GRNFormer: A Biologically-Guided Framework for Integrating Gene Regulatory Networks into RNA Foundation Models

Mufan Qiu[1], Xinyu Hu[2], Fengwei Zhan[2,3], Sukwon Yun[1], Jie Peng[1], Ruichen Zhang[1], Bhavya Kailkhura[4], Jiekun Yang[2], Tianlong Chen[1]

[1]University of North Carolina at Chapel Hill, [2]Rutgers University,
[3]Barnard College, [4]Lawrence Livermore National Laboratory

Correspondence: tianlong@cs.unc.edu

Abstract

Foundation models for single-cell RNA sequencing (scRNA-seq) have shown promising capabilities in capturing gene expression patterns. However, current approaches face critical limitations: they *ignore biological prior knowledge* encoded in gene regulatory relationships and *fail to leverage multi-omics signals* that could provide complementary regulatory insights. In this paper, we propose GRNFormer, a new framework that systematically integrates multi-scale *Gene Regulatory Networks (GRNs)* inferred from multi-omics data into RNA foundation model training. Our framework introduces two key innovations. First, we introduce a pipeline for constructing *hierarchical GRNs* that capture regulatory relationships at both *cell-type-specific* and *cell-specific* resolutions. Second, we design a *structure-aware integration framework* that addresses the *information asymmetry* in GRNs through two technical advances: ❶ A graph topological adapter using multi-head cross-attention to weight regulatory relationships dynamically, and ❷ a novel *edge perturbation strategy* that perturb GRNs with biologically-informed co-expression links to augment graph neural network training. Comprehensive experiments have been conducted on three representative downstream tasks across multiple model architectures to demonstrate the effectiveness of GRNFormer. It achieves consistent improvements over state-of-the-art (SOTA) baselines: **3.6**% increase in drug response prediction correlation, **9.6**% improvement in single-cell drug classification AUC, and **1.1**% average gain in gene perturbation prediction accuracy.

1 Introduction

Recent advances in foundation models (FMs) for single-cell RNA sequencing (scRNA-seq) analysis has revolutionized our ability to decipher cellular states and gene expression patterns. Models like scGPT (Cui et al., 2024), Geneformer (Theodoris et al., 2023), and scFoundation (Hao et al., 2024) demonstrate remarkable capabilities in capturing transcriptomic relationships through large-scale pretraining on millions of cells. These models achieve state-of-the-art performance in critical tasks, including cell type annotation, perturbation prediction, and multi-omic integration. Particularly noteworthy is scPaLM (Chen et al., 2024), which introduces biological pathway-aware representations to address computational challenges in transformer-based approaches.

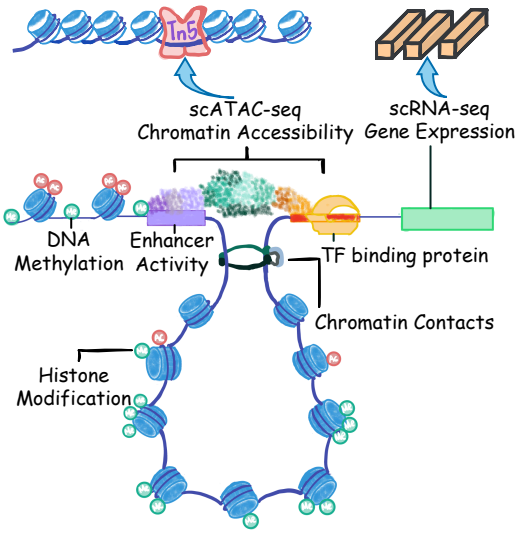

Figure 1: Gene regulatory process in scATAC-seq and scRNA-seq modalities. Image credit to Bonev et al. (2024).

However, despite their successes, current RNA FMs face fundamental limitations rooted in their reliance on expression data alone. Three key challenges persist in existing approaches. First, as shown in Fig. 1, while current models learn gene-gene correlations implicitly, they lack explicit integration of *regulatory causality* derived from chromatin accessibility data – a crucial determinant of cellular identity (Bravo González-Blas et al., 2023). Second, existing methods struggle to capture the multi-scale nature of gene regulation, where relationships operate at both *cell-type-specific* and *cell-specific* (Kamimoto et al., 2020). Third, se-

Findings of the Association for Computational Linguistics: ACL 2025, pages 3805–3819
July 27 - August 1, 2025

vere *information asymmetry* plagues regulatory networks: for some cell types, transcription factors (TFs) exhibit dense connectivity while $\sim 40\%$ of genes lack reliable regulatory links (Aibar et al., 2017; Bravo González-Blas et al., 2023), creating a topological imbalance that standard architectures cannot effectively handle (Chen et al., 2021).

To handle these challenges, we present `GRNFormer`, a novel architecture that integrates multi-scale Gene Regulatory Networks (GRNs) into current RNA FMs through three key innovations. First, we introduce a systematic pipeline for constructing *cell-specific* and *cell-type-specific* GRNs capturing *regulatory causality* derived from chromatin accessibility data by integrating single-cell ATAC-seq (scATAC-seq) and scRNA-seq data utilizing SCENIC+ (Bravo González-Blas et al., 2023). As shown in Fig. 2A and Appendix A, our method leverages chromatin accessibility to identify enhancer-driven regulatory units (eRegulons) through motif enrichment analysis and multi-modal linkage (Bravo González-Blas et al., 2023), enabling discovery of context-specific regulatory relationships across biological scales.

Then, we introduce a universal structure-aware integration framework that utilizes the multi-scale gene regulation information and addresses GRNs topological challenges through: i) an adaptive cross-attention layer that dynamically weights regulatory signals based on node centrality and ii) a biologically informed edge perturbation strategy that supplements sparse connections with co-expression relationships as shown in Fig. 2C. This design enables effective knowledge transfer from GRNs while mitigating *information asymmetry* – a critical advancement over naive fusion approaches such as addition or concatenation.

Lastly, we establish comprehensive benchmarks across three clinically-relevant tasks: gene perturbation prediction, drug response classification, and single-cell sensitivity analysis. Our experiments demonstrate that `GRNFormer` achieves consistent improvements over base models (scGPT $+3.6\%$ Pearson Correlation Coefficient (PCC) on the drug response prediction task, scFoundation $+4.1\%$ Area Under the ROC Curve (AUC) on the single cell drug response classification task). Notably, the model reveals interpretable attention patterns aligning with known biological regulations. Our key contributions are three folds:

❶ *Multi-scale GRN Construction Pipeline:* The first systematic framework integrating scATAC-seq and scRNA-seq data to build cell-type-specific and single-cell-resolution regulatory networks through enhancer-driven eRegulons analysis pipelines.

❷ *Structure-aware Model Architecture:* An integration strategy combining adaptive cross-attention with novel biological guided edge perturbation strategy, effectively resolving GRNs topological imbalance while maintaining computational efficiency.

❸ *Extensive Biological Validation:* State-of-the-art performance across three therapeutic development tasks, with demonstrated improvements in drug response prediction (*e.g.*, 3.6% of $\mathrm{PCC}_{\mathrm{delta}}$ gain against baselines) and single cell drug sensitivity classification (*e.g.*, 0.122 of AUC gain against baselines)

The success of `GRNFormer` underscores the transformative potential of integrating regulatory prior knowledge from different modalities into foundation models. Our work establishes a new paradigm for developing biologically grounded AI systems in computational genomics, with immediate applications in the discovery of drug targets and the improvement of existing gene therapies.

2 Related Works

Single-cell Data Analysis. Single-cell RNA sequencing (scRNA-seq) has revolutionized genomics by enabling profiling of cell-level gene expressions (Saliba et al., 2014; Kolodziejczyk et al., 2015). Providing hints on cellular heterogeneity, scRNA-seq transforms how to understand complex biological systems such as neural tissues, immune responses, and tumor micro-environments. Advances in perturbation sequencing techniques, such as Perturb-seq, have further allowed researchers to discover the causal relations between gene perturbations and cellular phenotypes by utilizing CRISPR-based editing alongside scRNA-seq (Dixit et al., 2016; Adamson et al., 2016; Norman et al., 2019). However, integrating information from other omics modalities, such as scATAC-seq or spatial omics, remains a significant challenge despite the remarkable progress in scRNA-seq technologies (Cui et al., 2023; Xiong et al., 2023).

Foundation Models in Single-cell Omics. FMs, first developed for natural language processing, have become powerful tools for learning hidden

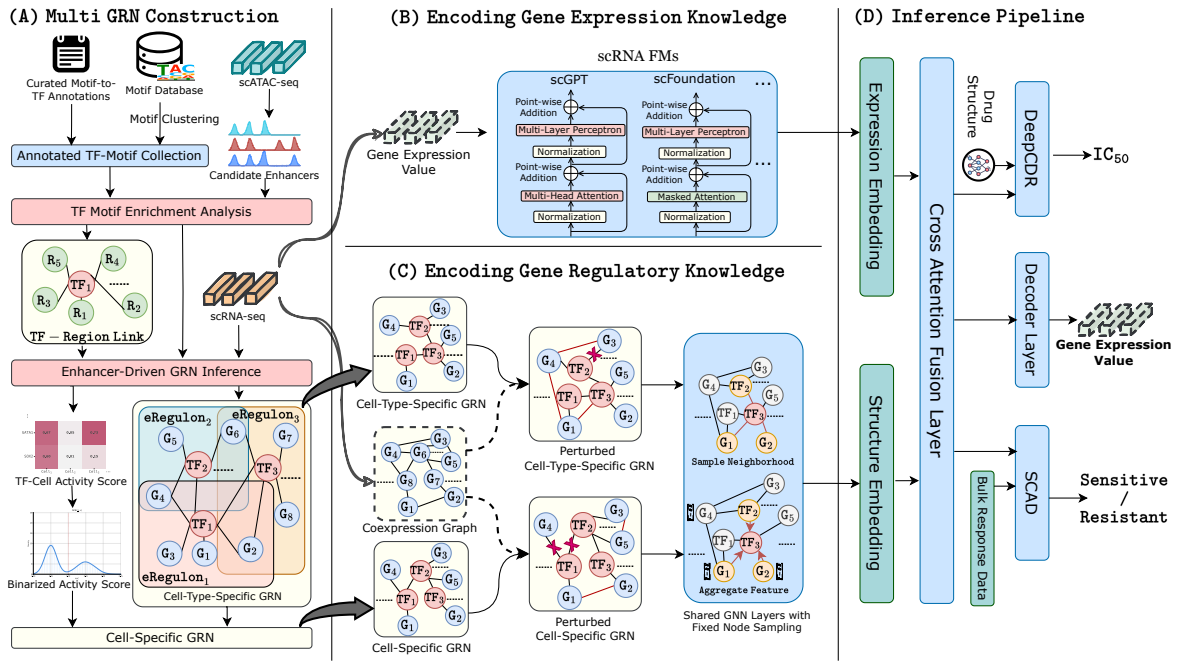

Figure 2: Overview of **GRNFormer** framework: (A) Multi-scale GRN construction from scATAC/scRNA-seq data utilizing additional Motif databases; (B) Our framework employs single-cell RNA foundation models (scRNA FMs) to encode gene expression profiles into expression embeddings, supporting three model architectures as backbones: *scGPT*, *scFoundation*, and *scPaLM*; (C) The multi-scale GRNs are perturbed using co-expression graphs and subsequently processed through GNN modules, with the resulting embeddings aggregated via summation to generate the structure embedding; (D) The expression embedding and structure embedding obtained from the previous two stages are fused through a cross-attention layer. The resulting hybrid embedding can be fed into the decoder for pretraining via masked language modeling objectives, or directly utilized for diverse downstream tasks.

embeddings of large-scale biological data. These models, typically pretrained on vast datasets, can be fine-tuned for downstream tasks such as classifications and translations, offering extensive flexibility and scalability (Bommasani et al., 2021; Moor, 2023). In single-cell biology, foundation models are pre-trained on large single-cell datasets, and then applied to downstream tasks like cell type annotation, perturbation prediction, and multi-omic integration (Cui et al., 2023; Theodoris et al., 2023). During the fine-tuning process, model parameters are further optimized using task-specific datasets typically of much smaller size than the training data, resulting in much lower computational cost(Gururangan, 2020; Qiu, 2020). Additionally, foundation models can recognize various data types , such as transcriptomics and epigenomics, providing a more generalized view of cell biology (Brown et al., 2020; OpenAI, 2023).

Modern scRNA-seq generates gene expression profiles as a cell-by-gene matrix $X \in \mathbb{R}^{N \times G}$, where each element X_{ij} represents the expression count of gene j in cell i. RNA foundation models typically employ masked language modeling objectives adapted to transcriptomic data. Given an input expression vector $x \in \mathbb{R}^G$, these models randomly mask a subset of genes $\mathcal{M} \subset \{1, ..., G\}$ and optimize reconstruction via $\mathcal{L} = \mathbb{E}_x \left[\sum_{i \in \mathcal{M}} ||f_\theta(x^{\text{masked}})_i - x_i||^2 \right]$, where f_θ denotes the foundation model. Key architectural variants include: (1) **scGPT**(Cui et al., 2024) employs generative pretraining with specialized attention masking for non-sequential omics data; (2) **scFoundation** (Hao et al., 2024) introduces a read-depth-aware (RDA) pretraining task using an asymmetric transformer architecture. **scPaLM** (Chen et al., 2024) also tries extending current architecture through pathway-aware architectures. Despite these architectural explorations, current foundation models remain predominantly focused on scRNA-seq data, lacking systematic integration of multi-omics signals such as chromatin accessibility profiles from scATAC-seq data.

3 Methodology – GRNFormer

Overview of GRNFormer. Our approach addresses the challenge of integrating biological prior knowledge of RNA foundation models through a two-stage framework as shown in Fig. 2. First, we leverage multi-omics data to construct reliable gene

regulatory networks (GRNs) at multiple scales - *cell-specific* and *cell-type-specific* levels. These networks capture the complex regulatory relationships between transcription factors and their target genes. Second, we develop a structure-aware integration mechanism that uses cross-attention to incorporate GRNs information into RNA foundation model training while handling the inherent sparsity and topological imbalance of regulations.

3.1 Construction of Multi-scale GRNs

Gene regulatory networks (GRNs) from the computational blueprint of cellular identity, encoding how transcription factors (TFs) – proteins that bind DNA to control gene expression – orchestrate transcriptional programs through *cis*-regulatory elements. Traditional GRN inference methods face two critical limitations: (1) reliance on expression correlations alone, missing causal chromatin accessibility signals; (2) inability to resolve regulatory relationships at both population (cell type) and single-cell levels (Aibar et al., 2017). Our framework addresses these through multi-modal integration and multi-scale analysis as shown in Fig. 2 A. We first begin with cell-type-specific GRNs generation, which we mainly followed SCENIC+ (Bravo González-Blas et al., 2023) pipeline and the details can be found at Appendix A.

Single-cell GRNs via Activity Thresholding. To resolve regulatory heterogeneity within cell types, we quantify eRegulon activity at single-cell resolution using AUCell (Aibar et al., 2017). This algorithm calculates an *Area Under the recovery Curve* (AUC) score by ranking genes or regions and measuring target set enrichment. Critically, the AUC distribution across cells reveals fundamental biological patterns: (1) **Bimodal distributions** indicate two distinct cell subpopulations (active/inactive), while (2) **Skewed Gaussian distributions** reflect graded activation across a continuum (Van de Sande et al., 2020). We model these patterns using a two-component Gaussian mixture:

$$p(x) = \pi_1 \mathcal{N}(x|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(x|\mu_2, \sigma_2^2) \quad (1)$$

where π_i are mixing coefficients. For bimodal cases, the threshold is set at the Gaussians' intersection, cleanly separating active and inactive cells. For skewed distributions with a single dominant component, we label cells in the right tail ($\mu + 2\sigma$) as active, capturing cells with exceptionally strong regulon activity. This biologically-grounded thresholding ensures each cell's GRN comprises only context-relevant regulatory interactions. Examples of the activity distribution of transcription factors and the corresponding thresholds in our pre-training data are illustrated in Appendix E.

Cross-modality Integration. Recognizing that most downstream tasks involve single-modality scRNA-seq datasets, we enable GRN integration through reference mapping. For single omics downstream datasets, we leverage embeddings from pre-trained single multi-omics foundation models (scGPT (Cui et al., 2024), scFoundation (Hao et al., 2024)) to map query cells to their nearest neighbors in the reference space. This method establishes connections between downstream cells and precomputed multi-scale GRNs from paired scATAC-seq and scRNA-seq data, ensuring broad applicability across diverse biological contexts.

3.2 Structure Adapter to Incorporate Gene Regulation

Our structure-aware integration framework focuses on addressing three fundamental challenges in incorporating multi-scale GRNs: (1) *topological imbalance* where TFs dominate connectivity while $\sim 40\%$ of genes lack reliable regulations (Aibar et al., 2017); (2) *information asymmetry* between TF-rich and isolated gene representations; (3) *multi-scale regulatory dynamics* requiring simultaneous modeling of cell-type and single-cell contexts (Bravo González-Blas et al., 2023).

Architecture Adaptation for Different Backbones. As shown in Fig. 2B, our framework utilizes existing RNA FMs to encode gene expressions and demonstrates universal applicability across major variants: (1) For *decoder-only* models (*i.e.*, scGPT (Cui et al., 2024)), we utilize the embedding before the last transformer layer as our expression embedding; (2) For *encoder-decoder* models (*i.e.*, scFoundation (Hao et al., 2024), scPaLM (Chen et al., 2024)), we fuse structural embeddings with encoder outputs before feeding them to the decoder.

Multi-scale GRN Processing. Following the pipeline in Section 3.1, we process cell-specific and cell-type-specific GRNs using GraphSAGE (Hamilton et al., 2017), chosen for its ability to handle *degree imbalance* through fixed-size neighborhood sampling as shown in Fig. 2C. Traditional GNNs that aggregate all neighbors would amplify magnitude differences between high-degree TFs (average degree 81.3 in our data) and remaining genes (average degree 1.3 in our data). For each node v at layer k, the aggregation follows:

$$h^k_{\mathcal{N}(v)} = \text{AGGREGATE}_k\left(\{h^{k-1}_u, \forall u \in \mathcal{N}(v)\}\right)$$
$$h^k_v = \sigma\left(W^k \cdot \text{CONCAT}\left(h^{k-1}_v, h^k_{\mathcal{N}(v)}\right)\right), \quad (2)$$

where $\mathcal{N}(v)$ denotes a fixed-size uniform sample of neighbors, addressing degree imbalance through neighbor sampling as in Hamilton et al. (2017). The final structural embedding $h_{\text{struct}} = h_{\text{cell}} \oplus h_{\text{type}}$ combines regulation information at both scales through element-wise summation.

Cross-modal Fusion. Direct concatenation of GRN embeddings (h_{struct}) with expression features (h_{expr}) amplifies information asymmetry. Instead, our multi-head cross-attention dynamically reweights features. The query-key mechanism prioritizes TF-gene interactions with high topological centrality while attenuating noise from unconnected genes. This mechanism produces context-aware fusion embedding h_{fusion} that complements expression patterns with regulatory constraints.

Edge Perturbation for Topological Balance. Conventional graph augmentations (Zhao et al., 2022) risk involving biologically meaningless connections. Our *biologically-informed perturbation* replaces $\alpha|E|$ edges ($\alpha = 0.2$) with co-expression links from G_{co}, constructed per cell as:

$$G_{\text{co}} = \{(u, v)|x_u > 0 \wedge x_v > 0\}, \forall u, v \in \mathcal{G} \quad (3)$$

where x denotes normalized gene expression, $\mathcal{G}$ denotes the gene vocabulary. This perturbation strategy preserves connectivity for genes lacking regulatory annotations while maintaining biological plausibility – co-expressed genes in the same cell are more likely to share functional relationships (Van de Sande et al., 2020; Roohani et al., 2022). Compared to random edge perturbation, our approach ensures that node embeddings for all non-zero-expressed genes receive sufficient training through the sampling of co-expression graph.

3.3 Pretraining and Inference Pipeline

The training and inference pipeline of our model is illustrated in Fig. 2D. For each backbone architecture, the pretraining objectives and data processing pipelines remain consistent with their original implementations, which primarily involve variants of masked language modeling tasks.

We implemented downstream task pipelines based on scGPT and scFoundation frameworks, with additional integration of scPaLM. Detailed descriptions of these downstream task workflows are provided in the Experiments Section.

4 Experiments

We conducted extensive experiments to evaluate `GRNFormer` across three biologically significant tasks: ❶ **Gene perturbation prediction** examines the model's ability to capture regulatory mechanisms by predicting gene expression changes following gene perturbations. This task is particularly relevant for therapeutic development and understanding disease mechanisms. ❷ **Drug response prediction** evaluates the model's clinical utility by predicting cellular responses to therapeutic compounds. The model integrates gene expression profiles with drug structural information to predict IC50 values (half-maximal inhibitory concentrations). ❸ **Single-cell drug response classification** tests the model's ability to transfer knowledge from bulk cell line to single-cell resolution, a critical capability for personalized medicine. The task involves predicting drug sensitivity for individual cells. Across all these tasks, we will compare our approach against SOTA baselines and conduct comprehensive ablation studies to evaluate our GRN integration strategy's effectiveness systematically. This multi-faceted evaluation framework ensures a thorough assessment of our approach's biological accuracy and practical utility.

4.1 Implementation Details.

Pretraining Data. We pre-trained our model using the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) dataset (Hawrylycz et al., 2024), which provides paired scRNA-seq and scATAC-seq measurements for $113,209$ **cells** from 28 **donors**. The scRNA-seq data captures expression profiles for $18,984$ **protein-coding genes**, while the scATAC-seq data provides chromatin accessibility information across the genome. Detailed statistics about the dataset can be found in Appendix C.

Architectures. Our framework comprises three core components: A *transformer-based RNA foundation model* backbone processing gene expression embeddings; A *GraphSAGE encoder* (Hamilton et al., 2017) generating gene structural embeddings from multi-scale GRNs; and A *cross-attention fusion layer* replacing the final transformer layer to integrate structural and expression features. The architecture preserves the original backbone dimensions (*e.g.*, 768 hidden units for scFoundation).

Training Settings. For scGPT (Cui et al., 2024) and scPaLM (Chen et al., 2024) backbone, we conducted full pretraining on SEA-AD multiome data (Hawrylycz et al., 2024). For scFoundation (Hao et al., 2024) backbone, we performed continued pretraining from their public checkpoint, validating our method's *plug-and-play* capability. All models used backbone-specific hyperparameters from original implementations, including optimizer type, learning rate, and batch size. Training completed on $8\times$A100 GPUs with full reproducibility. Details of the pretraining algorithm with multi-scale GRNs can be found in Appendix B.

Benchmarks Data. We established three evaluation paradigms: (1) *Gene perturbation prediction* using Adamson (Adamson et al., 2016)(87 single gene perturbations in protein response pathway), Dixit (Dixit et al., 2016) (single and combinatorial LPS response gene perturbations), and Norman (Norman et al., 2019) (131 gene pairs and 105 single genes in K562 cells) datasets; (2) *Bulk drug response prediction* via CCLE (Barretina et al., 2012) (24 drugs, 947 cell lines) and GDSC (Iorio et al., 2016) (297 compounds, 969 cell lines); (3) *Single-cell drug classification* following scFoundation's (Hao et al., 2024) protocol for four commonly cancer targeted therapies (Sorafenib, NVP-TAE684, PLX4720, Etoposide). Detailed statistics about these datasets can be found in Appendix C.

Baselines. We established three fundamental baselines: Our implementations of scGPT (Cui et al., 2024) and scPaLM (Chen et al., 2024) pre-trained on SEA-AD multiome data, and the officially pre-trained scFoundation (Hao et al., 2024) checkpoint. For drug response prediction, we additionally compared against DeepCDR (Liu et al., 2020) as a specialized baseline. For single-cell sensitivity classification, we included SCAD (Roohani et al., 2022) to benchmark cell resolution capabilities. All scFoundation results report the maximum performance between its original pre-trained version and our continued pretraining variant for fair comparison.

4.2 Gene Perturbation Prediction

Gene perturbation prediction represents a critical task in computational biology with direct implications for therapeutic development and disease understanding. The task involves predicting genome-wide transcriptional changes following genetic interventions, which is essential for understanding gene function and identifying potential drug targets. A key challenge in this task is capturing the complex, non-linear effects of gene perturbations on cellular transcriptional programs.

Table 1: Gene perturbation prediction evaluation.

Model	Adamson	Dixit	Norman	Avg. $\mathrm{PCC}_{\mathrm{delta}}\uparrow$
scGPT	0.609	0.130	0.405	$0.381_{\pm 0.240}$
+ GRN (ours)	**0.622**	**0.138**	**0.418**	$\mathbf{0.393}_{\pm 0.243}$
scFoundation	0.483	0.239	0.255	$0.326_{\pm 0.137}$
+ GRN (ours)	**0.487**	**0.241**	**0.283**	$\mathbf{0.337}_{\pm 0.132}$

Our evaluation utilized three widely-used benchmark datasets (Adamson (Adamson et al., 2016), Norman (Norman et al., 2019), and Dixit (Dixit et al., 2016)). The input comprises unperturbed gene expression profiles and perturbation gene targets, while the output comprises predicted post-perturbation expression levels. We focused on the Pearson correlation coefficient on differential expression ($\mathrm{PCC}_{\mathrm{delta}}$), which measures how well the model predicts expression changes directions.

As shown in Table 1, `GRNFormer` achieves consistent improvements across all datasets. The GRN-enhanced scGPT variant attains a 1.1% average PCC increase (0.393 vs. 0.381 baseline), with particularly robust gains on the Norman dataset ($+3.1\%$). We adapted each model's native pipeline for gene perturbation prediction, with critical divergence in fine-tuning strategies: *scFoundation* employed parameter freezing for most layers due to GPU memory constraints, while *scGPT* permitted full parameter updates. This architectural distinction likely contributes to scFoundation's relatively lower performance, as partial fine-tuning may limit its adaptability to perturbation patterns.

4.3 Cancer Drug Response Prediction

Accurate prediction of cancer drug responses enables personalized treatment strategies and accelerates therapeutic development (Barretina et al., 2012; Iorio et al., 2016). We evaluate our model on CCLE and GDSC datasets using IC50 values (half-maximal inhibitory concentration) as ground truth. All experiments were repeated four times with identical settings except for random seed variations, with means and standard deviations calculated. We integrate gene expression profiles with drug structural information through DeepCDR-style architecture (Liu et al., 2020; Hao et al., 2023).

Our evaluation utilized data from the Cancer Cell Line Encyclopedia (CCLE) and Genomics of Cancer Drug Sensitivity (GDSC) databases (Iorio et al., 2016; Barretina et al., 2012). The model integrates gene expression profiles with drug structural

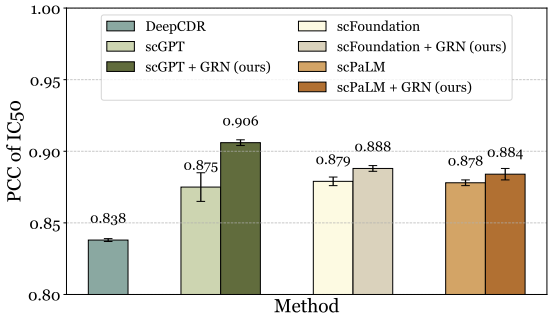

Figure 3: Cancer drug response prediction evaluation.

information to predict drug sensitivity. As shown in Fig. 3, our GRN-enhanced approach achieves superior performance across different experimental settings. Our model achieves a correlation coefficient of $0.906 \pm 0.0.002$, significantly outperforming both DeepCDR (0.838 ± 0.001) and the baseline scGPT model (0.875 ± 0.010). Furthermore, as shown in Fig. 4, our GRN-integrated model demonstrates superior performance over the baseline across all cancer types. The enhanced model exhibits consistently better predictive capability than baseline approaches under most cell lines and drug conditions, achieving robust performance improvements across different experimental settings.

4.4 Single-Cell Drug Response Classification

Single-cell drug response classification presents a unique challenge in cancer research, requiring drug sensitivity prediction at the individual cell resolution. This task is particularly challenging due to the limited availability of single-cell drug response data and the need to transfer knowledge from bulk-level pharmacogenomic data to single cells (Zheng et al., 2023; Hao et al., 2023).

Table 2: Single-cell drug response classification. Superior model between backbone and GRN (ours) is bolded, while the best performance for each drug is underlined.

Model	Etoposide	NVP-TAE684	PLX4720	Sorafenib	Avg. AUC ↑
SCAD	0.696	0.613	0.380	0.572	0.565±0.134
scGPT	**0.511**	0.415	0.563	0.346	0.459±0.097
+ GRN (ours)	0.510	**0.663**	**0.678**	**0.474**	**0.581**±0.104
scFoundation	0.596	0.750	**0.694**	0.807	0.712±0.090
+ GRN (ours)	**0.663**	**0.760**	0.598	**0.953**	**0.743**±0.155
scPaLM	0.471	**0.730**	0.502	0.299	0.500±0.177
+ GRN (ours)	**0.483**	0.468	**0.689**	**0.602**	**0.561**±0.105

We evaluated our model on four drugs (Sorafenib, NVP-TAE684, PLX4720, and Etoposide). Performance was assessed using the Area Under the ROC Curve (AUC) for classification accuracy. Table 2 demonstrates `GRNFormer`'s superiority across most settings. Benefiting from the integration of GRN information, our model achieved a 4.4% performance improvement on scFoundation, surpassing the previous SOTA. For each drug, we report average performance metrics computed through five-fold cross-validation.

4.5 Ablation Studies

Effectiveness of GRN Types. We first investigate how different GRN construction strategies influence model performance. We evaluate four variants: (1) *Random GRN*: Randomly generated networks with matched edge counts; (2) *Cell-type Specific*: GRNs constructed using SCENIC+ at cell population level; (3) *Cell-specific*: Single-cell resolution GRNs via AUCell thresholding; (4) *Hybrid*: Our proposed combination of cell-type and cell-specific GRNs. Experiments are conducted on the scGPT backbone with identical hyperparameters across all variants.

Table 3: Variants of GRN types (Backbone: scGPT)

GRN Type	Drug Response PCC ↑
No GRN	0.875 ±0.010
Random	0.892 ±0.006
Cell-type Specific	0.901 ±0.003
Cell-specific	0.902 ±0.002
Hybrid (Ours)	**0.906** ±0.002

Table 3 demonstrates that our hybrid approach achieves superior performance, with relative improvements of 0.5% in drug response prediction compared to single-scale GRNs. The cell-specific and cell-type-specific variants show better performance than random networks, suggesting the importance of capturing regulatory information.

Impact of Edge Perturbation Strategies. We next analyze the effectiveness of our biologically informed edge perturbation strategy. Two variants are compared: (1) *Random Perturbation*: 20% edges randomly replaced; (2) *Co-expression Guided*: Our proposed strategy using gene co-expression patterns. Experiments are conducted on scPaLM using identical training protocols.

Table 4: Edge perturbations (Backbone: scPaLM)

Edge perturbation strategies	Drug Response PCC ↑	Response Classification AUC ↑
No Augmentation	0.870 ±0.006	0.555 ±0.113
Random Perturbation	0.867 ±0.002	0.548 ±0.108
Co-expression Guided (Ours)	**0.884** ±0.004	**0.561** ±0.105

As shown in Table 4, our co-expression guided perturbation achieves 1.6% relative improvements over the baseline in the drug response prediction tasks. It is noteworthy that simple random perturbation-based data augmentation may degrade model performance, highlighting the necessity of our co-expression guided perturbation strategy.

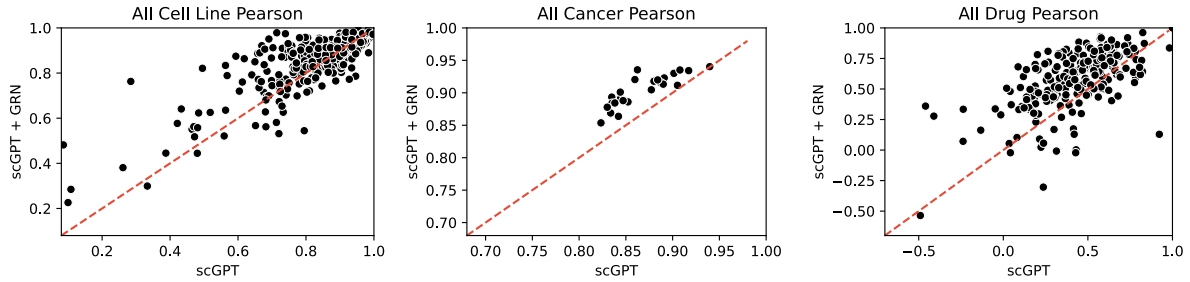

Figure 4: Pairwise visualization of the Pearson correlation coefficient of scGPT and scGPT + GRN based on different grouping strategies. Left: grouping with respect to the cell lines; Middle: grouping with respect to the cancer type; Right: grouping with respect to the drug type. The red lines indicate the relationship of $y = x$.

Analysis of GNN Architectures. We further examine how different GNN architectures affect model performance when integrated with scFoundation. We compare three popular GNN variants: (1) *GCN*: Standard graph convolutional networks (Kipf and Welling, 2016); (2) *GIN*: Graph isomorphism networks (Xu et al., 2018); (3) *GraphSAGE*: Our choice with the neighbor sampling approach.

Table 5: Variants of GNN types (Backbone: scFoundation)

GNN Type	Drug Response PCC ↑	Response Classification AUC ↑
GCN	0.881 ±0.007	0.675 ±0.014
GIN	0.876 ±0.006	0.623 ±0.138
GraphSAGE (Ours)	**0.888**±0.002	**0.743** ±0.155

Table 5 reveals that GraphSAGE performs best while maintaining computational efficiency. The 1.4% improvement in response prediction over GIN demonstrates the effectiveness of neighbor sampling for handling GRN sparsity.

4.6 Analysis of Attention Patterns

To investigate how our model leverages gene regulatory relationships, we analyze the attention patterns in the cross-attention fusion layer. Let $\mathbf{A}^{(h)} \in \mathbb{R}^{N\times N}$ denote the attention matrix for head h in the multi-head cross-attention mechanism, where N is the number of genes. Each entry $a_{ij}^{(h)}$ represents the attention weight between query gene i (from the RNA FM) and key gene j (from the GNN encoder). We compute the *gene-wise attention importance score* ϕ_j for each gene j by averaging across all heads and query genes as $\phi_j = \frac{1}{H\cdot N}\sum_{h=1}^{H}\sum_{i=1}^{N} a_{ij}^{(h)}$, where H is the number of attention heads. This score quantifies how frequently a gene's regulatory embedding influences other genes' expression representations.

To identify biologically meaningful patterns, we calculate the *transcription factor (TF) enrichment ratio* ρ:

$$\rho = \frac{\mathbb{E}[\phi_j | j \in \mathcal{T}]}{\mathbb{E}[\phi_j | j \notin \mathcal{T}]}, \tag{4}$$

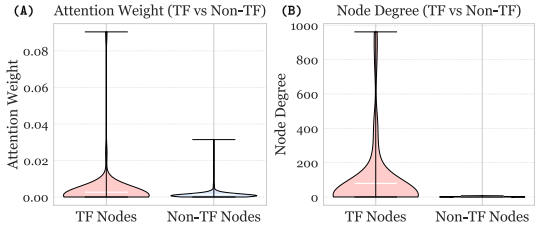

Figure 5: (A) Distribution of average attention scores for transcription factor (TF) and non-transcription factor (non-TF) nodes; (B) Node degree distributions for these two types of nodes. TF nodes appear to connect to more genes and also exhibit higher attention weights.

where $\mathcal{T}$ denotes the set of transcription factors in our GRNs. $\rho > 1$ indicates preferential attention to TFs. Our analysis reveals $\rho = 2.011$ across all cell types on the drug response prediction task, indicating the model attends disproportionately to TFs. The distributions of node degrees for TF and non-TF nodes, as well as the cross attention weights in the fusion layer, are shown in Figure 5.

5 Conclusion

In this paper, we propose `GRNFormer`, a framework that systematically integrates *multi-scale gene regulatory networks* into RNA foundation models through two key innovations: (1) hierarchical GRN construction via multi-omics fusion, and (2) a structure-aware adapter combining adaptive cross-attention with biologically informed edge perturbation to resolve the topological imbalance. `GRNFormer` achieves consistent performance improvement across therapeutic development tasks. Attention analysis reveals biologically meaningful patterns of our edge perturbation strategy. The framework's universal applicability is validated through the integration with major RNA foundation architectures, establishing a new paradigm for biologically grounded AI in computational genomics.

Limitations

Dependency on Regulatory Databases. The quality of our constructed GRN relies heavily on existing motif databases and chromatin accessibility data. Similar to SCENIC+ (Bravo González-Blas et al., 2023), our approach cannot fully resolve ambiguous TF binding patterns within shared motif families. Future integration of emerging techniques like GET-style pseudobulk chromatin profiles (Fu et al., 2025) probably could further improve the reliability of gene regulatory information.

Multi-modal Data Requirement. While our framework theoretically supports single-modality data through reference mapping, optimal GRN construction requires paired scRNA-seq/scATAC-seq data. Future work could try integrate lifelong learning strategies to reduce multi-modal dependency through atlas-scale data integration (Yuan and Duren, 2024). Additionally, inspired by GET (Fu et al., 2025), constructing pseudo-paired multi-omics data from existing resources may better leverage heterogeneous datasets.

Ethics Statement

Our work on integrating multi-scale gene regulatory networks into RNA foundation models demonstrates a commitment to advancing biomedical AI while adhering to ethical research practices. All datasets used in this study listed in Table 6 are publicly available and fully anonymized, with all donor identities and sensitive metadata removed in compliance with privacy regulations. While our model shows promise in accelerating drug discovery and improving gene therapies, any clinical application must undergo rigorous ethical review to ensure compliance with genomic data protection standards. We emphasize that biological foundation models built upon our methodology should incorporate safeguards against misuse, such as restricting access to potentially harmful gene-editing predictions. Furthermore, our implementation prioritizes transparency—all code and preprocessing workflows are designed for public auditability, reproducibility, and explainability.

Acknowledgment

This manuscript has been authored by Lawrence Livermore National Security, LLC under Contract No. DE-AC52-07NA27344 with the U.S. Department of Energy. This material is based upon work supported by the Department of Energy, Office of Science, Office of Advance Scientific Computing Research. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This research was, in part, funded by the National Institutes of Health (NIH) under other transactions 1OT2OD038045-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the NIH.

References

Britt Adamson, Thomas M Norman, Marco Jost, Min Y Cho, James K Nuñez, Yuwen Chen, Jacqueline E Villalta, Luke A Gilbert, Max A Horlbeck, Marco Y Hein, et al. 2016. A multiplexed single-cell crispr screening platform enables systematic dissection of the unfolded protein response. Cell, 167(7):1867–1882.

Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, et al. 2017. Scenic: single-cell regulatory network inference and clustering. Nature methods, 14(11):1083–1086.

Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A Margolin, Sungjoon Kim, Christopher J Wilson, Joseph Lehár, Gregory V Kryukov, Dmitriy Sonkin, et al. 2012. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature, 483(7391):603–607.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Boyan Bonev, Castelo-Branco Gonçalo, Fei Chen, Simone Codeluppi, M Ryan Corces, Jean Fan, Myriam Heiman, Kenneth Harris, Fumitaka Inoue, Manolis Kellis, et al. 2024. Opportunities and challenges of single-cell and spatially resolved genomics methods for neuroscience discovery. Nature neuroscience, 27(12):2292–2309.

Carmen Bravo González-Blas, Seppe De Winter, Gert Hulselmans, Nikolai Hecker, Irina Matetovici, Valerie Christiaens, Suresh Poovathingal, Jasper Wouters, Sara Aibar, and Stein Aerts. 2023.

Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. Nature methods, 20(9):1355–1367.

Carmen Bravo González-Blas, Liesbeth Minnoye, Dafni Papasokrati, Sara Aibar, Gert Hulselmans, Valerie Christiaens, Kristofer Davie, Jasper Wouters, and Stein Aerts. 2019. cistopic: cis-regulatory topic modeling on single-cell atac-seq data. Nature methods, 16(5):397–400.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. NeurIPS.

Deli Chen, Yankai Lin, Guangxiang Zhao, Xuancheng Ren, Peng Li, Jie Zhou, and Xu Sun. 2021. Topology-imbalance learning for semi-supervised node classification. Advances in Neural Information Processing Systems, 34:29885–29897.

Xuxi Chen, Zhangyang Wang, Marinka Zitnik, Manolis Kellis, and Tianlong Chen. 2024. Pre-training of single-cell language models through genetic pathway learning. In ICML 2024 Workshop on Efficient and Accessible Foundation Models for Biological Discovery.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. 2024. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. Nature Methods, pages 1–11.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, and Bo Wang. 2023. scgpt: Towards building a foundation model for single-cell multi-omics using generative ai. bioRxiv.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, et al. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. cell, 167(7):1853–1866.

Xi Fu, Shentong Mo, Alejandro Buendia, Anouchka P Laurent, Anqi Shao, Maria del Mar Alvarez-Torres, Tianji Yu, Jimin Tan, Jiayu Su, Romella Sagatelian, et al. 2025. A foundation model of transcription across human cell types. Nature, pages 1–9.

Suchin et al. Gururangan. 2020. Don't stop pretraining: adapt language models to domains and tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 8342–8360.

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

Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Le Song, and Xuegong Zhang. 2023. Large scale foundation model on single-cell transcriptomics. bioRxiv.

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

Michael Hawrylycz, Eitan S Kaplan, Kyle J Travaglini, Mariano I Gabitto, Jeremy A Miller, Lydia Ng, Jennie L Close, Rebecca D Hodge, Brian Long, Tyler Mollenkopf, et al. 2024. Sea-ad is a multimodal cellular atlas and resource for alzheimer's disease. Nature Aging, pages 1–4.

Francesco Iorio, Theo A Knijnenburg, Daniel J Vis, Graham R Bignell, Michael P Menden, Michael Schubert, Nanne Aben, Emanuel Gonçalves, Syd Barthorpe, Howard Lightfoot, et al. 2016. A landscape of pharmacogenomic interactions in cancer. Cell, 166(3):740–754.

Kenji Kamimoto, Christy M Hoffmann, and Samantha A Morris. 2020. Celloracle: Dissecting cell identity via network inference and in silico gene perturbation. BioRxiv, pages 2020–02.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. 2015. The technology and biology of single-cell rna sequencing. Molecular cell, 58(4):610–620.

Qiao Liu, Zhiqiang Hu, Rui Jiang, and Mu Zhou. 2020. Deepcdr: a hybrid graph convolutional network for predicting cancer drug response. Bioinformatics, 36(Supplement_2):i911–i918.

Sundeep Malik, Chang-Fen Huang, and Jakob Schmidt. 1995. The role of the canntg promoter element (e box) and the myocyte-enhancer-binding-factor-2 (mef-2) site in the transcriptional regulation of the chick myogenin gene. European journal of biochemistry, 230(1):88–96.

Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. 2019. Grnboost2 and arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics, 35(12):2159–2161.

Michael et al. Moor. 2023. Foundation models for generalist medical artificial intelligence. Nature, 616(7955):259–265.

Thomas M Norman, Max A Horlbeck, Joseph M Replogle, Alex Y Ge, Albert Xu, Marco Jost, Luke A Gilbert, and Jonathan S Weissman. 2019. Exploring genetic interaction manifolds constructed from

rich single-cell phenotypes. Science, 365(6455):786–793.

OpenAI. 2023. Gpt-4 technical report. Preprint, arXiv:2303.08774.

Xipeng et al. Qiu. 2020. Pre-trained models for natural language processing: a survey. Science China Technological Sciences, 63(10):1872–1897.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. 2022. Gears: Predicting transcriptional outcomes of novel multi-gene perturbations. BioRxiv, pages 2022–07.

Antoine-Emmanuel Saliba, Alexander J Westermann, Stanislaw A Gorski, and Jörg Vogel. 2014. Single-cell rna-seq: advances and future challenges. Nucleic acids research, 42(14):8845–8860.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. 2023. Transfer learning enables predictions in network biology. Nature, 618(7965):616–624.

Tian Tian, Jiquan Wan, Qing Song, and Zemin Wei. 2019. Clustering single-cell rna-seq data with a model-based deep learning approach. Nature Machine Intelligence, 1(4):191–198.

Bram Van de Sande, Christopher Flerin, Kristofer Davie, Maxime De Waegeneer, Gert Hulselmans, Sara Aibar, Ruth Seurinck, Wouter Saelens, Robrecht Cannoodt, Quentin Rouchon, et al. 2020. A scalable scenic workflow for single-cell gene regulatory network analysis. Nature protocols, 15(7):2247–2276.

Jialu Wang, Anjun Ma, Yuzhe Chang, Jingyi Gong, Yuqiao Jiang, Ruisheng Qi, and Dong Xu. 2021. scgnn is a novel graph neural network framework for single-cell rna-seq analyses. Nature communications, 12(1):1882.

Woodring E Wright. 1992. Muscle basic helix-loop-helix proteins and the regulation of myogenesis. Current Opinion in Genetics & Development, 2(2):243–248.

Lei Xiong, Tianlong Chen, and Manolis Kellis. 2023. scCLIP: Multi-modal Single-cell Contrastive Learning Integration Pre-training.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826.

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

Tong Zhao, Wei Jin, Yozen Liu, Yingheng Wang, Gang Liu, Stephan Günnemann, Neil Shah, and Meng Jiang. 2022. Graph data augmentation for graph machine learning: A survey. arXiv preprint arXiv:2202.08871.

Zetian Zheng, Junyi Chen, Xingjian Chen, Lei Huang, Weidun Xie, Qiuzhen Lin, Xiangtao Li, and Ka-Chun Wong. 2023. Enabling single-cell drug response annotations from bulk rna-seq using scad. Advanced Science, 10(11):2204113.

A Cell-type-specific GRNs via eRegulon Inference.

We construct hierarchical GRNs using SCENIC+ (Bravo González-Blas et al., 2023), which integrates scATAC-seq and scRNA-seq through three phases:

❶ *Candidate Enhancer Identification:* Chromatin accessibility profiles from scATAC-seq reveal genomic regions where DNA is unwound, indicating potential regulatory elements. **Co-accessible regions** are detected using pycisTopic (Bravo González-Blas et al., 2019), which employs topic modeling – a probabilistic method that groups genomic loci with similar accessibility patterns across cells. These regions, enriched near genes with correlated expression, serve as candidate enhancers – non-coding DNA elements that promote gene transcription.

❷ *TF-Motif Enrichment Analysis:* Transcription factors bind DNA through specific sequence patterns called **motifs** (e.g., the E-box "CANNTG" for basic helix-loop-helix TFs (Wright, 1992; Malik et al., 1995)). Enhancer candidates are scanned against a curated database of $32,765$ TF-binding motifs (aggregated from 29 collections (Bravo González-Blas et al., 2023)) using pycisTarget. Two algorithms identify statistically overrepresented motifs: i) The *cisTarget* algorithm ranks motifs by how early their target regions appear in accessibility-based rankings; ii) The *DEM* algorithm identifies motifs differentially enriched between cell types. These algorithms establish **TF-to-enhancer links** (NES > 3.0, FDR < 0.1) while mitigating false positives through motif clustering.

❸ *eRegulon Construction:* For each TF, we link its target enhancers to genes using three criteria: (1) genomic proximity (± 150kb from gene), (2) expression correlation (Pearson $|r| > 0.03$), and (3) gradient-boosted regression importance scores (GRNBoost2 (Moerman et al., 2019)). This forms **enhancer-driven regulons (eRegulons)** – triplets connecting TFs, enhancers, and target genes that function as regulatory units. Cell-type specificity is determined by joint accessibility of enhancers and expression of target genes (Bravo González-Blas et al., 2023).

B Algorithm

Algorithm 1 formalizes our structure-aware pretraining process, implementing the key components described in §3.2 and §3.3. The pseudocode explicitly shows the edge perturbation strategy (Lines $3-12$) that addresses topological imbalance through co-expression guided augmentation, and the multi-scale fusion mechanism (Lines $14-20$) combining cell-specific and cell-type-specific GRN embeddings. This algorithm complements Fig. 2 in the main text by detailing how biological priors are injected during training while maintaining compatibility with various backbone architectures.

Algorithm 1 Structure-Aware (Continue) pretraining with Multi-scale GRN

Input: Masked gene expression vector x, cell-specific GRN G_{cell}, cell-type-specific GRN G_{type}, GNN encoder F, Transformer backbone H, cross-attention module P, perturbation ratio α, fusion weight β
Output: Reconstructed expression $\bar{x}$
function PERTURBGRN(G, G_{co}, α)
    $V \leftarrow \text{nodes}(G)$
    $E_{\text{original}} \leftarrow \text{edges}(G)$
    $E_{\text{drop}} \leftarrow \text{Sample}(E_{\text{original}}, \alpha|E_{\text{original}}|)$
    $E_{\text{co}} \leftarrow \text{Sample}(\text{edges}(G_{\text{co}}), \alpha|E_{\text{original}}|)$
    return $(V, E_{\text{original}} \setminus E_{\text{drop}} \cup E_{\text{co}})$
// Stage 1: Graph Augmentation
$G_{\text{co}} \leftarrow \text{ConstructCoExpressionGraph}(x)$
$\tilde{G}_{\text{cell}} \leftarrow \text{PERTURBGRN}(G_{\text{cell}}, G_{\text{co}}, \alpha)$
$\tilde{G}_{\text{type}} \leftarrow \text{PERTURBGRN}(G_{\text{type}}, G_{\text{co}}, \alpha)$
// Stage 2: Structural Encoding
$h_{\text{cell}} \leftarrow F(\tilde{G}_{\text{cell}})$ ▷ Cell-specific encoding
$h_{\text{type}} \leftarrow F(\tilde{G}_{\text{type}})$ ▷ cell-type-specific encoding
$h_{\text{struct}} \leftarrow h_{\text{cell}} \oplus h_{\text{type}}$ ▷ Element-wise sum
// Stage 3: Cross-modal Fusion
$h_{\text{expr}} \leftarrow H(x)$ ▷ Gene expression embedding
$h_{\text{fusion}} \leftarrow P(h_{\text{expr}}, h_{\text{struct}})$ ▷ Cross-attention fusion
$h_{\text{combined}} \leftarrow h_{\text{expr}} + \beta h_{\text{fusion}}$ ▷ Weighted combination
$\bar{x} \leftarrow \text{Decoder}(h_{\text{combined}})$
return $\bar{x}$

C Datasets

Table 6 summarizes key statistics for all experimental datasets. The SEA-AD multiome dataset provides paired scRNA-seq/scATAC-seq profiles for pretraining, while the perturbation benchmarks (Adamson, Dixit, Norman) and drug response datasets (CCLE, GDSC) enable comprehensive downstream evaluation across different biological contexts.

Table 6: Summary of datasets used in different tasks.

Task	Dataset	# of cells/# of cell lines	# of genes
Training: Mask Language Modeling	SEA-AD (multiome part)(Hawrylycz et al., 2024)	113,209	18,984
Gene Perturbation Prediction	Adamson(Adamson et al., 2016)	68,603	5,060
	Dixit(Dixit et al., 2016)	447,35	5,012
	Norman(Norman et al., 2019)	91,205	5,045
Drug Response Prediction/ Single Cell Drug Response Classification	CCLE(Barretina et al., 2012)	947	1651
	GDSC(Iorio et al., 2016)	969	$\sim$ 22,000

D Additional Experiment Results

Unsupervised Cell-type Clustering. To further validate our work's effectiveness, we supplemented our evaluation with clustering tasks. We conducted unsupervised cell-type clustering experiments on several datasets, obtaining cell embeddings from our model followed by k-means clustering, and calculating the Adjusted Rand Index (ARI) using cell type labels. We compared our approach with two classic machine learning clustering methods, scDeepCluster (Tian et al., 2019) and scGNN (Wang et al., 2021). Our method achieved good results on most datasets, as shown in Table 7. This addition ensures our work covers all downstream tasks addressed in the scFoundation paper.

Table 7: Supplementary Table: Unsupervised cell-type clustering results (ARI). Performance of our method (scFoundation + GRN) compared to other methods across various datasets. Average ARI indicates overall clustering performance. The $\pm$ values indicate standard deviation.

Method	Baron	Chen	Endothelium	Muto	Pancreas	Average ARI $\uparrow$
scDeepCluster (Tian et al., 2019)	0.502 ± 0.000	0.287 ± 0.005	0.637 ± 0.000	0.625 ± 0.031	0.414 ± 0.000	0.493
scGNN (Wang et al., 2021)	0.556 ± 0.027	0.322 ± 0.010	0.581 ± 0.007	0.576 ± 0.023	0.457 ± 0.057	0.4984
scFoundation	0.453 ± 0.051	0.322 ± 0.014	0.512 ± 0.003	0.385 ± 0.005	0.358 ± 0.026	0.406
scFoundation + GRN (Ours)	$\mathbf{0.490 \pm 0.042}$	$\mathbf{0.447 \pm 0.015}$	$\mathbf{0.649 \pm 0.033}$	$\mathbf{0.549 \pm 0.032}$	$\mathbf{0.436 \pm 0.064}$	**0.5142**

Comparison of Different Regulatory Knowledge Integration Methods. We investigated various methods for integrating Gene Regulatory Network information with gene expression embeddings. In the early stages of our project, we experimented with simple concatenation and addition methods but found they did not yield optimal performance. Analysis of embedding magnitude distributions for each gene revealed significant differences, likely due to information imbalance between nodes. This observation motivated our adoption of cross-attention for fusion. We conducted ablation experiments on the scFoundation backbone for Drug Response Prediction and Single Cell Drug Sensitivity Classification tasks. As shown in Supplementary Table 8, the cross-attention method demonstrated significant superiority.

Table 8: Supplementary Table 1: Comparison of different fusion methods for integrating GRN information with gene expression embeddings on the scFoundation backbone. Performance is evaluated on Drug Response Prediction (PCC) and Single Cell Drug Sensitivity Classification (AUC). The $\pm$ values indicate standard deviation.

Fusion Method	Drug Response PCC $\uparrow$	Response Classification AUC $\uparrow$
Baseline (scFoundation)	0.875 ± 0.010	0.712 ± 0.090
+ GRN (Addition)	0.884 ± 0.002	0.621 ± 0.133
+ GRN (Concatenation)	0.881 ± 0.003	0.538 ± 0.075
+ GRN (Cross-Attention) (Ours)	$\mathbf{0.906 \pm 0.004}$	$\mathbf{0.743 \pm 0.155}$

E Transcription Factor Activity Distribution

Fig. 6 visualizes the bimodal and skewed AUC distributions underlying the single-cell GRN construction, supporting the thresholding methodology from §3.1. The clear separation of active/inactive states for TFs like PURA empirically validates the gaussian mixture modeling approach. These distributions directly inform the cell-specific regulatory networks that drive our model's performance improvements in downstream tasks (§4.5).

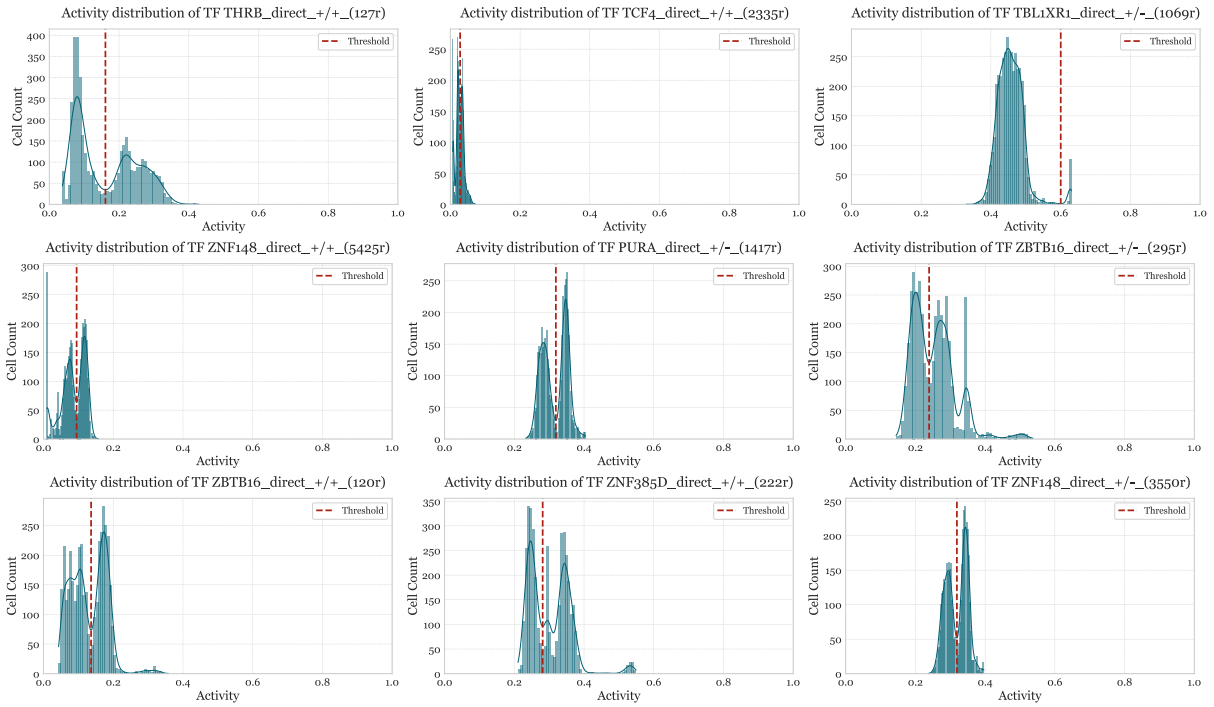

Figure 6: The distribution of activity levels for nine randomly selected transcription factors (TFs) within a single cell type. The threshold distinguishing active versus inactive states are demarcated by red vertical lines.

F Potential Risks

While GRNFormer advances computational genomics, three key risks warrant consideration: **(1) Data Bias Propagation:** Reliance on existing motif databases may propagate biases in TF-gene interactions, particularly for understudied cell types or minor populations, potentially leading to skewed therapeutic predictions. **(2) Privacy Vulnerabilities:** Although using anonymized data, integration of multi-omics profiles could theoretically enable cell identity re-identification through rare regulatory signatures. **(3) Dual-Use Concerns:** Enhanced prediction of gene regulatory outcomes might be misused to design targeted biological agents, though our current implementation focuses only on therapeutic contexts. We mitigate these risks through (1) transparent documentation of data sources, and (2) controlled access to regulatory network components. Responsible deployment requires ongoing collaboration with bioethicists and clinical reviewers.