

Review of Single-Cell Perturbation Response Prediction Methods Based on Generative Models

Yi Deng^{1*}

¹ School of Information, Yunnan Normal University, Kunming, China

Email Address

2323410009@ynnu.edu.cn (Yi Deng)

*Correspondence: 2323410009@ynnu.edu.cn

Received: 27 February 2026; **Accepted:** 25 March 2026; **Published:** 13 April 2026

Abstract:

Recent advances in high-throughput single-cell perturbation technologies have provided a critical data foundation for systematically studying gene function and cellular response patterns. However, high experimental costs and the vast combinatorial space have made computational prediction methods an essential complement. Focusing on single-cell perturbation response prediction, this paper presents a systematic review of existing deep learning methods from the perspective of modeling paradigms. Based on how perturbation effects are characterized, existing methods are categorized into two classes: latent space additive effect methods and direct perturbation distribution generation methods. Representative works within each category are examined, along with their respective strengths and limitations. A comparative analysis is then conducted across key dimensions including modeling assumptions, perturbation condition encoding, and incorporation of biological priors. Finally, core challenges facing the field are discussed-including out-of-distribution generalization, interpretability, and standardization of evaluation criteria-followed by an outlook on future research directions.

Keywords:

Single-cell Transcriptomics, Perturbation Response Prediction, Variational Autoencoder, Out-of-distribution Generalization

1. Introduction

Gene expression regulation is a fundamental mechanism governing cellular function. Under perturbations such as drug treatment, gene editing, or infection, intracellular transcriptional regulatory networks undergo dynamic remodeling, leading to substantial changes in cellular states and phenotypes[1]. Elucidating gene expression response patterns under perturbation conditions is essential for understanding gene function, uncovering disease mechanisms, and advancing precision medicine[2]. Recent advances in high-throughput single-cell perturbation technologies, including Perturb-seq[1], CROP-seq[3] and Sci-Plex[4], have enabled parallel profiling of gene knockouts, drug treatments, and diverse dosage combinations at single-cell

resolution[5][6]. Despite these advances, single-cell perturbation experiments remain subject to significant practical limitations. First, high experimental costs and prolonged time cycles hinder comprehensive exploration of large-scale perturbation combinatorial spaces[4]. This challenge is particularly acute under multi-gene knockout or multi-dosage conditions, where the number of combinations grows exponentially, rendering exhaustive experimental enumeration virtually infeasible[7]. Second, single-cell transcriptomic data inherently exhibits high dimensionality, sparsity, and pronounced batch effects, imposing stringent demands on downstream modeling[8]. In this context, researchers have increasingly turned to generative deep learning models to learn perturbation response patterns from observed data and predict gene expression changes under unobserved conditions. However, existing methods generally exhibit insufficient out-of-distribution (OOD) generalization capability when confronted with unseen cell types, novel perturbations, or new perturbation combinations[7].

Current single-cell perturbation prediction models primarily address three categories of tasks: genetic perturbation prediction, which forecasts transcriptomic changes following knockout or overexpression of specific genes using technologies such as CRISPR, supporting gene function analysis and regulatory network reconstruction; cellular perturbation prediction, which models transcriptional responses of different cell types to external stimuli, facilitating research in immunology and developmental biology; and drug perturbation prediction, which characterizes the effects of small-molecule compounds and their dosage variations on transcriptional states, with applications in drug screening and dosage optimization. Although all three task categories share common challenges including high dimensionality, sparsity, and batch effects, they differ in perturbation representation and the availability of prior knowledge: genetic perturbations can leverage structured priors such as Gene Ontology and protein–protein interaction networks; cellular perturbations are described in more general terms and rely more heavily on data-driven approaches; and drug perturbations can incorporate molecular structure priors and involve dose-dependent continuous modeling. This paper categorizes existing methods into two classes based on how perturbation effects are characterized: latent space additive effect methods and direct perturbation distribution generation methods. We provide a systematic review of the core ideas, representative works, and respective strengths and limitations of each class, followed by a comparative analysis across key technical dimensions including modeling assumptions, perturbation condition encoding, and incorporation of biological priors. The remainder of the paper is organized as follows. Chapter 2 systematically reviews existing methods according to the two modeling paradigms. Chapter 3 presents a comparative discussion of the two method classes from multiple technical perspectives. Chapter 4 summarizes current challenges and outlines future directions. Chapter 5 concludes the paper.

2. Method Classification and Technical Approaches

Based on how perturbation effects are characterized, this paper categorizes existing representative methods into two classes. The first class, termed latent space additive effect methods, characterizes perturbation effects through vector arithmetic or approximately linear transformations in latent space. The second class, termed direct perturbation distribution generation methods, directly learns the mapping from perturbation conditions to gene expression distributions via conditional generation or

distribution alignment. The following sections provide a systematic review of each class.

2.1. Latent Space Additive Effect Methods

Latent space additive effect methods represent the earliest and most widely adopted modeling paradigm in single-cell perturbation prediction. These methods typically employ generative models such as variational autoencoders (VAEs) to map high-dimensional gene expression data into a low-dimensional latent space, assuming that perturbation effects exhibit an additive or approximately linear structure. Accordingly, the post-perturbation cellular state can be represented as a vector combination of the basal state and the perturbation effect.

Representative works such as scGen[2], scPRAM[9] and scVIDR[10] all model state changes before and after perturbation based on vector differences or additive shifts in latent space. Specifically, scGen achieves cross-cell-type prediction by estimating the mean vector difference between latent representations under perturbed and control conditions. scPRAM combines optimal transport with attention mechanisms to align latent spaces and compute individualized perturbation vectors. scVIDR introduces linear regression-based additive parameter learning for dose-dependent scenarios, characterizing expression shifts across perturbation conditions. These methods perform well in single-perturbation or weakly nonlinear scenarios but struggle to capture nonlinear interactions within gene regulatory networks under complex perturbation conditions.

Some studies enhance the expressive power of the additive framework by incorporating external structural priors. GEARS[11] integrates gene co-expression networks with Gene Ontology pathway information, employing graph neural networks to learn gene and perturbation embeddings and predicting perturbation responses through composable embeddings, achieving notable progress in combinatorial perturbation prediction. Another direction introduces disentangled representation learning to explicitly separate basal cellular states from perturbation effects in latent space, thereby alleviating the entanglement of multiple biological factors. The Compositional Perturbation Autoencoder (CPA)[12] achieves this by partitioning the latent space into cell-intrinsic state and perturbation-related subspaces. chemCPA[13] further incorporates compound structural priors for prediction of unseen drugs, while MultiCPA[14] extends the framework to multi-modal perturbation scenarios. scDisInFact[15] decomposes expression into three components: shared biological factors, condition-specific factors, and batch effects. CINEMA-OT[16] combines causal inference with optimal transport to achieve causal disentanglement.

Overall, latent space additive effect methods are structurally concise and training-stable, offering good interpretability of perturbation effects. However, they are generally constrained by linear modeling assumptions and struggle to capture complex nonlinear interactions. Although disentanglement mechanisms can alleviate factor entanglement, their effectiveness depends on sufficient training data coverage, and out-of-distribution generalization remains limited.

2.2. Direct Perturbation Distribution Generation Methods

Unlike latent space additive effect methods, direct perturbation distribution generation methods do not rely on explicit additive decomposition of basal state and

perturbation effect. Instead, they take perturbation conditions as conditional inputs to the generative process, directly learning the mapping from perturbation conditions to gene expression distributions via conditional generation or distribution alignment.

Early works such as Dr.VAE[17] achieve generative prediction of cellular drug responses by jointly modeling gene expression data before and after drug treatment. Building upon this, trVAE[7] introduces a conditional variational autoencoder framework that injects perturbation conditions as explicit conditional variables into the generative process, directly generating expression distributions under specific conditions and supporting multi-condition combinatorial modeling. ScPreGAN[18] combines autoencoders with generative adversarial networks, extracting common latent features from unperturbed and perturbed data through a shared encoder and employing adversarial training to constrain the generated distribution for more precise fitting under observed conditions.

Several studies incorporate external priors or multi-modal information to enhance conditional generation. PRnet[19], targeting drug perturbation prediction, introduces compound molecular structure priors by encoding SMILES representations into perturbation embeddings through molecular fingerprint extraction, improving generalization to novel perturbation conditions. CellDrift[20], designed for time-series perturbation data, jointly models perturbation conditions and temporal information to identify gene-level dynamic responses and reveal key changes during cellular state evolution. Disentanglement designs have also emerged within this method class. Biolord[21] encodes known and unknown attributes into independent subspaces, achieving multi-attribute decoupled modeling and cross-perturbation generalization. scCausalVI[22], grounded in structural causal models, decomposes expression into background state and treatment effect latent variables, supporting counterfactual inference.

Overall, by relaxing structural assumptions on latent space, these methods enhance the capacity to model complex expression distributions and achieve good generation quality under observed conditions. However, their predictive performance is highly dependent on training data coverage, with accuracy declining noticeably when conditions deviate from the training distribution. Furthermore, the lack of explicit constraints on the biological semantics of latent variables limits interpretability and renders predictions prone to insufficient biological consistency under out-of-distribution scenarios.

3. Comparison and Discussion of Key Technical Issues

The preceding chapters have systematically reviewed existing methods from the perspective of two modeling paradigms. This chapter presents a comparative analysis across three dimensions: modeling assumptions, perturbation condition encoding, and incorporation of biological priors.

3.1. Comparison of Modeling Assumptions

The core distinction between the two method classes lies in their prior assumptions regarding the structure of perturbation effects, which directly determines their respective applicability boundaries. The linear additivity assumption of additive effect methods provides a reasonable approximation in single-gene perturbation or mild-effect scenarios. However, under multi-gene combinatorial perturbations, nonlinear gene interactions cause vector superposition to deviate from true expression changes.

Although conditional generation methods do not presuppose an additive structure and offer greater fitting flexibility, this flexibility is contingent on sufficient training data coverage—generation quality is satisfactory under observed conditions, but extrapolation capability degrades notably when conditions deviate from the training distribution. The two method classes thus exhibit complementary characteristics: additive methods provide a foundation for combinatorial extrapolation through structured assumptions but are constrained by the precision of linear approximation, whereas conditional generation methods possess stronger fitting capability within observed distributions but lack the structural inductive bias necessary to support extrapolation.

3.2. Comparison of Perturbation Condition Encoding Strategies

The representation of perturbation conditions directly affects model generalization performance. Existing methods primarily adopt three encoding strategies. The first is one-hot encoding or learnable embeddings, as employed in CPA[12] and trVAE[7], which map each perturbation to an independent vector. This approach is straightforward to implement but, since encodings of different perturbations are mutually independent, cannot support transfer to unseen perturbations. The second is structured embeddings based on graph neural networks, represented by GEARS[11], which leverages GO graphs and co-expression networks to ensure that functionally related genes obtain similar representations, providing a structured foundation for prediction of unseen genes or gene combinations. The third is prior encoding based on molecular structures, as employed in chemCPA[13] and PRnet[19], which encode compound SMILES representations into perturbation embeddings through molecular fingerprint extraction, inherently supporting generalization to unseen drugs but applicable only to drug perturbation scenarios. Overall, one-hot encoding is stable under closed-set settings but lacks generalization capability, whereas the latter two strategies enhance cross-perturbation generalization through external knowledge, though their effectiveness depends on the quality and completeness of the available priors.

3.3. Approaches and Effects of Incorporating Biological Priors

Incorporating biological priors is an important avenue for enhancing model generalization, and the priors employed in existing research primarily encompass three categories. Gene Ontology (GO) functional relationship graphs are the most widely applied prior source in genetic perturbation prediction. Their hierarchical structure captures functional dependencies among genes and enhances prediction for unseen genes and gene combinations, though GO annotations suffer from incompleteness and considerable variation in depth across genes. Protein–protein interaction (PPI) networks provide information on physical interactions among gene products and hold potential value for characterizing cascade effects following gene knockout, but their direct application in single-cell perturbation prediction remains limited. Compound molecular structure priors play an important role in drug perturbation prediction, offering conveniently accessible and precisely described structural information, but they encode only the chemical properties of drugs and do not reflect actual drug–target binding affinity. In summary, prior knowledge provides models with structured constraints extending beyond training data, helping to alleviate OOD generalization challenges, though its incompleteness, noise, and compatibility with specific tasks affect practical effectiveness.

4. Challenges and Outlook

Although existing single-cell perturbation response prediction methods have achieved significant progress in algorithm design and application, several key challenges remain when confronting more complex biological contexts and higher generalization demands. This chapter analyzes current core issues and outlines future research directions.

4.1. Current Core Challenges

Exponential growth of combinatorial perturbation space and out-of-distribution generalization bottleneck. In genetic perturbation prediction, the number of multi-gene combinatorial perturbations grows exponentially with the number of genes, yet training data can cover only a small fraction of the combinatorial space. Without explicit functional similarity constraints, existing methods struggle to transfer learned perturbation effects to unseen genes or gene combinations, resulting in significant performance degradation. Achieving reliable extrapolation to unobserved perturbation conditions from limited training data remains a core bottleneck in the field.

Batch effects, latent representation entanglement, and insufficient interpretability. Single-cell transcriptomic data is simultaneously influenced by multiple factors, including cell type, developmental state, technical noise, and external perturbations. Most existing models fail to adequately distinguish these factors in latent space, leading to mixed encoding of cell-intrinsic attributes and perturbation effects. Although disentanglement mechanisms can alleviate this issue, their effectiveness depends on sufficient training data coverage, and stability under out-of-distribution scenarios remains limited. Relatedly, current generative processes generally lack explicit constraints on the biological semantics of latent variables, making it difficult to distinguish inherent cellular differences from perturbation-induced changes and thereby limiting the biological interpretive value of predictions. Moreover, a systematic validation framework linking model predictions to known regulatory pathways has yet to be established.

4.2. Future Research Directions

Deep integration of biological priors and multi-modal data fusion. Current utilization of prior knowledge such as GO graphs and protein–protein interaction networks remains relatively preliminary. Future research may explore incorporating richer biological knowledge—including gene regulatory network topology and chromatin accessibility information—into perturbation modeling, enabling models to understand perturbation effect propagation at the structural level and enhancing generalization to unseen perturbations. Meanwhile, with advances in multi-omics technologies, joint modeling integrating multi-modal data such as ATAC-seq, proteomics, and spatial transcriptomics promises a more comprehensive characterization of cellular states than single transcriptomics alone, offering a new data foundation for multi-level analysis of perturbation effects.

Large-scale pretraining and foundation models. Single-cell foundation models such as scGPT and GeneFormer have demonstrated knowledge transfer capability across various downstream tasks through pretraining on large-scale transcriptomic data. Introducing the pretrain-finetune paradigm into perturbation prediction holds promise for alleviating scarce labeled data and out-of-distribution generalization

challenges. However, designing finetuning strategies well-suited to perturbation prediction tasks remains an open question.

Causal inference and causal representation learning. Most existing methods rely on associative prediction and struggle to reveal the causal mechanisms underlying perturbation effects. Combining structural causal models, do-calculus, and other causal inference frameworks with representation learning has the potential to advance models from predicting post-perturbation outcomes toward understanding why perturbations produce such effects, providing more valuable computational support for mechanistic research.

5. Conclusions

This paper presents a systematic review of deep learning methods for single-cell perturbation response prediction from the perspective of modeling paradigms. Based on how perturbation effects are characterized, existing methods are categorized into two classes: latent space additive effect methods and direct perturbation distribution generation methods, with a comparative analysis conducted across dimensions including modeling assumptions, perturbation condition encoding, incorporation of biological priors, and disentanglement mechanisms. The two classes of methods each possess distinct advantages: additive effect methods are structurally concise with good interpretability, but the linear assumption limits their capacity to capture complex nonlinear interactions; conditional generation methods offer stronger fitting capability, but exhibit high dependence on training distributions and limited out-of-distribution generalization. Building on this, the paper provides a comparative analysis of the two classes of methods across dimensions including modeling assumptions, perturbation condition encoding, and incorporation of biological priors. Currently, how to learn representations that are biologically consistent, capable of capturing nonlinear perturbation mechanisms, and equipped with stable out-of-distribution generalization capability based on limited perturbation observation data remains the core challenge in this field. With the deep integration of biological priors, multi-modal data fusion, the introduction of pretrained foundation models, and causal inference methods, single-cell perturbation response prediction is expected to achieve further breakthroughs in generalization capability and application value.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Dixit, A.; Parnas, O.; Li, B, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, 2016, 167(7), 1853-1866. e17.
- [2] Lotfollahi, M.; Wolf, F.A.; Theis, F.J. scGen predicts single-cell perturbation responses. *Nature methods*, 2019, 16(8), 715-721.

- [3] Bielczyk-Maczynska, E.; Sharma, D.; Blencowe, M, et al. CROP-Seq: a single-cell CRISPRi platform for characterizing candidate genes relevant to metabolic disorders in human adipocytes. *BioRxiv*, 2022, 2022.06. 27.497796.
- [4] Srivatsan, S.R.; McFaline-Figueroa, J.L.; Ramani, V, et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science*, 2020, 367(6473), 45-51.
- [5] Replogle, J.M.; Norman, T.M.; Xu, A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature biotechnology*, 2020, 38(8), 954-961.
- [6] Norman, T.M.; Horlbeck, M.A.; Replogle, J.M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*, 2019, 365(6455), 786-793.
- [7] Lotfollahi, M.; Naghipourfar, M.; Theis, F.J, et al. Conditional out-of-distribution generation for unpaired data using transfer VAE. *Bioinformatics*, 2020, 36(Supplement_2): i610-i617.
- [8] Lopez R, Regier J, Cole M B, et al. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 2018, 15(12), 1053-1058.
- [9] Jiang, Q.; Chen, S.; Chen, X, et al. scPRAM accurately predicts single-cell gene expression perturbation response based on attention mechanism. *Bioinformatics*, 2024, 40(5), btae265.
- [10] Kana, O.; Nault, R.; Filipovic, D, et al. Generative modeling of single-cell gene expression for dose-dependent chemical perturbations. *Patterns*, 2023, 4(8).
- [11] Roohani, Y, Huang, K, Leskovec, J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nature Biotechnology*, 2024, 42(6), 927-935.
- [12] Lotfollahi, M.; Klimovskaia, Susmelj, A.; De, Donno, C, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular systems biology*, 2023, 19(6), e11517.
- [13] Hetzel, L.; Boehm, S.; Kilbertus, N, et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Advances in Neural Information Processing Systems*, 2022, 35, 26711-26722.
- [14] Inecik, K.; Uhlmann, A.; Lotfollahi, M, et al. Multicpa: Multimodal compositional perturbation autoencoder. *BioRxiv*, 2022, 2022.07. 08.499049.
- [15] Zhang, Z.; Zhao, X.; Bindra, M, et al. scDisInFact: disentangled learning for integration and prediction of multi-batch multi-condition single-cell RNA-sequencing data. *Nature Communications*, 2024, 15(1), 912.
- [16] Dong, M.; Wang, B.; Wei, J, et al. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nature Methods*, 2023, 20(11), 1769-1779.
- [17] Rampášek, L.; Hidru, D.; Smirnov, P, et al. Dr. VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*, 2019, 35(19), 3743-3751.

- [18] Wei, X.; Dong, J.; Wang, F. scPreGAN, a deep generative model for predicting the response of single-cell expression to perturbation. *Bioinformatics*, 2022, 38(13), 3377-3384.
- [19] Qi, X.; Zhao, L.; Tian, C, et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nature Communications*, 2024, 15(1), 9256.
- [20] Jin, K.; Schnell, D.; Li, G, et al. CellDrift: inferring perturbation responses in temporally sampled single-cell data. *Briefings in Bioinformatics*, 2022, 23(5), bbac324.
- [21] Piran, Z.; Cohen, N.; Hoshen, Y, et al. Disentanglement of single-cell data with biolord. *Nature Biotechnology*, 2024, 42(11), 1678-1683.
- [22] An, S.; Cho, J.W.; Cao, K, et al. scCausalVI disentangles single-cell perturbation responses with causality-aware generative model. *BioRxiv*, 2025, 2025.02.02.636136.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)