Augmented Neural ODEs Mitigate Distribution Shifts in Gene Expression Prediction Under Single-Gene Perturbations

Jingkai Wang[1], H. Wu[1], and J. Fang[1]

[1]Columbia University

Abstract

Understanding gene expression responses under genetic perturbations is crucial for unraveling regulatory mechanisms and identifying therapeutic targets. Single-cell RNA sequencing, particularly large-scale Perturb-seq datasets, has catalyzed the emergence of machine learning approaches for predicting transcriptional outcomes of gene perturbations. We evaluate the performance of three benchmark models in capturing gene expression responses under single-gene perturbation: GEARS model, the linear model and graph-informed Neural ODE (NODE) model. The graph-informed NODE model is inspired on the PerturbODE model while training on the GNN encoder. We highlight the role of Gene Regulatory Network (GRN) in predicting gene expression responses by replacing the original network with a randomized GRN. To overcome the limitations of Neural ODE model in modeling nonlinear dynamics, we propose a novel model, which is called AugPerturbODE model. The AugPerturbODE model introduces an auxiliary latent matrix to capture disproportionately large or non-linear expression shifts in predicting the gene expression responses to genetic perturbation. Our findings show that AugPerturbODE attains strong MSE performance and achieves the lowest KL divergence, indicating that ANODE framework offers a biologically interpretable, distributionally faithful solution for predicting the expression distributions under single-gene perturbations. This work suggests that ANODE serves as a strong baseline for mitigating shift behaviors in gene expression prediction and highlights the potential of incorporating gene ontology to enhance generalizability and interpretability.

1 Introduction

The approach to modeling transcriptional responses to single-gene perturbation and multiple-gene perturbation lays the groundwork for the development of modern biology. As biological technologies like CRISPR-based perturbation screens have become increasingly accessible, and given the availability of Perturb-seq datasets, researchers have begun to explore the computational algorithms that predict the gene expression responses to chemical perturbations, leading to the proliferation of advanced models and deeper insights into gene regulatory dynamics. Because accurately capturing the gene expressions after chemical or genetic perturbations is essential for designing targeted therapies and dissecting regulatory mechanics, there is a pressing need to minimize the prediction error. To address this, models are suggested to extract biological patterns and represent the nonlinear dependencies and dynamic behaviors. As a result, the advanced models aim to determine biological regulatory effects and incorporate biological features utilizing perturbation-level data or embeddings, though different models vary in their ability to balance generalization performance and represent algorithm structures.

In this work, we introduce three benchmark models that capture gene expressions under single-gene perturbations. The GEARS model represents the structure of the gene regulatory networks (GRN) through GNN-based encoders; the linear model uti-

lizes the basic regression structure for gene expression prediction; and the graph-informed Neural ODE model is based on Neural Ordinary Differential Equation (NODE) to measure the continuous dynamics of gene expression responses to genetic perturbations. In GEARS model, the role of Gene Regulatory Network (GRN) is analyzed by comparing the performance of the original structure and a random structure. To address the limitations of these approaches in modeling the non-linear dynamics, we propose a novel model, AugPerturbODE model, which is based on the Augmented Neural Ordinary Differential Equation (ANODE). The ANODE framework has advantages in learning the non-linear responses, analyzing high-order dependencies, and measuring large gene expression changes under perturbations. The auxiliary latent matrix in the AugPerturbODE model improves the effectivity and avoids shift behaviors in capturing transcriptional expression responses to single-gene perturbations.

Our results show that the linear model is competitive in predicting gene expressions under single-gene perturbations, and the AugPerturbODE model achieves the best performance in preserving the overall distributions of gene expression responses, especially in avoiding the spurious shifts in the predicted expression distributions. This is essential for preventing the biologically implausible artifacts and capturing regulatory heterogeneity at the single-gene level.

2 Related Work

Various approaches have been proposed to capture the predicted gene expression responses to the chemical or genetic perturbations, prompting the development of distinct methods ranging from data-based models to biology-inspired machine learning models. Among one of the earliest models to simulate the perturbation outcomes, baseline regression models are introduced by assuming the linearity in how perturbations affect gene expressions, serving as the benchmark for evaluating more complicated models. The utilization of simple baseline frameworks as comparisons is demonstrated in researches such as Lotfollahi etc. (2019) and Duan & Lintao (2023). According to Haque etc. (2017), single-cell RNA sequencing (scRNA-seq) provides insights into gene expression variations across individual cells. The scRNA-seq method is a powerful tool that provides gene expression data leveraging gene-to-cell expression matrices. Pretrained on large-scale gene expression profiles based on scRNA-seq, the scGPT is based on the transformer algorithm, utilizing self-attention to study the relationships between different genes and predict perturbation-induced transcriptomic responses, referring to Roohani etc. (2023). Meanwhile, motivated by the complexity of regulatory networks and Graph Neural Networks (GNN), GEARS model was designed to capture the biological relationships using a graph-based architecture of gene regulatory networks. This approach can effectively increase the accuracy of predicting gene expression changes resulting from chemical perturbations compared to baseline models, referring to Roohani etc. (2023). As presented in Ahlmann-Eltze etc. (2024), the simple linear model trained on the perturbation embeddings can guarantee better performance in measuring unseen perturbations compared to GEARS and scGPT.

A growing number of approaches predict the gene expression responses using neural differential equations. Neural Ordinary Differential Equation (NODE), first proposed by Chen etc. (2018), is motivated by the continuous-time dynamical system. Leveraging NODE, the hidden states are transformed continuously to determine the optimal coefficients or weight matrices. The model is trained with the adjoint sensitivity method, supporting gradient-based optimization through backward ODE solvers such as backward Euler's method or linear multi-step methods. As presented by Lin etc. (2025), PerturbODE extends the NODE to predict perturbed responses in single-gene perturbation. The predicted gene expressions are modeled by time-evolving ODE integration governed by the learned dynamics.

3 Data

We leveraged publicly available single-gene perturbation data from the GEO dataset GSE90063. The dataset integrates CRISPR knockouts with single-cell RNA sequencing (scRNA-seq) to profile gene expression responses. In our work, we use the Perturb-seq subset that measures the transcriptional responses in dendritic cells (DCs) subjected to single-gene perturbations.

To prepare the dataset for training and predicting, we merged the raw sparse matrix, gene and cell

annotations, and mapped cell barcodes to CRISPR-based gene perturbation labels. The gene expression counts across all the samples were normalized, and the 5000 most informative genes were selected to reduce noise. To ensure the coverage of all knockouts, the missing perturbation genes excluded from the 5000 most informative genes were restored.

The dataset was split into the training dataset and the testing dataset. The training dataset is used for parameter identification of different models, while the testing dataset is leveraged to examine the performance of predicting gene expressions. The separation is crucial to avoid overfitting and supports comparison across different modeling techniques. This dataset is highly relevant to our goal of predicting gene expression responses under single-gene perturbation. When dealing with the gene embeddings, the embeddings are separated based on the corresponding samples in the training dataset.

4 Methods

We leverage four different models to predict single-gene perturbation and multiple-gene perturbation responses, including GEARS, linear model, PerturbODE and ANODE. Each model employs the perturbation embedding as the training input, derived from the GNN-based encoder. Different model frameworks exhibit distinct model priors and learning biases, influencing their performance in capturing gene expression patterns.

4.1 Benchmarks Overview

Based on Graph Neural Networks (GNNs), the GEARS model introduces an innovative approach to capture the gene expressions after perturbations. Utilizing GNN-based encoders, the gene embedding and the perturbation gene embedding are derived based on covariances. The two embeddings are merged through a compositional module to evaluate the post-perturbation gene expressions. From the integration of gene ontology (GO) graphs, the ability of the model in examining the gene regulatory mechanisms is enhanced. Then the autofocus direction-aware loss is leveraged during the training process of the model.

Utilizing a straightforward regression framework, linear model exhibits simple structures compared to the GEARS model. The perturbation embeddings, containing graph-informed perturbation features, serve as the training input for the linear model. The linear model presumes linear relationships between perturbation genes and corresponding transcriptional responses. The structure for the linear model is defined as

$$\mathbf{x} = \mathbf{x}_0 + uW^T,$$

where $W \in \mathbb{R}^{d\times k}$ denotes the weight matrix, $u \in R^{n\times d}$ means the perturbation embedding matrix, and x_0 is the initial control expressions. The model assumes that gene expression responses can be detected by the linear projection of the perturbation embedding matrix.

The traditional Neural ODE (NODE) is proposed to model the hidden state dynamics as a continuous transformation governed by an ODE:

$$\frac{du}{dt} = f(u, t; \theta).$$

The NODE model enables modeling of smooth, time-continuous evolution and allows adaptive computation via ODE solvers. Biological structure and perturbation control is added in the NODE model, which is specific to gene expression prediction under perturbations

$$\dot{X} = A \cdot \sigma(BX + \beta) \odot M(u) - WX$$

where σ is the activation function (RELU or sigmoid); A, B represents encoded regulatory influence between gene modules and genes; $M(u)$ denotes the linear activation function in terms of the gene embeddings with learnable parameters; W means diagonal decay matrix stabilizing gene expression over time. Graph-informed NODE models gene regulation dynamics under interventions while allowing GRN extraction from the ODE parameters.

The central distinction between the graph-informed NODE model and the PerturbODE model is how the perturbation structures are represented. PerturbODE introduces a latent dynamics of a Neural ODE. The ODE is solved forward and backpropagated alternately given each perturbation, indicating that the training batches are subsetted corresponding to specific perturbations. In contrast, our model integrates the Gene Regulatory Network (GRN) through a GNN encoder, and the perturbation embedding is provided

based on the GRN. The learnable linear structure of the perturbation embeddings is modulated during the training process.

5 Results

5.1 Performance Comparison Across Metrics for Benchmarks

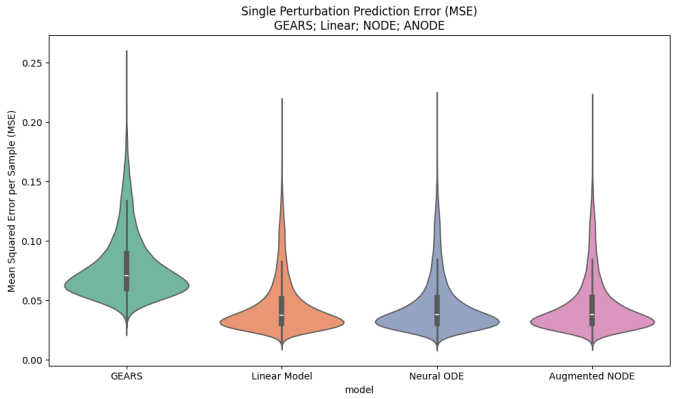
We evaluate the performance of three benchmarks: GEARS, linear model, graph-informed NODE in predicting perturbation responses. The GEARS model is trained on the compositional structure that combines perturbation embedding and gene embedding to predict perturbation responses. Both the linear model and NODE employ perturbation embedding derived from a Graph Neural Network (GNN) trained on perturbation-perturbation mappings, serving as the input to predict genetic responses to perturbations. Different metrics are introduced to analyze the performance of these models. Specifically, Mean Square Error (MSE) measures the average square error between the predicted and real gene expression vectors. The Wasserstein-2 distance (W2) quantifies the difference between the predicted expression distribution and real expression distribution, reflecting the variability in perturbation responses. And Kullback-Leibler (KL) divergence evaluates the efficiency of predicted gene expression by measuring the lost information when approximating the actual expression.

The model performance of three benchmarks in single-gene perturbation is evaluated by analyzing their distribution nuances between the predicted and real gene expressions. The comparison results in evaluation metrics are shown in figure 1. Since the models are trained leveraging the MSE metric, we compute the MSE loss by comparing the mean square error between predicted gene expressions and the ground truth in the testing dataset. Among the benchmark models, the linear model achieves the lowest MSE loss, with a value of 0.0455, slightly outperforming the NODE model in the MSE metric (0.0461), in figure 1a. The GEARS model, unexpectedly, lagged behind with a higher MSE of 0.0787. The cumulative MSE error across top N genes, ranking by either mean expression or standard deviation, indicates the difference in model performance along the most informative genes in each model. All the benchmark models, GEARS model, the linear model and the NODE model, exhibit similar performance in the predicted perturbation expressions that closely resemble the ground truth expressions. However, as N increases, the accumulated MSE loss for the GEARS model is higher compared to the linear model and the NODE, indicating limitations in modeling complex and noisy expressions, as displayed in figure 1b. As a comparison, the linear model and the PerturbODE maintain lower cumulative errors and share similar patterns, with the linear model marginally outperforming the PerturbODE in higher variable genes with substantial differential responses.

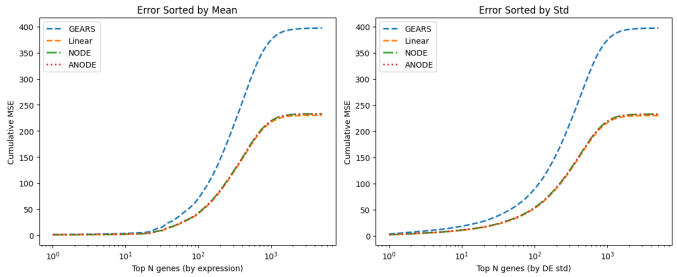
Beyond prediction accuracy, the performance of models capturing the distribution structures across individual genes is displayed in figure 1c. Though the GEARS model attains the largest MSE loss (0.0787), it yields the lowest Wasserstein-2 (W2) distance (0.026), suggesting its strength in capturing the overall distribution of gene-level expression predictions. However, despite the advantages in demonstrating the overall distributions of gene expressions, GEARS model contains the highest KL divergence loss (6.463) among all benchmark models, reflecting the significant mismatches in the shape of the predicted gene expressions. The linear model shows notable improvement in KL divergence (5.956), while the NODE model performs even better (5.577). The KL divergence metric is essential in biological settings, since lower KL divergence indicates the fidelity in gene expression variability and modality. The GEARS model attains higher KL divergence, reflecting its misalignment in predicting distributions, such as shifted modes or missed gene expression states. Poor performance in KL divergence would cause reduced resolution in capturing the underlying patterns of cellular responses. Among the three benchmark models, NODE demonstrates modest MSE loss (0.461) and lowest KL divergence (5.577), highlighting its advantages in increasing prediction accuracy, and in preventing distortions that could impair the biological interpretability of gene expression responses to single-gene perturbation.

5.2 Role of Gene Regulatory Network (GRN) in Prediction Performance

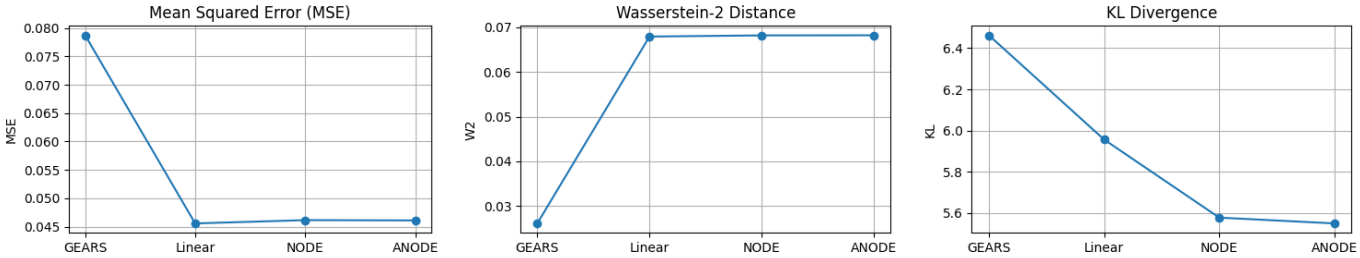
In the GEARS model, the Gene Regulatory Network (GRN) serves as a structured prior, enabling the gene



(a) Prediction MSE loss across models.



(b) Cumulative MSE over top genes sorted by mean and standard deviation.



(c) Mean squared error, Wasserstein-2 distance, and KL divergence.

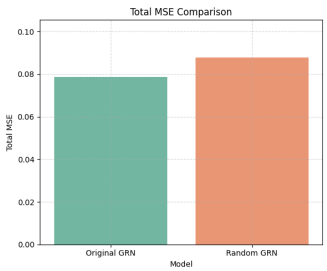
Figure 1: Evaluation of model performance across different metrics.

embeddings to update from the Graph Neural Network (GNN) layers. And the graph-informed embeddings are essential to the prediction of gene expression responses. The GRN is a sparse, graph-based structure including nodes and edges deriving from Gene Ontology (GO) similarity, allowing the model to propagate information through the GNN encoder. To test the role of the GRN in the prediction performance, we compare the performance of the GEARS model in the testing dataset by using the original GRN and a randomized GRN with the same edges.

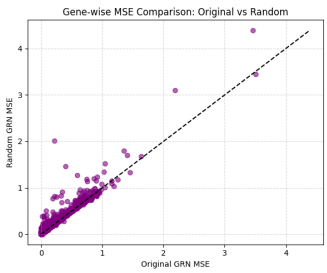
From figure 2, the MSE loss with the random GRN matrix (0.088) is higher, compared to the original GRN structure (0.079), indicating that the original GRN structure leads to better prediction performance. Figure 2b shows that the majority of comparison points stay above the identity line, demonstrating that the MSE loss increases as a random GRN structure is leveraged across most of the genes. While most of the genes exhibit slight changes in MSE loss by using a random GRN, several display a significant increase, and none of them demonstrate a marked reduction. The per gene distribution is also shown in figure 3. The overall poor performance of the GEARS model with a randomized GRN is even more noticeable when considering the W2 loss. The GEARS model with the original GRN achieves a W2 loss of value 0.026, while the loss increases to 0.054 when the random GRN structure is replaced, referring to Appendix C.

The original GRN structure shows overall higher prediction accuracy in MSE loss and W2 loss, suggesting that the GRN is crucial to preserve the shape and distributions of gene expression responses. Under GRN randomization, the MSE loss rises slightly, but the W2 loss grows rapidly, indicating a significant impact on the expression distributions. With a random GRN, the model retains a baseline performance of prediction accuracy, while replicating the biological structure is challenging. The reason is that the original GRN structure encodes the regulatory patterns and latent dependencies, enabling the model to facilitate the underlying biological topology.

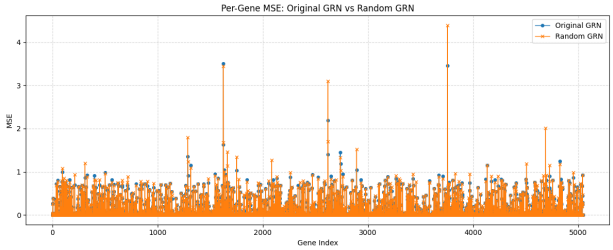
Overall, the Gene Regulatory Network (GRN) plays a crucial role in the GEARS model, as its capabilities in capturing the underlying biological



(a) Total MSE comparison between GRNs.



(b) Gene-wise MSE comparison plot (original vs random GRN).



(c) Per-gene MSE across gene indices.

Figure 2: Evaluation of the Gene Regulatory Network (GRN) structure in the GEARS model. (a) Total MSE bar plot indicating aggregate performance. (b) Gene-wise comparison plot comparing MSE between original and random GRNs. (c) Per-gene MSE comparison across gene indices for original and random GRNs.

patterns and gene interactions enable accurate prediction of the shape and distribution of gene expression responses under single-gene perturbations. The use of the random GRN structure leads to degraded performance, especially in W2, highlighting its failure in capturing biological expression patterns and regulatory dependencies. This underlines the significance of network structure with biological patterns to ensure the robust prediction results.

5.3 Augmented Neural Ordinary Differential Equations (ANODE) Capturing Non-linearity

Despite the recent developments in evaluating perturbation-induced gene expressions, it remains unclear how to effectively measure the dynamics and distribution of gene expressions influenced by the chemical or genetic perturbations. The NODE model attains a superior W2 score in predicting transcriptional responses since single-gene perturbation effects are effectively captured using a continuous dynamical system. As previously suggested by Domingo, Julia etc. (2024), non-monotonic behavior

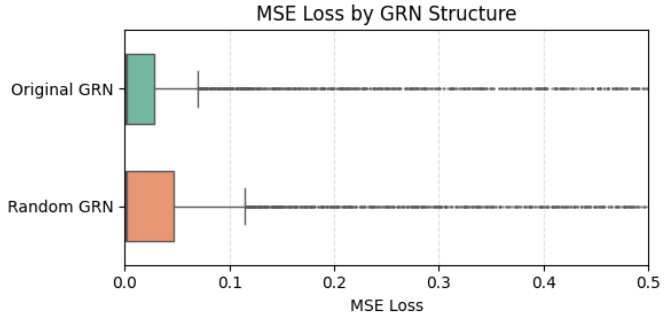


Figure 3: Scatter plot of the original GRN and the random GRN in MSE loss

is noticeable in the responses to dosage changes even when the regulator changes are restricted to a copy number. This effect can lead to gene expression shifts in predicted expressions, compared to the ground truth, since the models fail to capture the non-linearity in gene expressions under perturbations. We aim to address the issue by adding the latent space into the dynamics. Since the normal NODE model is performed in the fixed-dimensional latent space, struggling to learn non-linear responses and high-order dependencies, we present the Augmented Neural Ordinary Differential Equation (ANODE) model to capture possible disproportionately large gene expression changes caused by perturbations, and compare the results with benchmark models.

The Augmented Neural Ordinary Differential Equation (ANODE) model is trained on the perturbation embeddings, introducing additional latent dimensions to measure the gene expression shifts. Compared with the NODE model, the ANODE framework operates on an augmented matrix given by

$$\begin{bmatrix}\mathbf{x}(t) & \mathbf{a}(t)\end{bmatrix} \in \mathbb{R}^{n\times(d+k)},$$

where $x(t) \in R^{n\times d}$ represents the gene expression state and $a(t) \in R^{n\times k}$ is an auxiliary latent vector that provides additional dimensions to improve flexibility and enable high-dimensional representations. And n, d denotes the sample size and number of genes respectively. Then the cellular dynamics can be derived as

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}(t) & \mathbf{a}(t)\end{bmatrix} = \sigma\left(\begin{bmatrix}\mathbf{x}(t) & \mathbf{a}(t)\end{bmatrix} B^\top + \boldsymbol{\beta}\right) \odot M(u)A^\top - \begin{bmatrix}\mathbf{x}(t) & \mathbf{a}(t)\end{bmatrix} W^\top,$$

where $A, B, W \in \mathbb{R}^{(d+k)\times(d+k)}$ are weight matrices that update during each iteration, $\boldsymbol{\beta} \in \mathbb{R}^{d+k}$ denotes

the bias matrix, $M(u) \in \mathbb{R}^{N\times(d+k)}$ is linear activation function of perturbation embeddings with learnable parameters, and $\sigma(\cdot)$ is the RELU activation function. Utilizing the ANODE model, the limitations of the traditional NODE model can be addressed since the latent augmentation can capture the complex biological patterns of cellular responses, enabling the representation of the skewed and multimodal distributions while reducing prediction shifts in gene expression responses.

To demonstrate the effectiveness of the ANODE model, we compare the evaluated metrics on the testing dataset with three benchmark models: the GEARS model, the linear model and the PerturbODE model. As shown in figure 1c, the ANODE model attains improved performance in terms of MSE and KL divergence (MSE: 0.0461; KL: 5.549), compared to the NODE model (MSE: 0.0461; KL: 5.577). Although the MSE performance of the ANODE model is slightly below that of the linear model, it significantly outperforms the GEARS model in the MSE metric. In the KL divergence, we observe that the ANODE model outperforms all the benchmark models, indicating its ability to predict the shape of gene expression responses under single-gene perturbation. Lower KL divergence in the ANODE model also reveals its ability to preserve the transcriptional features under single-gene perturbation and prevent the shifts in expression distributions. Appendix B demonstrates how KL divergence effectively penalizes shift behaviors in gene expression prediction, along with its biological interpretation.

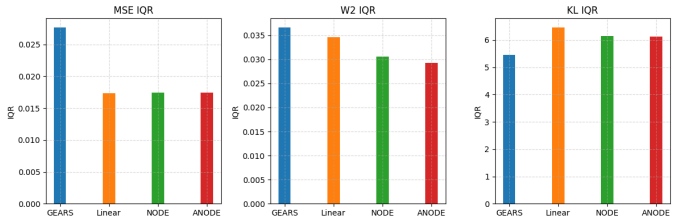


Figure 4: Interquartile range (IQR) of MSE, Wasserstein-2, and KL divergence for GEARS, Linear, NODE, and ANODE models. Lower IQR indicates more stable performance across perturbations.

To analyze the stability of different methods in MSE, W2 and KL, the interquartile range (IQR) measures the spread of a metric by calculating the difference between 25% and 75% of the values. In figure 4, the performance of IQR over all the metrics across genes is examined. The ANODE model attains the lowest IQR in W2 divergence (0.0293) and maintains a competitive MSE IQR (0.01739), slightly higher than that of the linear model (0.01735). Also, ANODE model outperforms the linear model (6.445) and NODE model (6.141) in KL IQR, with a value of 6.118. Although the GEARS model attains the lowest KL IQR among all the models, the ill performance in KL divergence undermines the reliability of the result. Overall, the results highlight the stability performance of the ANODE model in predicting gene expressions under single-gene perturbation. The ANODE model shows its strength in modeling capabilities among all the models, while minimizing prediction loss, maintaining biologically significant distributions, and sustaining stability over all the metrics.

6 Codes

Please refer to `https://github.com/wang2024467/Augmented-Neural-ODE`.

7 Conclusion

Among the three benchmark models, the linear model exhibits better performance in single-gene perturbation compared to the GEARS model and the graph-based NODE model achieves the best performance in KL divergence. Also, the structure of the graph-based NODE model is different from that of the PerturbODE by introducing the linear activation on the perturbation embeddings, which is derived from the integrated Gene Regulatory Network (GRN) through a GNN encoder. The GRN is crucial to capture the regulatory dependencies across the genes. After replacing the GRN with a randomized one in the GEARS model, the MSE loss increases and the degraded performance is noticeable in W2 loss, demonstrating its weakness in representing coherent expression responses and capturing regulatory influences.

Our findings reveal that the ANODE provides a biologically interpretable framework in predicting the gene expression responses to single-gene perturbation. The performance of ANODE on MSE loss is strong, much better than the GEARS model. The ANODE model exhibits the lowest KL divergence, indicating

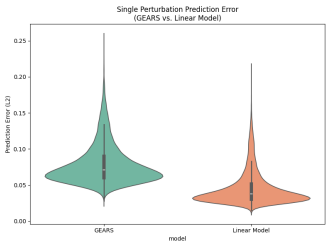
the advantages in capturing the underlying patterns of gene expression responses. From a biological perspective, it is particularly essential to capture the shape of expression distributions and avoid shifts in the prediction results. The ground truth of gene expression is a sparse matrix with integer entries, reflecting the number of transcriptional genes. In prediction results, minor variations in distributions across different genes are inconsequential, while the shift behaviors such as mode displacements can distort gene-level conclusions and significantly confound the gene expression relationships. By attaining the best performance in KL divergence and mitigating the shift behaviors, the ANODE model can predict the gene expression responses under single-gene perturbation in biologically meaningful contexts.

By adding the auxiliary latent dimensions, the ANODE model prevents regulatory influences while maintaining the structured formulation that measures biological dynamics. Compared to the other three benchmark models, ANODE model improves biological interpretability and provides a compelling approach for perturbation modeling.

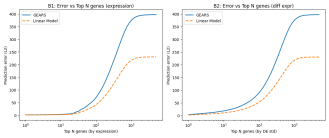
Appendix A Linear Model Exhibits Better Performance than GEARS Model in Single-gene Perturbation

Our results reveal that, on single-gene perturbation tasks, the linear model consistently outperforms GEARS and even matches the performance of PerturbODE in terms of mean squared error (MSE). This supports recent findings that linear models often capture direct biological effects more robustly than complex deep learning models in well-structured single-gene settings. However, GEARS demonstrates greater capacities to model non-additive effects, making it valuable for multi-gene interactions.

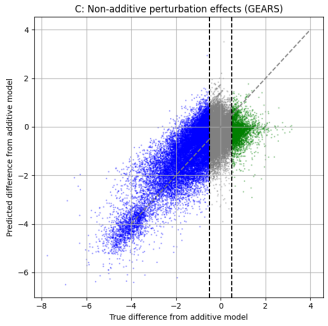
By comparing the performance between the GEARS model and the linear model on predicting transcriptional responses to single-gene perturbation, we utilize evaluation metrics to analyze the prediction results. As shown in figure 5, we observe that the linear model achieves lower MSE loss compared to the GEARS model. Also, in the cumulative prediction error, the linear model outperforms the GEARS model when N increases. From figures 5c and 5d, the linear model stays closer to real cellular responses across the perturbation strengths, indicating that the linear model performs better than the GEARS model in single-gene perturbation responses, which aligns with the results in the paper by Ahlmann-Eltze etc. (2024).



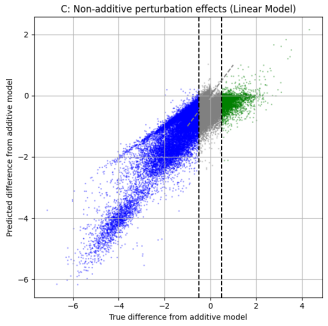
(a) Violin plot of MSE errors.



(b) Cumulative MSE over top-ranked genes.



(c) Non-additive effects (GEARS).



(d) Non-additive effects (Linear Model).

Figure 5: Comparison between GEARS and the Linear Model across evaluation axes: (a) prediction error distribution, (b) cumulative prediction error over top-ranked genes, and (câĂŞd) non-additive perturbation response fitting.

Appendix B Simple Samples Explaining The Reason for Lower W2 but Higher KL Divergence and Their Biological Significance

The results in predicting gene expressions under single-gene perturbations show that the GEARS model attains lower W2 distance but incurs a much higher KL divergence. Although both W2 distance and KL divergence measure the distribution similarities of the predicted gene expressions compared to the ground truth, they capture different aspects of distribution behaviors.

To demonstrate the distribution between W2 distance and KL divergence, we construct a simple example with gene expressions of five samples. Suppose that

Model 1 and Model 2 predict the gene expression responses, as shown in table 1 . A notable difference arises in gene 3, where model 1 shifts the expression modes, but model 2 produces a smoother but centered distribution, both different from the ground truth.

Gene	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Ground Truth Distributions					
Gene 1	0.15	0.25	0.33	0.18	0.05
Gene 2	0.05	0.10	0.65	0.10	0.05
Gene 3	0.00	0.00	1.00	0.00	0.00
Gene 4	0.20	0.20	0.20	0.20	0.20
Model 1 Predictions (shifted)					
Gene 1	0.10	0.20	0.40	0.25	0.05
Gene 2	0.05	0.05	0.70	0.15	0.05
Gene 3	0.00	0.00	0.00	1.00	0.00
Gene 4	0.20	0.20	0.20	0.20	0.20
Model 2 Predictions (centered)					
Gene 1	0.05	0.20	0.30	0.20	0.25
Gene 2	0.10	0.25	0.40	0.15	0.10
Gene 3	0.10	0.20	0.40	0.20	0.10
Gene 4	0.20	0.20	0.20	0.20	0.20

Table 1: Ground truth and predicted distributions for four example genes. Model 1 predictions are sharper but slightly shifted, leading to lower Wasserstein-2 distance but higher KL divergence in some cases.

Gene	KL (Model 1)	W2 (Model 1)	KL (Model 2)	W2 (Model 2)
Gene 1	0.035	0.235	0.200	0.681
Gene 2	0.031	0.105	0.171	0.379
Gene 3	23.026	1.000	0.916	0.800
Gene 4	0.000	0.000	0.000	0.000
Mean	**7.697**	**0.447**	**0.429**	**0.620**

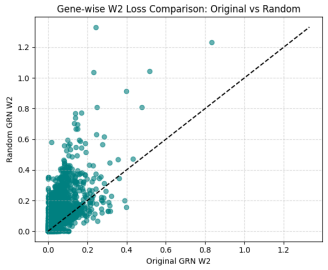
Table 2: KL divergence and Wasserstein-2 distance (W2) for each gene under both models. Model 1 shows sharper but spatially shifted predictions, while Model 2 better captures the overall shape, leading to lower KL values.

Table 2 demonstrates the results of prediction calculated mean of KL divergence and W2 distance. Although model 1 exhibits lower W2, the extremely large average KL divergence reveals the limitations of model 1 in capturing the non-linear cellular responses, especially the shifts in predicted distributions compared to the ground truth, which can be seen in Gene 3. This suggests that W2 distance quantifies the overall geometric difference between distributions while KL divergence penalizes the shift behaviors or mismatching.

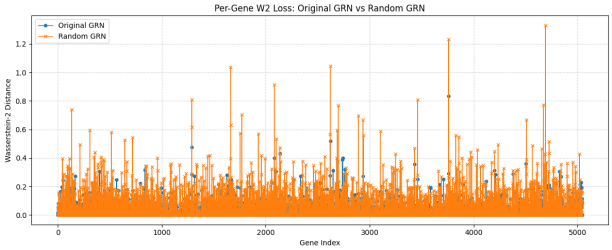
Appendix C Graphs of Comparison Between Original GRN and Random GRN

In the GEARS model, if the original GRN is replaced by a random GRN, the ill performance is noticeable in W2 loss, as shown in figure 6.

(a) Total W2 comparison between GRNs.



(b) Gene-wise W2 scatter plot (original vs random GRN).



(c) Per-gene W2 across gene indices.

Figure 6: Evaluation of the Gene Regulatory Network (GRN) structure in the GEARS model. (a) Total W2 bar plot indicating aggregate performance. (b) Gene-wise scatter plot comparing W2 between original and random GRNs. (c) Per-gene W2 comparison across gene indices for original and random GRNs.

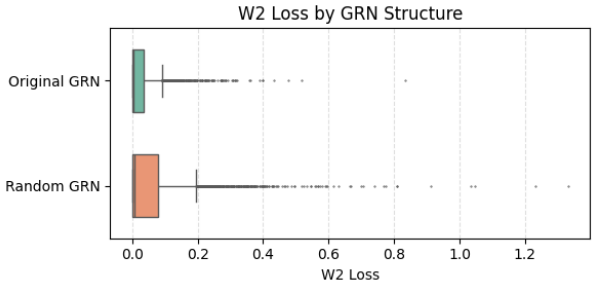


Figure 7: Scatter plot of the original GRN and the random GRN in W2 loss

Appendix D Methodologies of Neural Ordinary Differential Equations

Neural Ordinary Differential Equations (Neural ODE) combines numerical methods with deep neural networks in handling dynamic systems. It employs a backpropagation-based method to determine the derivative of the loss function for given coefficients, and afterwards, the optimal coefficients are calculated by root-finding methods such as Newton's method, which is aimed to minimize the loss for a given ODE system. Neural ODE addresses the initial value problem with a collection of coefficients θ:

$$\begin{cases} \frac{du(t)}{dt} = f(u, t; \theta), \\ u(0) = u_0, \end{cases}$$

where $u(t)$ is a function depending on time t and u_0 is the initial value. In this problem, the objective is to determine the optimal θ that accurately represents the model. Thus, we define the loss function that compares the output of $u(t)$ with given θ to the real value.

$$L(u, \theta) = \int_0^T g(u; \theta) dt$$

The primary purpose of Neural ODE is to calculate the gradient of loss function $dL/d\theta$ in terms of the coefficients θ. To achieve this, the adjoint state sensitivity is leveraged to track the derivative of loss in the initial state. Consider the constrained optimization problem and the Lagrangian equation.

$$\begin{cases} \min_\theta L(u; \theta) \\ s.t.\ \frac{du}{dt} = f(u, t; \theta) \end{cases} \tag{1}$$

$$\mathcal{L}[u, \lambda; \theta] = L(u; \theta) + \int_0^T \lambda(t)^T (f - \frac{du}{dt}) dt.$$

Given the ODE system $\frac{du}{dt} = f$, the Lagrangian $\mathcal{L}$ is equivalent to L. The adjoint is defined as

$$a(t) = \frac{\partial L}{\partial u}$$

Then it follows that

$$\frac{dL}{dt}(u; \theta) = \frac{\partial L}{\partial u} \cdot \frac{du}{dt} = a(t)^T f$$

Therefore,

$$\begin{aligned} \frac{da(t)}{dt} = \frac{d}{dt}\left(\frac{\partial L}{\partial \theta}\right) &= \frac{\partial}{\partial u}\left(\frac{dL}{dt}\right) \\ &= \frac{\partial}{\partial u}\left(a(t)^T f(u, t; \theta)\right) \\ &= a(t)^T \frac{\partial f}{\partial u} \end{aligned}$$

And similarly,

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial L}{\partial \theta}\right) &= \frac{\partial}{\partial \theta}\left(\frac{dL}{dt}\right) \\ &= \frac{\partial}{\partial \theta}\left(a(t)^T f(u, t; \theta)\right) \\ &= a(t)^T \frac{\partial f(u, t; \theta)}{\partial \theta} \end{aligned}$$

Integrating by parts,

$$\left.\frac{dL}{d\theta}\right|_{t_0} = \left.\frac{dL}{d\theta}\right|_T + \int_{t_0}^T a(t)^T \frac{\partial f}{\partial \theta} dt$$

Thus, we can evaluate the following ODE systems that

$$\begin{cases} \frac{du}{dt} = f(u, t; \theta), \\ \frac{da}{dt} = -a(t)^T \frac{\partial f(u,t;\theta)}{\partial \theta}, \\ \frac{d}{dt}\left(\frac{\partial L}{\partial \theta}\right) = -a(t)^T \frac{\partial f(u,t;\theta)}{\partial \theta}, \\ \left[u, a, \frac{dL}{d\theta}\right]\Big|_{t=t_1} = [u(t_1), a(t_1), 0] \end{cases}$$

References

Lotfollahi, Mohammad, F. Alexander Wolf, and Fabian J. Theis. *scGen predicts single-cell perturbation responses.* Nature methods 16.8 (2019): 715-721.

Hetzel, Leon, et al. *Predicting cellular responses to novel drug perturbations at a single-cell resolution.* Advances in Neural Information Processing Systems 35 (2022): 26711-26722.

Duan, Lintao, et al. *Analytical method for time-varying meshing stiffness and dynamic responses of modified spur gears considering pitch deviation and geometric eccentricity.* Mechanical Systems and Signal Processing 218 (2024): 111590.

Roohani, Yusuf, Kexin Huang, and Jure Leskovec. *Predicting transcriptional outcomes of novel multigene perturbations with GEARS.* Nature Biotechnology 42.6 (2024): 927-935.

Jovic, Dragomirka, et al. *SingleâĂŘcell RNA sequencing technologies and applications: A brief overview.* Clinical and translational medicine 12.3 (2022): e694.

Haque, Ashraful, et al. *A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.* Genome medicine 9 (2017): 1-12.

Ahlmann-Eltze, Constantin, Wolfgang Huber, and Simon Anders. *Deep learning-based predictions of gene perturbation effects do not yet outperform simple linear methods.* BioRxiv (2024): 2024-09.

Ferrell Jr, James E. *Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability.* Current opinion in cell biology 14.2 (2002): 140-148.

Domingo, Julia, et al. *Non-linear transcriptional responses to gradual modulation of transcription factor dosage.* BioRxiv (2024).

Chen, Ricky TQ, et al. *Neural ordinary differential equations.* Advances in neural information processing systems 31 (2018).

Lin, Zaikang, et al. *Interpretable Neural ODEs for Gene Regulatory Network Discovery under Perturbations.* arXiv preprint arXiv:2501.02409 (2025).

Dixit, Atray, et al. *Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens.* cell 167.7 (2016): 1853-1866.