From transcriptome reversal to network reprogramming: A new paradigm for drug screening driven by deep learning

Yijie Yang,[1,6] Hong Zhang,[2,6] Yixue Li,[1,3,4,5,*] Xuelan Gu,[2,*] and Tao Huang[1,*]

[1]Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2]Unilever R & D Shanghai, Shanghai 200335, China
[3]Guangzhou National Laboratory, Guangzhou 510000, China
[4]Guangdong-Hong Kong-Macau Joint Laboratory for Cell Fate Regulation and Diseases, Guangzhou Laboratory, Guangzhou Medical University, Guangzhou 510000, China
[5]Key Laboratory of Systems Health Science of Zhejiang Province, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[6]These authors contributed equally
*Correspondence: li_yixue@gzlab.ac.cn (Y. L.); xuelan.gu@unilever.com (X. G.); huangtao@sibs.ac.cn (T. H.)

Received: March 31, 2026; Accepted: May 11, 2026; Published Online: May 18, 2026; https://doi.org/10.59717/j.xinn-life.2026.100220

Citation: Yang Y., Zhang H., Li Y., et al. (2026). From transcriptome reversal to network reprogramming: A new paradigm for drug screening driven by deep learning. *The Innovation Life* **4**:100220.

The traditional drug and active substance screening mainly relies on target screening and phenotypic screening. Target screening requires a clear target structure and is prone to off-target effects, while phenotypic screening is difficult to elucidate mechanisms. Complex diseases involve multiple cell types and coordinated regulation across pathways, making traditional strategies often ineffective. Understanding diseases at the system level and identifying interventions has become an important research direction. With the development of high-throughput sequencing, gene expression characteristics of disease states can be obtained, enabling transcriptomics-based characterization. A recent study published in Cell proposed a deep learning drug screening framework centered on transcriptomics, which predicts the transcriptional changes induced by compounds and screens potential therapeutic molecules based on the core criterion of reversing the disease expression profile.[1]

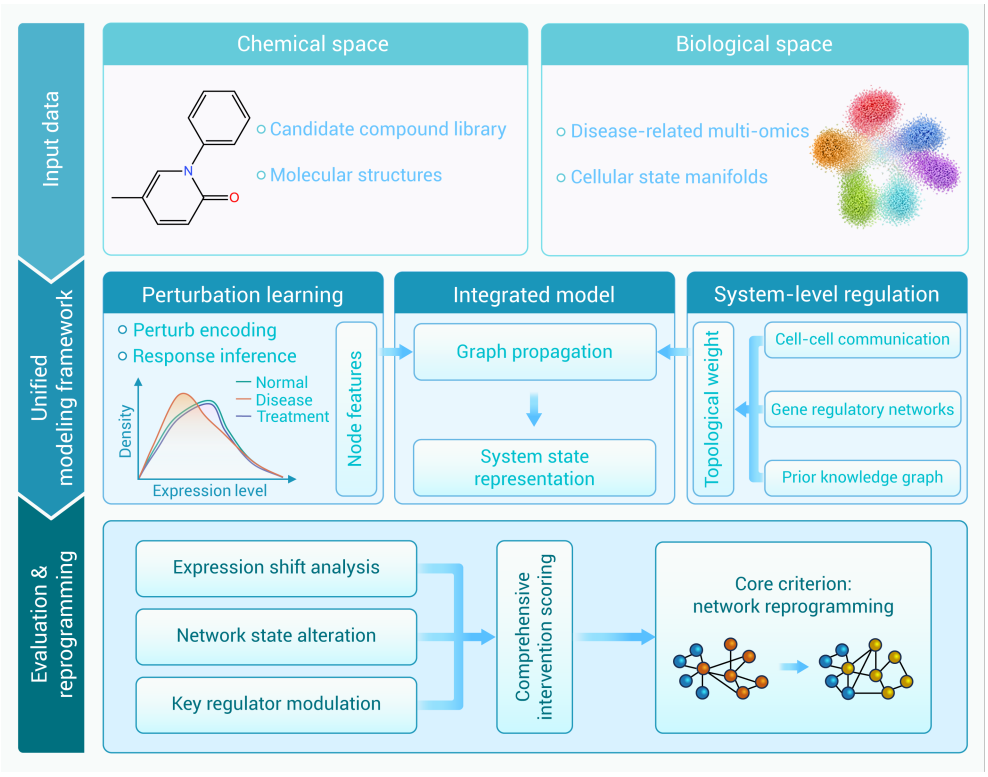

Figure 1. From expression reversal to network reprogramming in drug screening The framework integrates perturbation learning, network-aware modeling, and interpretable mechanisms. Given chemical structures and disease-related multi-omics data, a transformer-based perturbation learning model first predicts compound-induced transcriptional responses. These predicted expression profiles are further incorporated into network-aware modeling, where gene regulatory networks and cell–cell communication are explicitly considered to capture system-level regulatory effects. Interpretable modeling constrained by biological knowledge enables the identification of pathways and regulatory routes underlying the predicted responses. Finally, candidate compounds are evaluated through multiple screening tasks, including expression shift, network state alteration, and key regulator modulation, with network reprogramming as the core screening criterion.

A DEEP LEARNING FRAMEWORK FOR EXPRESSION-BASED SCREENING

The core of this study is to construct a prediction model based on the deduction of gene expression profiles from chemical structures. The model learned the mapping relationship between chemical structures and transcriptional expression changes through large-scale cell perturbation transcriptome data, thereby achieving generalized prediction of the effects of unknown compounds. The researchers introduced the "transcriptional phenotype reversal" strategy, comparing the gene expression patterns under disease conditions with the expression changes induced by compounds, and screening compounds that can reverse abnormal expression and tend towards a normal state, which are regarded as having therapeutic potential. This model integrates single-cell and bulk RNA sequencing data in idiopathic pulmonary fibrosis to construct disease-related expression features and screen out compounds that can reduce the expression of core fibroblast markers; in hepatocellular carcinoma, it is used to identify potential anti-tumor molecules.

This research is innovative in multiple aspects. Unlike traditional methods

that rely on known targets and limited phenotypic data, it uses transcriptional feature changes as the basis to reverse-screen compounds that can alleviate disease states. By introducing deep learning models, an activity prediction framework for untested compounds was established, expanding the space for early drug discovery and shortening the screening cycle. Moreover, integrating single-cell omics data can depict disease states at the cellular type level, making the screening more in line with the real biological environment.

LIMITATIONS OF EXPRESSION REVERSAL IN COMPLEX DISEASES

However, this research also has certain limitations. Complex physiological processes, such as fibrosis and aging, are not driven by the transcriptional dysregulation of a single cell type, but by the collapse of the complex multicell homeostasis in the tissue microenvironment. The current research mainly focuses on the expression changes within single cells, while the consideration of cell-to-cell communication is still relatively limited. Different cell types build a highly dynamic communication network through ligand-receptor interactions, and these cross-cell signals play a key role in the development of the disease. Limiting the screening process to the reversal of single-cell expression cannot fully capture the systemic characteristics of the disease and is prone to neglecting key bypass regulatory links. The transcriptional changes related to diseases are not all harmful; some of these changes may represent compensatory responses of the organism, and simply reversing the strategy might disrupt these responses. This screening strategy also has limited explanatory power for biological mechanisms. It cannot clearly indicate through which signaling pathways a certain compound exerts its effects, nor can it systematically assess its potential side effects or off-target risks. These limitations motivate a framework extending reversal to reprogramming (Figure 1).

TOWARD NETWORK-AWARE MODELING OF DISEASE PERTURBATIONS

In this situation, the current core issue is no longer merely predicting changes in gene expression, but rather how to incorporate molecular interactions linking compounds to transcriptional responses, such as protein-mediated processes, together with the regulatory network structure and the interactions between cells into the model. A series of related studies in recent years have provided an important foundation for this transformation. In the perturbation modeling aspect, models such as pertTF indicate that the architecture based on Transformer can learn the dependency between gene perturbations and transcriptional responses, thereby achieving generalized predictive capabilities across genes and conditions.[2] In the cell state representation aspect, CellNavi constructs an effective cell state representation by introducing gene relationships into the graph structure and embedding it into the low-dimensional manifold space of cell states, enabling the reverse and precise prediction of the key genes driving this cell state transition based on the given initial and post-perturbation cell states.[3]

Based on this, further introducing gene regulatory networks to explicitly model the real transcriptional regulatory logic is crucial for achieving mechanistic-level understanding. Methods such as GRNFormer represent the hierarchical relationships between genes through a graph Transformer structure, dynamically calculating the information transfer weights between genes and regulatory effects, enabling the model to upgrade from simple expression correlation prediction to modeling of structural-aware regulatory relationships, thus getting closer to the causal regulatory mechanisms in real biological systems.[4] The shift from modeling gene expression levels to modeling overall regulatory and intercellular communication structures has emerged as a general trend in modern biological research, reflecting a move toward system-level and structure-aware modeling paradigms.

After completing the modeling of the regulatory structure, how to provide mechanistic-level explanations for the model results becomes a further problem to be solved, and introducing interpretable artificial intelligence models becomes an important improvement direction. BFReg-NN embeds the verified hierarchical relationships between genes, proteins, and pathways into the neural layer for explicit modeling, and the network connection mode can be constrained by the known biological knowledge graph, thereby significantly enhancing the model's interpretability and enabling the prediction results to correspond to specific regulatory paths and biological processes.[5]

FROM EXPRESSION REVERSAL TO NETWORK REPROGRAMMING

Based on the above research progress, the next-generation transcriptome-driven drug screening framework is expected to achieve breakthroughs at three levels: (1) expression prediction capabilities based on perturbation learning, used to simulate the transcriptional responses of candidate compounds; (2) modeling of regulatory networks and cell communication networks based on graph structure, used to depict the information transmission regulatory mechanisms in multi-cellular systems; (3) interpretable models constrained by prior biological knowledge graphs, used to reveal the complete regulatory paths from molecular effects to cell state changes. In this framework, cell states are no longer simply represented by gene expression vectors, but are further described through their positions and interaction relationships in the gene regulatory network and cell communication network. This network-level representation enables the model to capture the compensatory pathways and buffering effects maintained by system robustness, thereby facilitating the identification of the key regulatory pathways that are easily overlooked in traditional expression reversal strategies, and providing a possibility for subsequent reprogramming of the networks targeting these pathways.

In conclusion, the transcriptome-driven drug screening method based on deep learning provides a feasible path for large-scale virtual screening of active substances and demonstrates important application prospects in complex disease research. However, to truly achieve effective intervention in complex diseases such as fibrosis and aging, it is still necessary to further develop from single expression reversal to a modeling framework for network reprogramming. With the integration of multi-omics data, graph learning methods, and interpretable artificial intelligence, the construction of a unified framework that integrates gene expression, regulatory networks, and intercellular communication can not only promote the development of precision medicine, but also reflect the conservative principles followed by biological systems in maintaining homeostasis and responding to disturbances. However, its applicability in non-model organisms with limited annotations still needs further exploration. Transfer learning may fill this gap.

REFERENCES

1. Xing J., Tan M., Leshchiner D., et al. (2026). Deep-learning-based de novo discovery and design of therapeutics that reverse disease-associated transcriptional phenotypes. *Cell* **189**:2556–2572.e19. DOI:10.1016/j.cell.2026.02.016
2. Su Y., Liu D., Menon V., et al. (2026). pertTF: Context-aware AI modeling for genome-scale and cross-system perturbation prediction. *bioRxiv*. DOI:10.64898/2026.03.12.711379
3. Wang T., Pan Y., Ju F., et al. (2025). CellNavi predicts genes directing cellular transitions by learning a gene graph-enhanced cell state manifold. *Nat. Cell Biol.* **27**:1863–1874. DOI:10.1038/s41556-025-01755-1
4. Hegde A. and Cheng J. (2026). GRNFormer: Accurate gene regulatory network inference using graph transformer. *Bioinformatics* **42**:btag144. DOI:10.1093/bioinformatics/btag144
5. Dai X., Shan C., Zheng J., et al. (2023). Biological factor regulatory neural network. *arXiv*. DOI:10.48550/arXiv.2304.04982

FUNDING AND ACKNOWLEDGMENTS

This work was supported by Major Project of Guangzhou National Laboratory (GZNL2024A01003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.