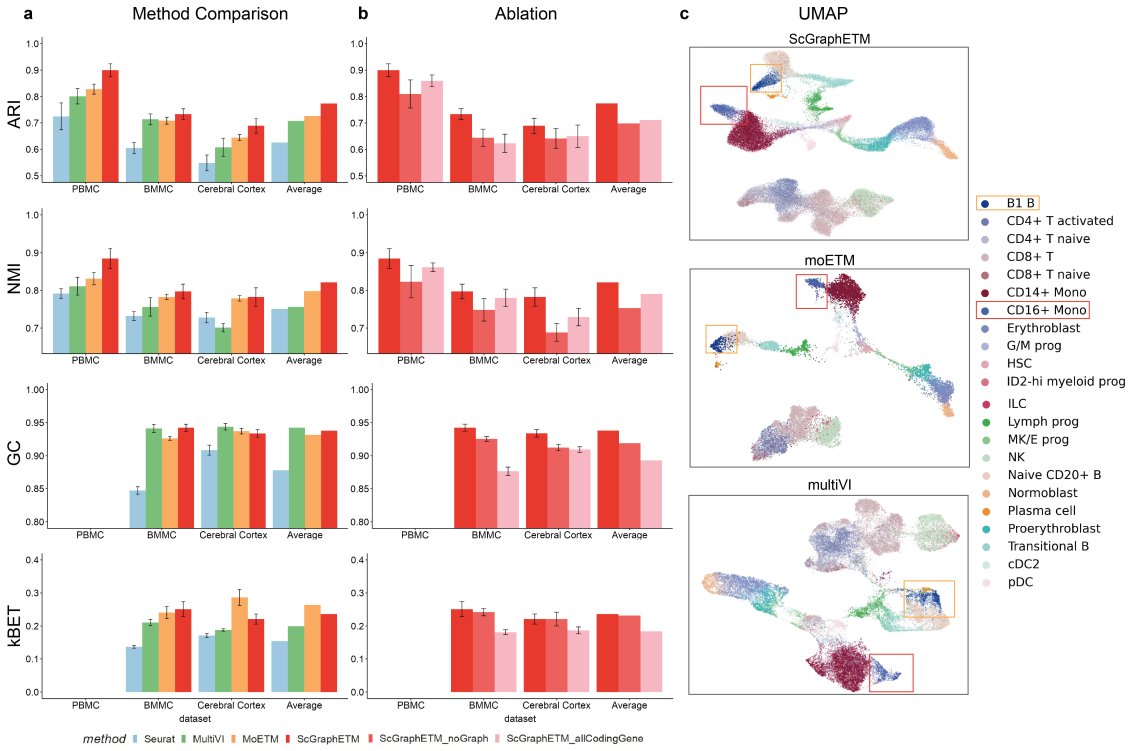

Figure 3.2: Methods comparison based on cell clustering.

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.

and 'Plasma celles', as highlighted by the orange box in **Fig. ??**c. These results indicate that scGraphETM improves cell clustering by incorporating the GRN component.

3.4.2 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-

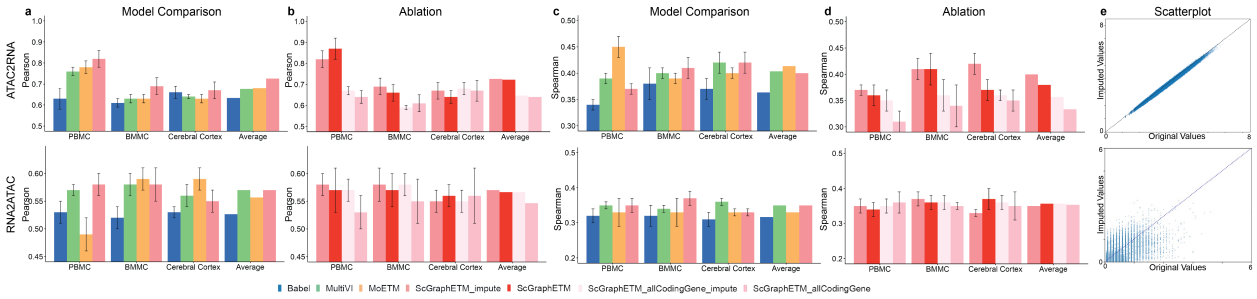

Figure 3.3: Methods comparison based on cross-modality imputation.
The upper panel displays performance on the ATAC2RNA imputation task, while the lower panel shows performance on the RNA2ATAC imputation task. **a, b** Pearson correlation for each method and scGraphETM's ablated versions on each dataset, 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.

topics matrix θ, 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 β.

For example, topic 28 and topic 88 were primarily associated with monocytes, while topic 93 was enriched for B cells (**Fig.** 3.4). 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).

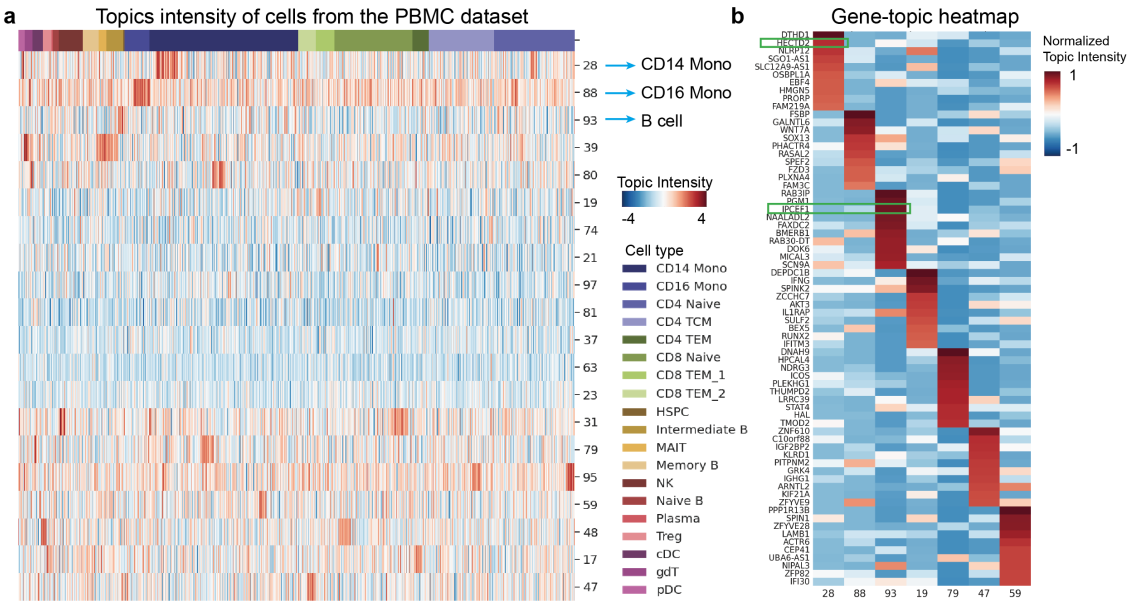

Figure 3.4: Topic analysis on the PBMC dataset

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.

3.4.3 scGraphETM enables accurate cross-modality imputation

The ETM approach represents both RNA and ATAC modalities in a shared topic space. During training, the joint reconstruction loss combines the reconstruction errors from both RNA and ATAC modalities, ensuring that the shared topic space captures information that is relevant for both data types. The linear decoders then reconstruct the input data to designated modality using these learned topic representations. 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.** 3.3, **Table** S2). 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.** 3.3c).

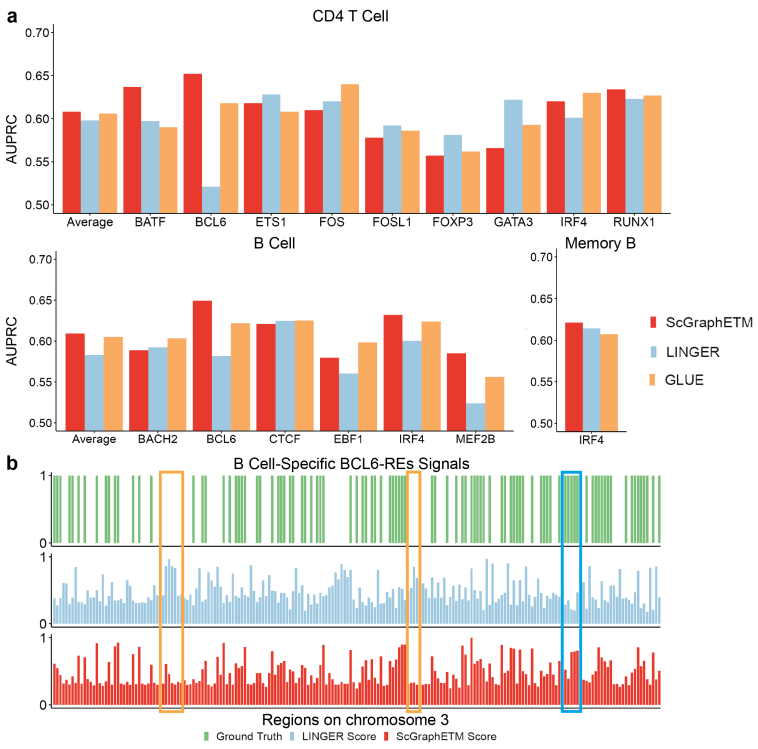

Figure 3.5: Methods comparison based on cell-type-specific GRN. a. Comparison across different cell types, TFs, and methods on the task of TF-RE binding potential inference. The x-axis represents transcription factors, and the y-axis represents AUPRC values. **b.** B cell-specific BCL6-RE predicted signals, compared with ChIP-Seq Atlas ground truth. The x-axis represents peaks on chr3, and y-axis shows the TF-RE binding potential score.

3.4.4 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. ChIP-seq Atlas

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 3.1: AUPRC Results for TF-RE prediction

(a) 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.

contains identified regions of significant enrichment for a specific TF binding in a specific tissue. The database provides a collection of ChIP-seq experiments that identify where particular transcription factors bind to DNA in 33368 specific cell types or tissues. 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.** 3.5, **Table** 3.1a). 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 0.6213, while LINGER averaged 0.6144 and GLUE achieved 0.6073. These results underscore scGraphETM's ability to consistently outperform other meth-

ods 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. 3.5b, scGraphETM produced 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.

3.5 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).

3.6 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/`.

3.7 Code Availability

All original code has been deposited at `https://github.com/li-lab-mcgill/scGraphETM`.

3.8 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.

Bibliography

Ashuach, T., Gabitto, M. I., Koodli, R. V., Saldi, G.-A., Jordan, M. I., and Yosef, N. (2023). Multivi: deep generative model for the integration of multimodal data. *Nature Methods*, 20(8):1222–1231.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods*, 20(9):1355–1367.

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

Cao, Z. and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*, 40:1458–1466.

Cha, J. and Lee, I. (2020). Single-cell network biology for resolving cellular heterogeneity in human diseases. *Experimental and Molecular Medicine*, 52(9):1428–1437.

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

Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005). Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*, 2005(09):P09008.

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

Dieng, A. B., Ruiz, F. J., and Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184–1191.

Fiers, M. W., Minnoye, L., Aibar, S., Bravo González-Blas, C., Kalender Atak, Z., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Briefings in functional genomics*, 17(4):246–254.

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of crispr perturbations. *Nature Genetics*, 51(12):1664–1669.

Gasperini, M., Tome, J. M., and Shendure, J. (2020). Genome-wide identification of regulatory elements by massively parallel reporter assays. *Nature Reviews Genetics*, 21(10):630–643.

Gong, J., Hao, M., Zeng, X., Liu, C., Ma, J., Cheng, X., Wang, T., Zhang, X., and Song, L. (2023). xtrimogene: An efficient and scalable representation learner for single-cell rna-seq data. biorxiv.

Gong, L., Huang, J., Bai, X., Song, L., Hang, J., and Guo, J. (2024). Expression of hectd2 predicts peritoneal metastasis of gastric cancer and reconstructs immune microenvironment. *Cancer Cell International*, 24(1):380.

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

Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., and Song, L. (2024). Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, pages 1–11.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2:193–218.

Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D., and Aerts, S. (2015). i-cistarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic acids research*, 43(W1):W57–W64.

Li, Y. E., Preissl, S., Miller, M., Johnson, N. D., Wang, Z., Jiao, H., Zhu, C., Wang, Z., Xie, Y., Poirion, O., et al. (2023). A comparative atlas of single-cell chromatin accessibility in the human brain. *Science*, 382(6667):eadf7044.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. *Nature neuroscience*, 19(9):1131–1141.

Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2):147–154.

Raghavan, V., Li, Y., and Ding, J. (2023). Harnessing agent-based modeling in cellagentchat to unravel cell-cell interactions from single-cell data. *bioRxiv*, pages 2023–08.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al. (2014). A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.

Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K. W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046.

Singh, A. J., Ramsey, S. A., Filtz, T. M., and Kioussi, C. (2018). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, 75:1013–1025.

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

Stubbington, M. J., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S. A. (2017). Single-cell transcriptomics to explore the immune system in health and disease. *Science*, 358(6359):58–63.

Tsoucas, D., Dong, R., Chen, H., Zhu, Q., Guo, G., and Yuan, G.-C. (2019). Accurate estimation of cell-type composition from gene expression data. *Nature communications*, 10(1):2975.

Unger Avila, P., Padvitski, T., Leote, A. C., Chen, H., Saez-Rodriguez, J., Kann, M., and Beyer, A. (2024). Gene regulatory networks in disease and ageing. *Nature Reviews Nephrology*, pages 1–18.

Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q., and Xu, D. (2021). scgnn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12(1):1–11.

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

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24.

Yin, D., Wang, K., Zhao, J., Yao, J., Han, X., Yan, B., Dong, J., and Liao, L. (2024). Ipcef1: Expression patterns, clinical correlates and new target of papillary thyroid carcinoma. *Journal of Cancer*, 15(19):6434.

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

Zhao, Y., Cai, H., Zhang, Z., Tang, J., and Li, Y. (2021). Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data. *Nature communications*, 12(1):5261.

Zhou, M., Zhang, H., Bai, Z., Mann-Krzisnik, D., Wang, F., and Li, Y. (2023). Single-cell multi-omics topic embedding reveals cell-type-specific and covid-19 severity-related immune signatures. *Cell Reports Methods*, 3:100563.

Zou, Z., Ohta, T., and Oki, S. (2024). Chip-atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Research*, page gkae358.

3.9 Appendix

The hyperparameters presented in Tables S4 and S5 were selected through grid search optimization. We performed a systematic grid search over embedding dimensions (128, 256, 512), learning rates (1e-4, 5e-4, 1e-3, 5e-3), and batch sizes (32, 64, 128, 256), similarly for the Node2Vec hyperparameters. The number of topics was determined by balancing computational efficiency with the biological requirement for capturing sufficient cellular heterogeneity,

Dataset	Method	ATAC2RNA		RNA2ATAC	
		Pearson Corr	Spearman Corr	Pearson Corr	Spearman Corr
PBMC	ScGraphETM	**0.87 (0.05)**	0.36 (0.02)	0.57 (0.04)	0.34 (0.02)
	ScGraphETM_allCodingGene	0.64 (0.03)	0.31 (0.02)	0.53 (0.03)	**0.36 (0.03)**
	ScGraphETM_impute	0.82 (0.04)	0.37 (0.01)	**0.58 (0.02)**	0.35 (0.02)
	ScGraphETM_allCodingGene_impute	0.67 (0.02)	0.35 (0.02)	0.57 (0.02)	0.35 (0.02)
	Babel	0.63 (0.05)	0.34 (0.01)	0.53 (0.02)	0.32 (0.02)
	MultiVI	0.76 (0.02)	0.39 (0.01)	0.57 (0.01)	0.35 (0.01)
	MoETM	0.78 (0.03)	**0.45 (0.02)**	0.49 (0.03)	0.33 (0.04)
BMMC	ScGraphETM	0.66 (0.04)	**0.41 (0.03)**	0.57 (0.03)	0.36 (0.02)
	ScGraphETM_allCodingGene	0.61 (0.04)	0.34 (0.04)	0.55 (0.04)	0.35 (0.01)
	ScGraphETM_impute	**0.69 (0.04)**	0.41 (0.02)	0.58 (0.03)	**0.37 (0.02)**
	ScGraphETM_allCodingGene_impute	0.59 (0.01)	0.36 (0.03)	0.58 (0.02)	0.36 (0.02)
	Babel	0.61 (0.02)	0.38 (0.03)	0.52 (0.02)	0.32 (0.03)
	MultiVI	0.63 (0.02)	0.4 (0.01)	0.58 (0.02)	0.34 (0.01)
	MoETM	0.63 (0.02)	0.39 (0.01)	**0.59 (0.02)**	0.33 (0.04)
Cerebral Cortex	ScGraphETM	0.64 (0.03)	0.37 (0.02)	0.56 (0.02)	**0.37 (0.03)**
	ScGraphETM_allCodingGene	0.67 (0.05)	0.35 (0.02)	0.56 (0.05)	0.35 (0.04)
	ScGraphETM_impute	0.67 (0.04)	**0.42 (0.02)**	0.55 (0.02)	0.33 (0.01)
	ScGraphETM_allCodingGene_impute	**0.68 (0.03)**	0.36 (0.01)	0.55 (0.02)	0.36 (0.02)
	Babel	0.66 (0.03)	0.37 (0.02)	0.53 (0.01)	0.31 (0.02)
	MultiVI	0.64 (0.01)	0.42 (0.02)	0.56 (0.02)	0.36 (0.01)
	MoETM	0.63 (0.02)	0.4 (0.01)	**0.59 (0.02)**	0.33 (0.01)

Table S2: Cross-modality imputation evaluation. We imputed gene expression values from chromatin accessibility values (ATAC2RNA) and vice versa (RNA2ATAC). Under each dataset, the best score per evaluation metric in each direction is in bold, and the second best score is in blue. The values represent the mean (standard deviation).

Dataset	Method	ARI	MNI	kBET	GC
PBMC	ScGraphETM	**0.900 (0.024)**	**0.885 (0.026)**	-	-
	ScGraphETM_allCodingGene	0.860 (0.022)	0.861 (0.012)	-	-
	ScGraphETM_noGraph	0.810 (0.053)	0.823 (0.043)	-	-
	MultiVI	0.801 (0.029)	0.811 (0.024)	-	-
	Seurat	0.725 (0.051)	0.791 (0.013)	-	-
	MoETM	0.828 (0.019)	0.831 (0.016)	-	-
BMMC	ScGraphETM	**0.733 (0.021)**	**0.797 (0.020)**	**0.251 (0.023)**	**0.942 (0.005)**
	ScGraphETM_allCodingGene	0.623 (0.035)	0.780 (0.023)	0.181 (0.007)	0.877 (0.006)
	ScGraphETM_noGraph	0.644 (0.033)	0.748 (0.030)	0.242 (0.011)	0.925 (0.004)
	MultiVI	0.714 (0.021)	0.756 (0.024)	0.211 (0.009)	0.941 (0.006)
	Seurat	0.605 (0.021)	0.732 (0.011)	0.136 (0.004)	0.847 (0.006)
	MoETM	0.709 (0.013)	0.783 (0.007)	0.241 (0.018)	0.926 (0.003)
Cerebral Cortex	ScGraphETM	**0.689 (0.029)**	**0.783 (0.025)**	0.221 (0.015)	0.934 (0.006)
	ScGraphETM_allCodingGene	0.650 (0.042)	0.729 (0.023)	0.187 (0.011)	0.909 (0.004)
	ScGraphETM_noGraph	0.642 (0.038)	0.689 (0.023)	0.221 (0.021)	0.912 (0.005)
	MultiVI	0.608 (0.034)	0.701 (0.011)	0.188 (0.004)	0.944 (0.005)
	Seurat	0.549 (0.030)	0.728 (0.014)	0.171 (0.006)	0.909 (0.008)
	MoETM	0.644 (0.011)	0.779 (0.007)	**0.287 (0.025)**	**0.937 (0.004)**

Table S3: Evaluation of cell clustering. Under each dataset, the best score per evaluation metric in each direction is in bold, and the second best score is in blue. The values represent the mean (standard deviation).

Hyperparameter	Values	Selected Value
Learning Rate	{1e-3, 1e-4, 1e-5 }	1e-4
Embedding Size	{128, 256, 512}	512
Topic Number	{20, 40, 60, 80, 100}	100
TSS Threshold	{150e3, 250e3, 1e6}	1e6
HVG	{2000, 3000, 4000, 5000}	3000
Latent Size	{10, 50, 100}	100

Table S4: scGraphETM Hyperparameters and Training Settings

Hyperparameter	Values	Selected Value
Learning Rate	{1e-2, 1e-3, 1e-4 }	1e-2
Embedding Size	{128, 256, 512}	512
p	{0.25, 0.5, 0.75, 1.0}	0.5
q	{1.0, 2.0, 3.0, 4.0}	2.0
Walk Length	{10, 20, 40, 80}	20
Number of Walks	{10, 20, 30, 50}	30
Window Size	{5, 10, 15, 20}	10
Iterations	{5, 10, 20, 50}	10

Table S5: Node2Vec Hyperparameters and Training Settings

Chapter 4

Discussion

The aim of this thesis is to explore the application of graph neural networks and Embedded Topic Models (ETMs) in the construction and analysis of cell-specific gene regulatory networks from single-cell multi-omic data. The necessity of the unified and interpretable deep learning framework scGraphETM arises from the complex and multifaceted nature of gene regulation, which cannot be fully captured through a single-omic approach. scGraphETM elucidates cellular regulation at the single cell level. Our technical contributions are threefold. First, scGraphETM employs a scalable graph encoder to effectively connect existing biological data and disparate omic data sources into a unified framework. This encoder projects the diverse data onto a common latent topic mixture representation, enabling a cohesive analysis of the intertwined molecular pathways that govern cellular behavior. Second, the model utilizes a linear decoder pivotal for extracting multi-omic signatures from the integrated data. These signatures represent the most influential features within each latent topic, revealing critical marker genes and phenotypic markers. Third, scGraphETM has enhanced capability of inferring cell-specific gene regulatory networks (GRNs) that contextualizes transcription factor binding events within specific cellular states. By incorporating prior knowledge from cisTarget databases and leveraging chromatin accessibility data alongside gene expression profiles, our model captures the dynamic regulatory interactions that vary across different cell types. Com-

prehensive evaluations on three datasets demonstrate that scGraphETM consistently outperforms state-of-the-art methods in cell clustering, cross-modality imputation, and cell type-specific TF-RE relationship identification.

4.1 Clustering and Cell Type Annotation

The evaluation of scGraphETM's performance in cell type clustering demonstrates its superior ability to identify biologically meaningful cell groups across multiple datasets. As shown in Table S3, scGraphETM consistently achieves the highest Adjusted Rand Index (ARI) scores among all tested methods, with values of 0.900, 0.733, and 0.689 for the PBMC, BMMC, and Cerebral Cortex datasets, respectively. These results represent a substantial improvement over existing methods such as multiVI, Seurat, and moETM. The high Normalized Mutual Information (NMI) scores further validate scGraphETM's clustering accuracy, indicating strong concordance with known cell type annotations.

When compared to Seurat's weighted nearest neighbor approach, which relies primarily on pairwise similarities between cells, scGraphETM benefits from modeling the shared latent topic space that captures co-expression patterns across modalities. Unlike multiVI's variational autoencoder architecture that treats the integration task independently from biological prior knowledge, scGraphETM incorporates known regulatory relationships through its graph neural network component, leading to more biologically informed embeddings. The performance gap between scGraphETM and scGraphETM_noGraph (Table S3) further confirms that the integration of regulatory knowledge significantly enhances clustering accuracy.

The UMAP visualization in Figure 3.2c provides qualitative evidence of scGraphETM's ability to generate well-separated clusters that correspond to distinct cell types. Notably, for the BMMC dataset, scGraphETM effectively distinguishes between closely related cell types such as CD16+ Monocytes and B1 B cells, which other methods struggled to separate. While moETM similarly employs a topic modeling approach, it lacks the graph-

based feature refinement that allows scGraphETM to better capture the regulatory relationships distinguishing closely related cell types.

The batch effect removal metrics (GC and kBET) further highlight scGraphETM's ability to integrate data from different experimental batches while preserving biological variation. With an average GC score of 0.938 across multiple datasets, scGraphETM effectively ensures that cells of the same type from different batches remain proximal in the embedding space, facilitating accurate cell type annotation in heterogeneous samples.

4.2 Cross Modality Imputation

The cross-modality imputation results demonstrate scGraphETM's remarkable ability to predict unmeasured modalities from measured ones, a critical capability for comprehensive multi-omic analysis. As detailed in Table S2, the scGraphETM_impute variant achieves the highest performance in predicting gene expression from chromatin accessibility (ATAC2RNA), with an average Pearson correlation of 0.73 across all datasets. This represents a significant improvement over specialized imputation tools such as BABEL (Wu et al., 2021), highlighting the advantage of this integrated approach.

Unlike BABEL, which employs separate encoder-decoder paths for each modality transition, scGraphETM benefits from a unified latent space that simultaneously captures relationships between all modalities. The incorporation of the graph neural network layer provides scGraphETM with explicit modeling of TF-RE relationships, which are crucial for accurately translating between chromatin accessibility and gene expression. This is particularly evident in the performance improvement over multiVI (**?**), which relies on a more general variational framework without specific regulatory modeling.

The imputation task from RNA to ATAC (RNA2ATAC) presents a greater challenge due to predicting from a lower-dimensional to a higher-dimensional space. Nevertheless, scGraphETM_impute achieves competitive performance with an average Pearson correlation of 0.57, outperforming other methods. The scatterplot in Figure 3.3 provides visual

confirmation of the strong correlation between original and imputed values, demonstrating minimal systematic bias in the predictions.

The ablation study comparing scGraphETM with and without imputation loss demonstrates the critical role of this component for enhancing prediction accuracy. Unlike moETM, which relies solely on shared topic spaces for imputation, scGraphETM's dedicated imputation loss function optimizes specifically for this task. The comparison with scGraphETM_allCodingGene variants reveals that focusing on highly variable genes significantly improves performance for high-to-low dimensional imputation tasks (ATAC2RNA), while showing comparable results for the more challenging RNA2ATAC task (Table S2). This finding suggests that non-variable genes contribute limited useful information for certain imputation directions, guiding future model development and application.

The comparison with scGraphETM_allCodingGene variants reveals that focusing on highly variable genes significantly improves performance for high-to-low dimensional imputation tasks (ATAC2RNA), while showing comparable results for the more challenging RNA2ATAC task (Table S2). Highly variable genes, by definition, contain more information about cellular states and differentiation processes than stably expressed housekeeping genes. When predicting gene expression from chromatin accessibility (ATAC2RNA), these highly variable genes serve as strong discriminative features that effectively capture cell type-specific regulatory patterns. The model benefits from focusing computational resources on these information-rich features rather than diluting its learning capacity across thousands of relatively less informative genes. Additionally, the dimensionality reduction inherent in the high-to-low dimensional task (ATAC2RNA) amplifies the importance of feature selection. With fewer output dimensions to predict, the model can leverage the concentrated signal in highly variable genes to achieve more accurate predictions.

4.3 GRN Inference

Table 3.1a summarizes the AUPRC scores for predicting TF-RE relationships across different cell types in the PBMC dataset. scGraphETM achieves an average AUPRC of 0.613 across all tested transcription factors and cell types, outperforming both LINGER (0.599) and GLUE (0.606). This performance is particularly noteworthy for B cells, where scGraphETM reaches an AUPRC of 0.609 compared to LINGER's 0.581 and GLUE's 0.605.

The superior performance of scGraphETM in cellular GRN inference can be attributed to several key innovations compared to existing approaches. Unlike LINGER, which requires bulk RNA-seq and ATAC-seq data for pretraining and employs TF-specific models, scGraphETM learns all TF-RE relationships simultaneously within a unified framework, enabling more efficient detection of shared regulatory patterns. While GLUE similarly employs a graph-based approach, its standard graph attention mechanism faces computational limitations with large-scale networks, whereas scGraphETM's GraphSAGE-inspired neighborhood aggregation strategy provides better scalability for the regulatory networks typically with millions of interactions in genomic data.

Figure 3.5b provides a detailed visualization of scGraphETM's predicted binding potential for BCL6 in B cells, compared against ChIP-seq ground truth from the ChIP-seq Atlas database. The figure clearly illustrates that scGraphETM achieves higher prediction scores in regions truly bound by BCL6 (blue box) while maintaining higher specificity in true negative regions (orange boxes) compared to LINGER. This fine-grained discriminative ability stems from scGraphETM's cell-specific dynamic node features, which integrate both global network properties via node2vec embedding and local expression patterns via xTrimoGene encoding (Gong et al., 2023).

Recent approaches such as SCENIC+ rely on separate steps for motif enrichment and co-expression analysis, potentially missing complex regulatory interactions that span these different data types. In contrast, scGraphETM's end-to-end training approach allows the model to leverage information flow between the graph neural network and the embed-

ded topic model components, leading to more coherent and accurate regulatory network predictions. The cell-specific nature of the inferred network provides insights into the regulatory dynamics that drive cellular identity. This cell-type resolution is particularly valuable for understanding complex tissues where different cell types employ unique regulatory programs despite sharing the same genome. By delineating these cell-specific regulatory landscapes, scGraphETM enables the identification of master regulators that maintain cell identity and reveals differential regulatory mechanisms. In disease contexts, these cell-specific GRNs can pinpoint which regulatory circuits are dysregulated and in which cell populations, allowing for a more nuanced understanding of pathogenesis

4.4 Model Limitations

A significant biological limitation of the current scGraphETM framework is its reliance on linear genomic distances for associating regulatory elements with potential target genes. The model primarily considers proximity-based connections, where regulatory elements are linked to genes based on their distance from transcription start sites, typically using a threshold of 1 Mbp as implemented in this study. This approach may fail to capture the complex three-dimensional organization of chromatin that enables distal regulatory elements.

The second limitation concerns the computational scalability of scGraphETM's graph neural network component when applied to datasets with extensive features or large numbers of potential regulatory interactions. While the model employs a GraphSAGE-inspired neighborhood aggregation strategy in contrast to the more computationally intensive graph attention mechanisms, memory constraints still become significant when modeling comprehensive genomic datasets.

In practical applications involving full transcriptomes (20,000+ genes) and genome-wide chromatin accessibility regions (100,000+ peaks), the resulting graph can contain millions of nodes and edges. The computational burden primarily stems from the mem-

ory required to store the full adjacency matrix and node features and the computational cost of message passing across multiple layers of the graph neural network. Leveraging distributed computing frameworks enabling parallel processing of large-scale datasets across multiple compute nodes would extend scGraphETM's applicability to atlas-scale datasets containing millions of cells.

Chapter 5

Conclusion and Future Work

This thesis has presented scGraphETM, a novel graph-based deep learning approach for unraveling cell specific gene regulatory networks from single-cell multi-omics data. The systematic comparison with state-of-the-art methods across multiple datasets confirms scGraphETM's superior performance in identifying cell types, predicting unmeasured modalities, and reconstructing regulatory networks. Particularly, the ablation studies further validate the importance of the graph neural network component and the selection of highly variable genes for optimal performance.

Future works include incorporating three-dimensional chromatin interaction data from Hi-C Belton et al. (2012) or similar technologies to enable more accurate modeling of distal regulatory interactions that play crucial roles in gene regulation. Additional approaches for improving the computational scalability of the graph encoder to facilitate analysis of increasingly large single-cell atlases such as more efficient graph sampling techniques or implementing distributed computing frameworks would be beneficial. As new single-cell technologies for measuring additional molecular modalities become more accessible, extending scGraphETM to incorporate protein expression, DNA methylation, or other epigenetic modifications would provide an even more comprehensive view of cellular regulation. The modular design of scGraphETM's architecture makes it well-suited for such extensions.

Bibliography

10x Genomics (2019). Pbmcs from c57bl/6 mice (v1, 150x150), single cell immune profiling dataset. Cell Ranger v3.1.0, 10x Genomics. Accessed: 2019, July 24.

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., and Aerts, S. (2017). Scenic: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11):1083–1086.

Argelaguet, R., Arnol, D., Bredikhin, D., et al. (2020a). Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21:111.

Argelaguet, R., Cuomo, A. S., Stegle, O., and Marioni, J. C. (2020b). Computational principles and challenges in single-cell data integration. *Nature Biotechnology*, 39(2):133–145.

Ashuach, T., Gabitto, M. I., Koodli, R. V., Saldi, G.-A., Jordan, M. I., and Yosef, N. (2023). Multivi: deep generative model for the integration of multimodal data. *Nature Methods*, 20(8):1222–1231.

Badia-i Mompel, P., Wessels, L., Müller-Dott, S., et al. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nature Reviews Genetics*, 24:739–754.

Belton, J., McCord, R., Gibcus, J., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-c: a comprehensive technique to capture the conformation of genomes. *Methods*, 58(3):268–276. Epub 2012 May 29. PMID: 22652625; PMCID: PMC3874846.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods*, 20(9):1355–1367.

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., and Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490.

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

Cao, Z. and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*, 40:1458–1466.

Cha, J. and Lee, I. (2020). Single-cell network biology for resolving cellular heterogeneity in human diseases. *Experimental and Molecular Medicine*, 52(9):1428–1437.

Chaudhary, K., Poirion, O. B., Lu, L., and Garmire, L. X. (2022). Deep learning–based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research*, 28(2):345–357.

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

Chen, Y., Li, Y., Narayan, R., Subramanian, A., and Xie, X. (2021). Gene regulatory network inference using graph neural networks. *Nature Communications*, 12(1):1–12.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., et al. (2016). A survey of best practices for rna-seq data analysis. *Genome Biology*, 17(1):1–19.

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

Consortium, H. (2019). The human body at cellular resolution: the nih human biomolecular atlas program. *Nature*, 574(7777):187–192.

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, 348(6237):910–914.

Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005). Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*, 2005(09):P09008.

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

Dieng, A. B., Ruiz, F. J., and Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184–1191.

Dutil, F., Cohen, J. P., Weiss, M., Derevyanko, G., and Bengio, Y. (2018). Towards gene expression convolutions using gene interaction graphs. *arXiv preprint arXiv:1806.06975*.

Fiers, M. W., Minnoye, L., Aibar, S., Bravo González-Blas, C., Kalender Atak, Z., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Briefings in functional genomics*, 17(4):246–254.

Franzén, O. and Björkegren, J. L. (2019). scatlas: a single-cell gene expression data portal for the analysis of embryonic and tissue development. *Database*, 2019.

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., et al. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of crispr perturbations. *Nature Genetics*, 51(12):1664–1669.

Gainza, P., Sverrisson, F., Monti, F., Rodola, E., Bronstein, M. M., and Correia, B. E. (2022). Protein surface representation via geometric deep learning. *Proceedings of the National Academy of Sciences*, 119(12):e2119260119.

Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Research*, 29(8):1363–1375.

Gasperini, M., Tome, J. M., and Shendure, J. (2020). Genome-wide identification of regulatory elements by massively parallel reporter assays. *Nature Reviews Genetics*, 21(10):630–643.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings.

Gong, J., Hao, M., Zeng, X., Liu, C., Ma, J., Cheng, X., Wang, T., Zhang, X., and Song, L. (2023). xtrimogene: An efficient and scalable representation learner for single-cell rna-seq data. biorxiv.

Gong, L., Huang, J., Bai, X., Song, L., Hang, J., and Guo, J. (2024). Expression of hectd2 predicts peritoneal metastasis of gastric cancer and reconstructs immune microenvironment. *Cancer Cell International*, 24(1):380.

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

Haghverdi, L., Lun, A. T., Morgan, M. D., and Marioni, J. C. (2018). Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5):421–427.

Hamilton, W., Ying, Z., and Leskovec, J. (2017a). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017b). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1024–1034.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017c). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034.

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C. Y., Lee, M., Kim, E., et al. (2018). Trrust v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, 46(D1):D380–D386.

Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., and Song, L. (2024). Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, pages 1–11.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., et al. (2021a). Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath,

J. M., Blish, C. A., Gottardo, R., Smibert, P., and Satija, R. (2021b). Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29.

Herrmann, C., de Sande, B. V., Potier, D., and Aerts, S. (2012). i-cistarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Research*, 40(15):e114. Epub 2012 Jun 20.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., and Leskovec, J. (2020a). Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems*, volume 33, pages 22118–22133. Provides benchmarks for scalability evaluation.

Hu, Z., Dong, Y., Wang, K., and Sun, Y. (2020b). Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, pages 2704–2710. Proposes HGT for heterogeneous graphs.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2:193–218.

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5(9):e12776.

Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D., and Aerts, S. (2015). i-cistarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic acids research*, 43(W1):W57–W64.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589.

Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., and Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201.

Korsunsky, I., Millard, N., Fan, J., et al. (2019). Fast, sensitive and accurate integration of single-cell data with harmony. *Nature Methods*, 16:1289–1296.

Langfelder, P. and Horvath, S. (2008). Wgcna: an r package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1):559.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., and Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739.

Li, C., Hong, Y., Li, B., et al. (2025). Benchmarking single-cell cross-omics imputation methods for surface protein expression. *Genome Biology*, 26:46.

Li, Y. E., Preissl, S., Miller, M., Johnson, N. D., Wang, Z., Jiao, H., Zhu, C., Wang, Z., Xie, Y., Poirion, O., et al. (2023). A comparative atlas of single-cell chromatin accessibility in the human brain. *Science*, 382(6667):eadf7044.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293.

Liu, L., Liu, C., Quintero, A., Wu, L., Yuan, Y., Wang, M., Cheng, M., Leng, L., Xu, L., Dong, G., et al. (2019). Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nature Communications*, 10(1):470.

Lopez, R., Regier, J., Cole, M., et al. (2018a). Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058.

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

Luecken, M., Büttner, M., Chaichoompu, K., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, 19:41–50.

Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular Systems Biology*, 15(6):e8746.

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.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Megill, C., Martin, B., Weaver, C., Bell, S., Prins, L., Badajoz, S., McCandless, B., Pisco, A. O., Kinsella, M., Griffin, F., et al. (2021). cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. *bioRxiv*, pages 2021–04.

Moerman, T., Santos, S. A., González-Blas, C. B., Simm, J., Moreau, Y., Aerts, J., and Aerts, S. (2019). Grnboost2 and arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35:2159–2161.

Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., and Welling, M. (2008). Fast collapsed gibbs sampling for latent dirichlet allocation. pages 569–577.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. *Nature neuroscience*, 19(9):1131–1141.

Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2):147–154.

Raghavan, V., Li, Y., and Ding, J. (2023). Harnessing agent-based modeling in cellagentchat to unravel cell-cell interactions from single-cell data. *bioRxiv*, pages 2023–08.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al. (2014). A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. *eLife*, 6:e27041.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80.

Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K. W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046.

Singh, A. J., Ramsey, S. A., Filtz, T. M., and Kioussi, C. (2018). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, 75:1013–1025.

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

Stark, R., Grzelak, M., and Hadfield, J. (2019). Rna sequencing: the teenage years. *Nature Reviews Genetics*, 20(11):631–656.

Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3):133–145.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. r., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.e21. Epub 2019 Jun 6. PMID: 31178118; PMCID: PMC6687398.

Stubbington, M. J., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S. A. (2017). Single-cell transcriptomics to explore the immune system in health and disease. *Science*, 358(6359):58–63.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., et al. (2009). mrna-seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382.

Tsoucas, D., Dong, R., Chen, H., Zhu, Q., Guo, G., and Yuan, G.-C. (2019). Accurate estimation of cell-type composition from gene expression data. *Nature communications*, 10(1):2975.

Unger Avila, P., Padvitski, T., Leote, A. C., Chen, H., Saez-Rodriguez, J., Kann, M., and Beyer, A. (2024). Gene regulatory networks in disease and ageing. *Nature Reviews Nephrology*, pages 1–18.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations (ICLR)*.

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., and Klein, A. M. (2020). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 360(6392):981–987.

Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q., and Xu, D. (2021a). scgnn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12(1):1–11.

Wang, X., Qian, B., Ye, J., and Davidson, I. (2021b). Using graph neural networks to predict cancer survival rates. *Nature Communications*, 12(1):1–13.

Wang, Z., Gerstein, M., and Snyder, M. (2009). Rna-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

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

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24.

Xiong, L., Xu, K., Tian, K., et al. (2019). Scale method for single-cell atac-seq analysis via latent feature extraction. *Nature Communications*, 10:4576.

Yin, D., Wang, K., Zhao, J., Yao, J., Han, X., Yan, B., Dong, J., and Liao, L. (2024). Ipcef1: Expression patterns, clinical correlates and new target of papillary thyroid carcinoma. *Journal of Cancer*, 15(19):6434.

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

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., La Manno, G., et al. (2022). Graph neural networks for single-cell analysis. *Nature Methods*, 19(2):178–185.

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019). Cellmarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research*, 47(D1):D721–D728.

Zhao, Y., Cai, H., Zhang, Z., Tang, J., and Li, Y. (2021). Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data. *Nature communications*, 12(1):5261.

Zhou, M., Zhang, H., Bai, Z., Mann-Krzisnik, D., Wang, F., and Li, Y. (2023). Single-cell multi-omics topic embedding reveals cell-type-specific and covid-19 severity-related immune signatures. *Cell Reports Methods*, 3:100563.

Zhu, C., Yu, M., Huang, H., Juric, I., Abnousi, A., Hu, R., Lucero, J., Behrens, M. M., Hu, M., and Ren, B. (2020). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral t cell exhaustion. *Nature Biotechnology*, 38(8):970–978.

Zou, Z., Ohta, T., and Oki, S. (2024). Chip-atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Research*, page gkae358.