Transfer Learning of Condition-Specific Perturbation in Gene Interactions: Towards Multi-modal Foundational Modeling of Drug Response

Dongmin Bang [1 2 †] **Bonil Koo** [1 2 †] **Sun Kim** [1 2 3 4]

Abstract

Effective modeling of drug response requires a multi-scale framework that bridges molecular perturbations with cellular viability. While emerging biological foundation models offer promise for cross-scale transfer, they are not explicitly designed to capture chemical-induced perturbations across omics and chemical modalities. Here, we present Condition-Specific Gene-Gene Attention (CSG^2A), a multi-modal transfer learning framework that integrates transcriptomic profiles with compound structure and treatment conditions (e.g., dosage, time) to model drug responses at multiple scales. CSG^2A is first pretrained on large-scale drug-induced gene expression perturbation dataset to learn condition-aware gene interaction patterns through its gene-gene attention module, guided by interactome network priors. It is then transferred to cell viability dataset, achieving state-of-the-art performance in cell line drug response prediction. Case studies support the biological interpretability of the learned attention maps, aligning with known drug mechanisms. CSG^2A also generalizes to patient-level prediction on TCGA, demonstrating its potentials in cross-scale transfer and offering promising directions for developing multi-modal foundation models in drug response. The source code for the CSG^2A network is available at: github.com/eugenebang/CSG2A.

[†]Equal contribution. [1]Interdiciplinary program in bioinformatics, Seoul National University, Seoul, Republic of Korea [2]AIGENDRUG Co., Ltd., Seoul, Republic of Korea [3]Department of computer science and engineering, Seoul National University, Seoul, Republic of Korea [4]Interdiciplinary program in artificial intelligence, Seoul National University, Seoul, Republic of Korea. Correspondence to: Sun Kim <sunkim.bioinfo@snu.ac.kr>.

Proceedings of the ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences, Vancouver, Canada. 2025.

1. Introduction

Predicting drug response plays a pivotal role in cancer treatment and personalized medicine. Drug response can be defined at multiple levels of perturbation, including the gene (transcriptome) level, cell line (in vitro) level, and patient (clinical) level. At the transcriptomic level, the task involves predicting perturbed gene expression profiles in response to chemical treatments (Zhu et al., 2021; Hetzel et al., 2022). Cell line-level drug response prediction utilizes basal gene expression profiles and chemical information to predict cell viability measures like inhibitory concentration 50 (IC50) values (Garnett et al., 2012), while patient-level prediction aims to distinguish responders from non-responders (Ding et al., 2016; Huang et al., 2020). Integrative modeling of this multi-scale flow of drug-induced perturbation, from gene to cell then systemic scale, may be an approach to precision medicine, addressing the scarce data on the patient level.

However, most existing drug response prediction models overlook the potential of this multi-scale integration, often restricting their focus to a single biological level. A key challenge in modeling such trans-level drug response lies in bridging the gap between gene-level perturbations and cell-level phenotypic outcomes. Recent advances in attention-based architectures and biological foundation models have demonstrated success in transferring knowledge from large-scale omics data to disease- and population-levels (Chen et al., 2024; Shen et al., 2025; Brixi et al., 2025). However, these models are primarily designed for modality integration within omics or to vision domains, and are not explicitly tailored to capture chemical perturbation dynamics.

In response, we introduce a novel approach to bridge the gap between large-scale gene-level perturbation and cell-level viability databases, namely LINCS L1000 (Subramanian et al., 2017) and GDSC (Garnett et al., 2012), respectively. Specifically, we employ Attention mechanism to model chemical-induced gene-gene network perturbations and adapt a transfer learning approach to learn from the transcriptomic landscape, transferring knowledge obtained from gene-level to higher level drug responses (Fig. 1).

Our trans-level transfer learning involves pretraining on LINCS L1000 dataset, then finetuning on the GDSC dataset.

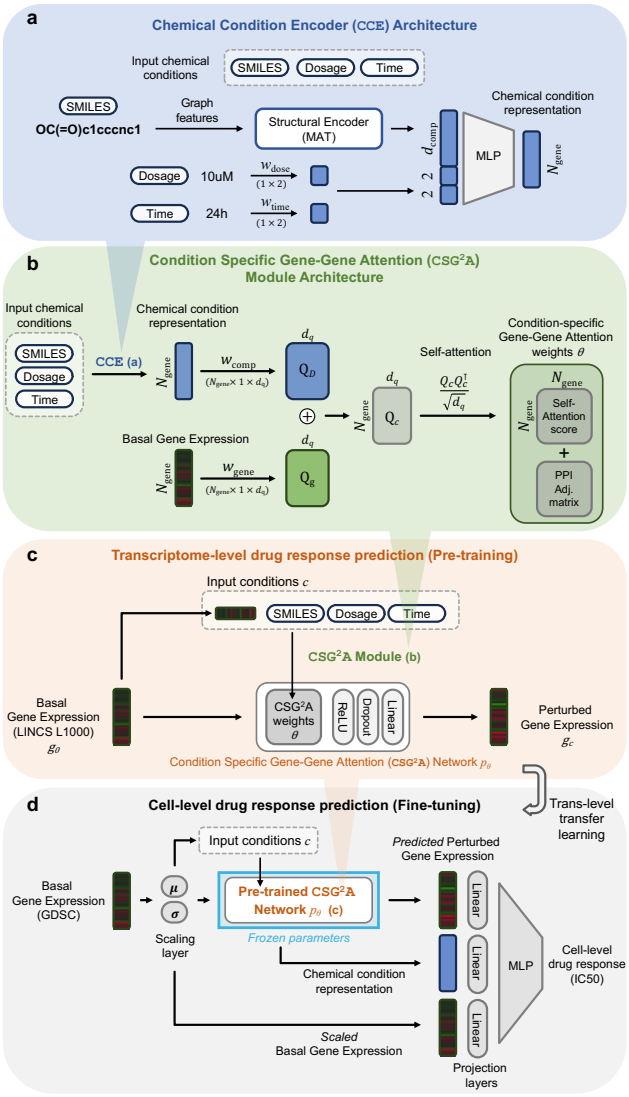

Figure 1: Overview and architecture of proposed framework. (**a**) Chemical Condition Encoder (CCE): Encodes chemical structure, dose, and duration into a 'chemical condition embedding'. (**b**) CSG^{2}A Module: Combines chemical and basal gene expression inputs to compute attention matrix, integrated with PPI to produce interaction weights θ. (**c**) Pretraining (LINCS L1000): CSG^{2}A predicts perturbed gene expression from basal profiles and chemical conditions. (**d**) Fine-tuning (GDSC): Adds a scaling layer and uses pretrained outputs to predict IC50 values.

The key component of our approach is the Condition-Specific Gene-Gene Attention (CSG^{2}A)–designed to dynamically learn gene interactions specific to input conditions, guided by both data and biological network priors. This procedure enables the our model to adeptly capture the chemical-induced perturbations in gene interactions, comprehensively understanding gene-level interactions for transfer to higher level drug response of cells and patients.

Our CSG^{2}A network, coupled with trans-level transfer learning strategy, achieves state-of-the-art performance compared to existing drug response prediction models on the GDSC dataset. Case studies further validate the alignment of learned condition-specific gene-gene attention with the known mode of action of drugs, highlighting the interpretability and biological relevance of our model. Additional experiments demonstrate its adaptability to predict drug responses in cancer patients from the TCGA dataset.

2. Related Work

2.1. Large-Scale Drug Response Datasets

Gene (Transcriptome) Level: The LINCS Connectivity Map (CMAP) L1000 (Subramanian et al., 2017) is a large-scale pharmacogenomics screening dataset from the Library of Integrated Network-based Cellular Signatures (LINCS). This dataset contains flow cytometry-based transcriptomic responses of cell lines to various perturbations, covering a wide spectrum of chemical treatments, cell types, and experimental conditions.

Cell Line (*in vitro*) Level: There are several comprehensive resources for understanding the relationship between genomic features and drugs at the cell line level: GDSC (Garnett et al., 2012), CTRP (Rees et al., 2016), CCLE (Barretina et al., 2012), and NCI-60 (Shoemaker, 2006). GDSC, one of the representative ones, provides drug sensitivity profiles for a variety of anti-cancer drugs, illustrating how different cancer cell lines respond to these drugs based on their genomic features.

Patient (Clinical) Level: The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) is a crucial resource that provides molecular characterization data for patients with various types of cancer. While TCGA primarily focuses on understanding the genomic landscape of cancers, it helps identify therapeutic targets, biomarkers, and patient subgroups, ultimately contributing to the development of more effective and personalized cancer treatments.

2.2. Drug Response Prediction in Multiple Levels

There has been various deep-learning approaches proposed for predicting drug responses on the cell level including autoencoders (Chiu et al., 2019; Rampášek et al., 2019), GNNs (Nguyen et al., 2021; Shin et al., 2022; Pak et al., 2024), and other architectures (Deng et al., 2020; Chawla et al., 2022; Jiang et al., 2022) (Further discussed in Appendix B).

Furthermore, models for predicting transcriptome-level perturbation, specifically at the single-cell resolution, has been gaining attention (Hetzel et al., 2022; Lotfollahi et al., 2023). However, multi-leveled transfer learning strategy for drug response prediction from gene level to higher levels has been rarely explored.

Table 1: Cell line drug response prediction performances on GDSC dataset with four different partitioning schemes. Model performances are assessed using RMSE and PCC metrics. The reported values represent averages and standard deviations across 10 cross-validation. The best performance is highlighted in bold, and the second-best performance is underlined. (RMSE: Root Mean Square Error; PCC: Pearson Correlation Coefficient)

Models	Mixed-set		Cell line-blind		Drug-blind		Disjoint-set	
	RMSE (↓)	PCC (↑)	RMSE (↓)	PCC (↑)	RMSE (↓)	PCC (↑)	RMSE (↓)	PCC (↑)
Random Forest	1.212 ± 0.017	0.905 ± 0.003	1.347 ± 0.058	0.881 ± 0.010	2.671 ± 0.579	0.406 ± 0.256	2.906 ± 0.220	0.370 ± 0.138
Support Vector Machine	1.126 ± 0.016	0.918 ± 0.002	<u>1.346 ± 0.062</u>	<u>0.881 ± 0.011</u>	<u>2.268 ± 0.437</u>	<u>0.520 ± 0.177</u>	<u>2.685 ± 0.350</u>	<u>0.450 ± 0.095</u>
GraphDRP (Nguyen et al., 2021)	1.217 ± 0.014	0.904 ± 0.002	1.457 ± 0.050	0.859 ± 0.010	2.354 ± 0.394	0.466 ± 0.163	2.844 ± 0.458	0.356 ± 0.108
PathDNN (Deng et al., 2020)	1.154 ± 0.011	0.928 ± 0.002	1.595 ± 0.076	0.862 ± 0.014	3.257 ± 0.666	0.336 ± 0.271	3.065 ± 0.471	0.383 ± 0.128
Precily (Chawla et al., 2022)	1.138 ± 0.016	0.917 ± 0.002	1.471 ± 0.063	0.856 ± 0.013	2.825 ± 0.400	0.362 ± 0.109	2.765 ± 0.344	0.426 ± 0.093
DRPreter (Shin et al., 2022)	1.104 ± 0.078	0.922 ± 0.011	1.495 ± 0.070	0.852 ± 0.013	2.473 ± 0.360	0.443 ± 0.175	2.745 ± 0.393	0.411 ± 0.148
DeepCoVDR (Huang et al., 2023)	1.019 ± 0.015	0.935 ± 0.002	1.394 ± 0.069	0.875 ± 0.012	2.754 ± 0.245	0.387 ± 0.200	3.001 ± 0.423	0.350 ± 0.139
DeepTTA(Jiang et al., 2022)	<u>0.974 ± 0.010</u>	<u>0.940 ± 0.001</u>	1.352 ± 0.060	0.881 ± 0.011	2.322 ± 0.496	0.502 ± 0.198	2.806 ± 0.512	0.404 ± 0.125
CSG^{2}A (Ours)	**0.942 ± 0.011**	**0.944 ± 0.001**	**1.342 ± 0.059**	**0.883 ± 0.010**	**2.119 ± 0.397**	**0.611 ± 0.140**	**2.442 ± 0.304**	**0.577 ± 0.082**

To the best of our knowledge, Dr.VAE (Rampášek et al., 2019) is the only prior work transferring drug response from gene expression to cell lines. It employs a VAE pretrained on L1000 and classifiers for cell-level prediction but trains separate models per drug, limiting generalization to unseen drugs and conditions.

3. CSG^{2}A: Condition-Specific Gene-Gene Attention Network

3.1. Model Architecture Overview

Our model comprises two key components: the Chemical Condition Encoder (CCE; Fig. 1a) and the Condition-Specific Gene-Gene Attention (CSG^{2}A; Fig. 1b) network, which together enable condition-aware modeling of transcriptomic perturbations.

First, the CCE encodes each treatment condition by integrating chemical structure, dosage, and time after exposure. The chemical structure is processed using a molecular foundational model (Maziarka et al., 2020), while dosage and time are scaled and projected via linear layers. These three components are concatenated and passed through a multilayer perceptron (MLP), producing a condition embedding aligned to the dimensionality of the gene space.

This embedding is combined with the basal gene expression profile to form a condition-specific gene representation. The CSG^{2}A module applies self-attention over these representations to compute pairwise gene-gene attention scores, which capture how genes interact under the given condition. These scores are further refined using a protein-protein interaction (PPI) adjacency matrix as a biological prior, producing an attention-informed adjacency matrix used as the weight matrix in the first linear layer of the network.

The downstream prediction network includes this attention-weighted linear layer, followed by a non-linear transformation, dropout, and output layer. By conditioning attention on both chemical and transcriptomic inputs, the model explicitly models how drug perturbations influence gene networks, improving interpretability and predictive power. The detailed description of the implementation details can be found in Appendix A.1.

3.2. Trans-Level Transfer Learning Strategy

Our framework comprises two stages: transcriptomic pretraining on LINCS L1000 and downstream fine-tuning on GDSC drug response data. The detailed explanations of the training strategy and dataset processing are provided in Appendix A.2 and A.3, respectively.

LINCS Pretraining: We train the model to predict perturbed gene expression from basal profiles and chemical conditions (structure, dosage, treatment time), minimizing the Mean-Squared Error (MSE) loss between the predicted and ground truth perturbed gene expression values. This phase can be interpreted as modeling the transition of transcriptomic cell states given a chemical condition. Central to this phase is the CSG^{2}A module, which learns condition-specific gene interactions from transcriptomic data.

Fine-Tuning on GDSC: Using the same inputs, we fine-tune the model to predict $\log$IC50 values, applying a trainable scaling layer with learnable mean and standard deviation to bridge dataset differences. All pretrained parameters are frozen, and bias terms are removed for better transferability. GDSC lacks dosage/time annotations, so we use standard values (10μM, 72h), commonly utilized for cell viability assays, across all experiments.

4. Results

4.1. Performance on Cell Line Drug Response Prediction with GDSC dataset

We first evaluated the cell-level drug response prediction performances on the GDSC dataset using 10-fold cross-validation under four data partitioning schemes aligning with the comprehensive investigation by Partin et al. (2023): (1) Mixed-set, where all cell lines and drugs can appear in both training and test sets; (2) Cell-blind, where test cell

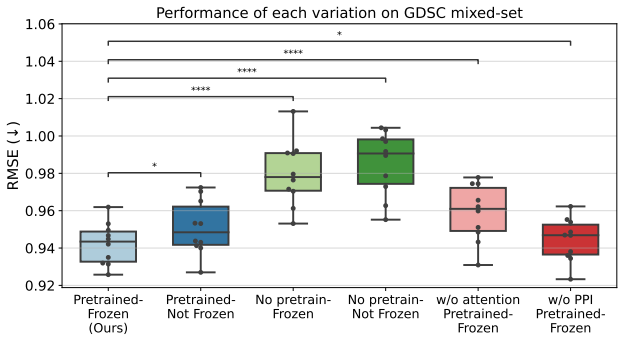

Figure 2: Performances of CSG^{2}A Network with variations in each modules. Our attention-based approach, equipped with LINCS L1000 pretraining, PPI network information and frozen parameters, shows statistically significant performance enhancements in GDSC dataset mixed-set. (*: paired t-test p-value < 0.05, ****: paired t-test p-value $<$ 1e-4).

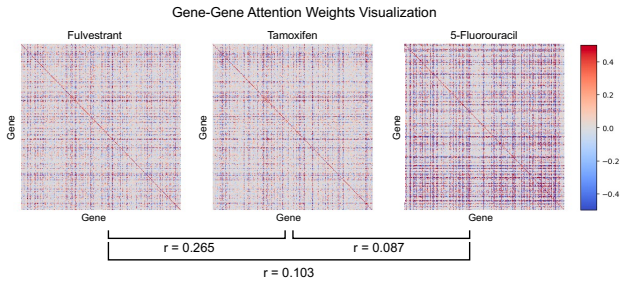

Figure 3: Visualization of gene–gene attention scores for three drugs: fulvestrant, tamoxifen (estrogen inhibitors), and 5-fluorouracil (DNA replication inhibitor). The pairwise correlation of the attention matrix is presented below.

lines are unseen during training; (3) Drug-blind, where test drugs are held out entirely; and (4) Disjoint-set, where neither drugs nor cell lines overlap between training and test sets. Performances were assessed in terms of root mean square error (RMSE) and Pearson correlation coefficient (PCC) metrics, measuring the distance and correlation between predicted and true logIC50 values, respectively.

Compared to eight baseline models—including Random Forest, SVM, and six state-of-the-art deep learning models—our model consistently achieved the lowest RMSE and highest Pearson correlation in all partitions. Notably, CSG^{2}A showed strong generalization in the more challenging drug-blind and disjoint-set settings, where other deep models typically degrade in performance due to overfitting.

Additional experiments on the NCI-60 dataset (Shoemaker, 2006) with 50% growth inhibitory concentration (GI50) prediction also demonstrated the outperformance of our model compared to baseline models (Appendix C.1.1).

4.2. Pretraining Gene-Gene Attention from Transcriptomic Landscape Improves Prediction

Our additional experiments revealed the significant impact of transfer learning from LINCS using our CSG^{2}A network on enhancing drug response prediction performances. By removing each component of CSG^{2}A, we observed that performance was highest when the model was pretrained on LINCS, and with all parameters frozen during fine-tuning except the prediction head (Fig. 2). This is contrary to common practice in transfer learning and suggests that the gene-gene interactions learned from transcriptomic perturbation are highly transferable. Removing pretraining, fine-tuning pretrained weights, omitting PPI priors, or replacing the attention mechanism with a linear model all led to significant drops in performance. Further discussion are provided in Appendix C.1.2.

In addition, we conducted a zero-shot investigation using the LINCS-pretrained CSG^{2}A network on the GDSC dataset. By evaluating the Euclidean distance between gene expressions before and after treatment, we observed that the perturbed gene expression points of the sensitive group were significantly farther from the basal gene expression points compared to the resistant group, demonstrating the capability of our LINCS-pretrained model to be capable of understanding the perturbation in cellular conditions at the gene level. Further experimental details and results are provided in Appendix C.2.

These results highlight the value of condition-specific pretraining and biological priors for robust, generalizable drug response prediction.

4.3. Gene Attention Scores and Predicted Expressions Reflect Drug Mechanisms

To evaluate biological interpretability of CSG^{2}A, we analyzed condition-specific gene attention maps and predicted expression profiles for several drugs. The methodological details of the case studies are provided in Appendix A.6.

Our results show that gene attention maps exhibit higher correlation between drugs with similar mechanisms of action (MoAs)—notably between fulvestrant and tamoxifen, both estrogen receptor inhibitors—compared to mechanistically distinct agents such as 5-fluorouracil (5-FU), a DNA replication inhibitor (Fig. 3). A more systemic scale investigation consistently display higher correlation between intra-class attention maps to inter-class drug pairs (Appendix C.3.1). Gene set enrichment analysis of the most highly attended gene pairs further revealed MoA-consistent pathway signatures: estrogen signaling and apoptosis pathways for fulvestrant, and DNA damage response pathways for 5-FU (Appendix C.3.2).

Additionally, we analyzed predicted gene expression profiles for drugs, with the up/down-regulated gene sets align-

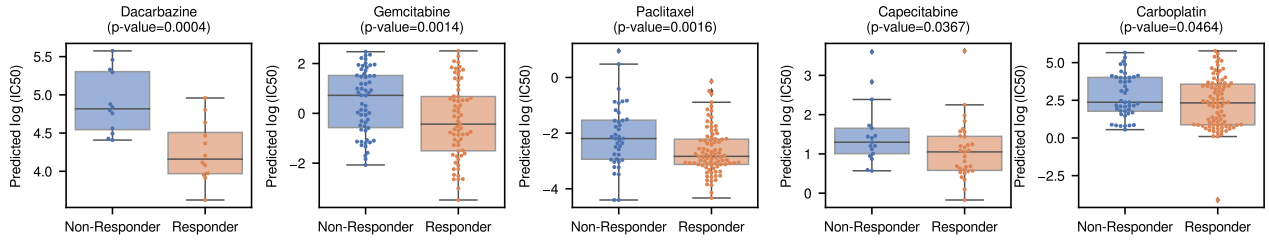

Figure 4: Patient response predictions in TCGA dataset. Using the GDSC-finetuned CSG^{2}A, we predicted drug responses in TCGA patients and significantly distinguished responders from non-responders ($p < 0.05$) for five drugs.

ing with known targets and downstream pathways (Appendix C.4). These findings show that CSG^{2}A's attention maps and predicted profiles are both predictive and biologically aligned with drug mechanisms.

4.4. Translation to Patient-level Transcriptome Data

In order to assess the model's ability to translate its knowledge gained from cell level to patient level, we applied the GDSC-finetuned models to predict IC50 values for TCGA patients based on tumor gene expression (Appendix A.5). Predicted responses were assessed by comparing IC50 distributions between responders and non-responders.

Among 13 drugs with sufficient patient response data, CSG^{2}A significantly differentiated responders for 5 drugs ($p < 0.05$) and 7 drugs ($p < 0.1$) through one-sided t-tests, outperforming DeepTTA and GraphDRP (Fig. 4 and Appendix C.5). These results highlight the model's capacity to transfer knowledge from cell lines to patient-level contexts.

5. Conclusion

In this work, we introduced CSG^{2}A, a condition-specific attention framework that models drug-induced gene interactions by integrating treatment conditions with transcriptomic data and biological priors. Through a trans-level transfer learning strategy—pretraining on perturbation profiles (LINCS) and fine-tuning on cell viability (GDSC)—our approach achieves state-of-the-art drug response prediction with strong generalization, particularly in drug-blind and disjoint settings. Crucially, freezing pretrained parameters preserves gene interaction contexts, enhancing performance. Attention scores and predicted gene expression profiles consistently align with known drug mechanisms, validating both interpretability and biological relevance. Potential future directions may include extending this approach to single-cell resolution data, broader patient cohorts, and multi-omics integration to deepen mechanistic understanding of drug response.

As we conclude, we believe our work underscores the promise of transfer learning of drug-induced gene interaction perturbation as a multi-modal foundation for robust and interpretable modeling in precision therapy.

References

Agrawal, A., Balcı, H., Hanspers, K., Coort, S. L., Martens, M., Slenter, D. N., Ehrhart, F., Digles, D., Waagmeester, A., Wassink, I., et al. Wikipathways 2024: next generation pathway database. *Nucleic acids research*, 52(D1): D679–D689, 2024.

Alcindor, T. and Beauger, N. Oxaliplatin: a review in the era of molecularly targeted therapy. *Current oncology*, 18(1):18–25, 2011.

Alves, C. L., Elias, D., Lyng, M., Bak, M., Kirkegaard, T., Lykkesfeldt, A. E., and Ditzel, H. J. High cdk6 protects cells from fulvestrant-mediated apoptosis and is a predictor of resistance to fulvestrant in estrogen receptor–positive metastatic breast cancer. *Clinical cancer research*, 22(22):5514–5526, 2016.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, 2012.

Brixi, G., Durrant, M. G., Ku, J., Poli, M., Brockman, G., Chang, D., Gonzalez, G. A., King, S. H., Li, D. B., Merchant, A. T., et al. Genome modeling and design across all domains of life with evo 2. *BioRxiv*, pp. 2025–02, 2025.

Carlson, R. W. The history and mechanism of action of fulvestrant. *Clinical breast cancer*, 6:S5–S8, 2005.

Chawla, S., Rockstroh, A., Lehman, M., Ratther, E., Jain, A., Anand, A., Gupta, A., Bhattacharya, N., Poonia, S., Rai, P., et al. Gene expression based inference of cancer drug sensitivity. *Nature communications*, 13(1):5680, 2022.

Chen, H., Venkatesh, M. S., Ortega, J. G., Mahesh, S. V., Nandi, T. N., Madduri, R. K., Pelka, K., and Theodoris,

C. V. Quantized multi-task learning for context-specific representations of gene network dynamics. *bioRxiv*, 2024.

Chen, J. and Zhang, L. A survey and systematic assessment of computational methods for drug response prediction. *Briefings in bioinformatics*, 22(1):232–246, 2021.

Chiu, Y.-C., Chen, H.-I. H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.-J., Huang, Y., and Chen, Y. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC medical genomics*, 12(1): 143–155, 2019.

Comella, P., Casaretti, R., Sandomenico, C., Avallone, A., and Franco, L. Role of oxaliplatin in the treatment of colorectal cancer. *Therapeutics and clinical risk management*, pp. 229–238, 2009.

Conroy, T., Pfeiffer, P., Vilgrain, V., Lamarca, A., Seufferlein, T., O'Reilly, E., Hackert, T., Golan, T., Prager, G., Haustermans, K., et al. Pancreatic cancer: Esmo clinical practice guideline for diagnosis, treatment and follow-up. *Annals of Oncology*, 34(11):987–1002, 2023.

De Angelis, P. M., Svendsrud, D. H., Kravik, K. L., and Stokke, T. Cellular response to 5-fluorouracil (5-fu) in 5-fu-resistant colon cancer cell lines during treatment and recovery. *Molecular cancer*, 5:1–25, 2006.

Deng, L., Cai, Y., Zhang, W., Yang, W., Gao, B., and Liu, H. Pathway-guided deep neural network toward interpretable and predictive modeling of drug sensitivity. *Journal of Chemical Information and Modeling*, 60(10):4497–4505, 2020.

Ding, Z., Zu, S., and Gu, J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics*, 32(19):2891–2895, 2016.

Garcia-Alonso, L., Iorio, F., Matchan, A., Fonseca, N., Jaaks, P., Peat, G., Pignatelli, M., Falcone, F., Benes, C. H., Dunham, I., et al. Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer research*, 78(3):769–780, 2018.

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., Greninger, P., Thompson, I. R., Luo, X., Soares, J., et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391):570–575, 2012.

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A. N., et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678, 2020.

Grivicich, I., Regner, A., Zanoni, C., Correa, L. P., Jotz, G. P., Henriques, J. A. P., Schwartsmann, G., and da Rocha, A. B. Hsp70 response to 5-fluorouracil treatment in human colon cancer cell lines. *International journal of colorectal disease*, 22:1201–1208, 2007.

Hetzel, L., Boehm, S., Kilbertus, N., Günnemann, S., Theis, F., et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Advances in Neural Information Processing Systems*, 35:26711–26722, 2022.

Huang, E. W., Bhope, A., Lim, J., Sinha, S., and Emad, A. Tissue-guided lasso for prediction of clinical drug response using preclinical samples. *PLoS computational biology*, 16(1):e1007607, 2020.

Huang, Z., Zhang, P., and Deng, L. Deepcovdr: deep transfer learning with graph transformer and cross-attention for predicting covid-19 drug response. *Bioinformatics*, 39(Supplement_1):i475–i483, 2023.

Jiang, L., Jiang, C., Yu, X., Fu, R., Jin, S., and Liu, X. Deeptta: a transformer-based model for predicting cancer drug response. *Briefings in bioinformatics*, 23(3): bbac100, 2022.

Johnson, W. E., Li, C., and Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007.

Lee, A. V., Cui, X., and Oesterreich, S. Cross-talk among estrogen receptor, epidermal growth factor, and insulin-like growth factor signaling in breast cancer. *Clinical Cancer Research*, 7(12):4429s–4435s, 2001.

Longley, D. B., Harkin, D. P., and Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nature reviews cancer*, 3(5):330–338, 2003.

Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I. L., Srivatsan, S. R., Naghipourfar, M., Daza, R. M., Martin, B., et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular systems biology*, 19(6):e11517, 2023.

Maziarka, Ł., Danel, T., Mucha, S., Rataj, K., Tabor, J., and Jastrzebski, S. Molecule attention transformer. *arXiv preprint arXiv:2002.08264*, 2020.

Nguyen, T., Nguyen, G. T., Nguyen, T., and Le, D.-H. Graph convolutional networks for drug response prediction. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(1):146–154, 2021.

Pak, M., Bang, D., Sung, I., Kim, S., and Lee, S. Dgdrp: drug-specific gene selection for drug response prediction via re-ranking through propagating and learning biological network. *Frontiers in Genetics*, 15:1441558, 2024.

Partin, A., Brettin, T. S., Zhu, Y., Narykov, O., Clyde, A., Overbeek, J., and Stevens, R. L. Deep learning methods for drug response prediction in cancer: predominant and emerging trends. *Frontiers in Medicine*, 10:1086097, 2023.

Rampášek, L., Hidru, D., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. Dr. vae: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*, 35(19):3743–3751, 2019.

Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. V., Gill, S., Javaid, S., Coletti, M. E., Jones, V. L., Bodycombe, N. E., et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature chemical biology*, 12(2):109–116, 2016.

Shen, W., Nguyen, T. H., Li, M. M. R., Huang, Y., Moon, I., Nair, N., Marbach, D., and Zitnik, M. Generalizable ai predicts immunotherapy outcomes across cancers and treatments. *medRxiv*, pp. 2025–05, 2025.

Shin, J., Piao, Y., Bang, D., Kim, S., and Jo, K. Drpreter: Interpretable anticancer drug response prediction using knowledge-guided graph neural networks and transformer. *International Journal of Molecular Sciences*, 23 (22):13919, 2022.

Shoemaker, R. H. The nci60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10): 813–823, 2006.

Skandalis, S. S., Afratis, N., Smirlaki, G., Nikitovic, D., Theocharis, A. D., Tzanakakis, G. N., and Karamanos, N. K. Cross-talk between estradiol receptor and egfr/igf-ir signaling pathways in estrogen-responsive breast cancers: focus on the role and impact of proteoglycans. *Matrix Biology*, 35:182–193, 2014.

Sterling, T. and Irwin, J. J. Zinc 15–ligand discovery for everyone. *Journal of chemical information and modeling*, 55(11):2324–2337, 2015.

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171 (6):1437–1452, 2017.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., et al. The string database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, 51(D1):D638–D646, 2023.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., et al. Gene set knowledge discovery with enrichr. *Current protocols*, 1 (3):e90, 2021.

Zhu, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., Liu, J., Yu, Y., Wang, L., Kong, W., et al. Prediction of drug efficacy from transcriptional profiles with deep learning. *Nature biotechnology*, 39(11):1444–1452, 2021.

A. Details in Methodology

Problem Definition We define the problem of predicting chemical-perturbed gene expression profile, denoted as g_c, as the design of a condition-specific prediction model p_θ. The predictive framework for the chemical-induced transcriptomic profile with the condition-specific neural network p_θ with parameters θ can be expressed as:

$$g_c = p_{(\theta|c)}(g_0) = p_{(\theta|g_0,S,d,t)}(g_0).$$

An essential aspect of this formulation is the explicit dependence of the parameters θ on the set of conditions $c = (g_0, S, d, t)$. Here, g_0 represents the basal gene expression profile before compound treatment, S denotes the chemical structure, and d and t as treatment dose and time, respectively. This formulation establishes a comprehensive framework for the prediction of chemical-induced gene expression profiles, where the neural network's behavior is intricately linked to the specific conditions.

In this section, we first describe the architectural details of our framework (Section A.1), then provide details in the trans-level transfer learning pipeline for translating from gene-level drug response to cell-level (Section A.2).

A.1. Architectural Details of the Condition-Specific Gene-Gene Attention Network

Chemical Condition Encoder Chemical-induced transcriptomic perturbation results from a biological cascade initiated by the binding of a chemical to its target proteins. The impact of this perturbation is mainly dictated by the chemical's structure, with additional dependencies on exposure dosage and time. Therefore, encoding the chemical condition necessitates incorporating not only structural information but also dosage and time factors.

To model the chemical condition comprehensively, we introduce a Chemical Condition Encoder (CCE) (Fig. 1a). This encoder takes structural features S, dosage d, and time t as input for producing a chemical condition representation $\mathcal{D}_{S,d,t}$.

The base component of our chemical structure encoder is the pretrained Molecular Attention Transformer (MAT, (Maziarka et al., 2020)). MAT is a Transformer-based structure encoder pretrained on a masked entity prediction task with 2 million compounds from the ZINC15 database (Sterling & Irwin, 2015). We leverage the pretrained weights provided by the authors throughout our experiments. The dosage and time conditions, scaled by $100\mu M$ and 72 hours respectively, are expanded to 2 dimensions by corresponding linear layers. The concatenated vector of these two expanded vectors and the MAT representation is then passed on to the final multi-layer perceptron (MLP) layer.

The mathematical representation of the chemical condition $\mathcal{D}_{S,d,t} \in \mathbb{R}^{N_{\text{gene}}}$ can be expressed as :

$$\mathcal{D}_{S,d,t} = \text{MLP}_{\text{CCE}}\big([\text{MAT}(S), d \cdot W_{\text{dose}}, t \cdot W_{\text{time}}]\big).$$

Here, $[\cdot,\cdot]$ refers to the concatenation operation, $W_{\text{dose}}, W_{\text{time}} \in \mathbb{R}^{1\times 2}$ are linear weights for dosage and time encoding, respectively, and MLP_{CCE} denotes the final MLP layer.

The overall process generates the final chemical condition representation aligned with the dimension of the number of genes (N_{gene}). This representation is further utilized by the Condition-Specific Gene-Gene Attention (CSG^2A) module for modeling the perturbed gene-gene network.

Condition-Specific Gene-Gene Attention Network To predict the chemical-induced gene expression profile, we incorporate condition-specific gene-gene attention CSG^2A values as our neural network parameters θ (Fig. 1b). This network is designed to capture the intricate relationships between genes in the treatment condition c under the influence of basal gene expression g_0, chemical structure S, dosage d and time t.

The process begins with the basal gene expression profile g_0 and chemical representation $\mathcal{D}_{S,d,t}$ obtained from the Chemical Condition Encoder (CCE). Then, the CSG^2A module introduces a novel paradigm by treating neural network parameters as condition-specific gene-gene attention values. This differentiation from conventional neural networks, where parameters are randomly initialized, enables the model to learn the conditional effects of both gene expression and drugs.

The calculation of attention values involves self-attention on the gene-level condition representation $Q_{(\mathcal{C}|g_0,S,d,t)} \in \mathbb{R}^{N_{\text{gene}}\times h}$. This self-attention module allows for the explicit learning of attention scores in a gene-specific manner, which is crucial for capturing gene expression perturbations effectively. The gene-level condition representation is defined as the sum of the

basal gene expression representation and the chemical condition representation:

$$Q_{(\mathcal{C}|g_0,S,d,t)} = Q_g + Q_{\mathcal{D}}$$
$$Q_g = W_{\text{gene}} \cdot g_0, \quad Q_{\mathcal{D}} = W_{\text{comp}} \cdot \mathcal{D}_{S,d,t},$$

where W_{gene} and $W_{\text{comp}} \in \mathbb{R}^{N_{\text{gene}} \times 1 \times h}$ represent learnable linear weights for the dimension expansion of gene expression and chemical condition representations into hidden dimension h. The utilization of a three-dimensional weight tensor $W \in \mathbb{R}^{N_{\text{gene}} \times 1 \times h}$ with distinct weights for each gene ensures that both the input expression value and the context of each gene are considered during dimension expansion.

Subsequently, the dot product attention score matrix $A_{Q_\mathcal{C}}$ is calculated as $\alpha_{i,j} = Q_{\mathcal{C}i} \cdot Q_{\mathcal{C}j}^{\intercal}$, where $\alpha_{i,j}$ denotes the attention score between genes i and j. To enhance ability to capture higher-order gene interactions in a biological context and reduce the search space by providing a starting point of the interactions and the scales, we integrate a PPI adjacency matrix (A_{PPI}) as prior knowledge as: $A_{\text{CSG}^2\text{A}} = A_{Q_\mathcal{C}} + A_{\text{PPI}}$.

The calculated attention score matrix directly serves as neural network weights. The first layer of the CSG^2A network is a linear layer with weights equal to $A_{\text{CSG}^2\text{A}}$. The network further includes an activation layer, a dropout layer, and a final linear layer. This comprehensive design allows the CSG^2A network to effectively capture and leverage condition-specific gene-gene interactions for precise predictions of perturbed gene expression.

A.2. Trans-Level Transfer Learning Strategy

In order to learn the chemical-perturbed gene network from the transcriptome data and leverage such model for higher level drug response predictions, we implement a trans-level transfer learning strategy.

LINCS Pretraining In the LINCS Pretraining phase (Fig. 1c), our model focuses on predicting the chemical-induced transcriptome profile, a crucial step in capturing the intricate dynamics of gene expression alterations under various chemical conditions. The inputs to this phase consist of the basal gene expression profile and the corresponding chemical structure for predicting the perturbed gene expression. Notably, our model also takes into account additional factors such as drug dosage and treatment time, enhancing its capacity to comprehend the nuanced aspects of chemical-induced transcriptomic changes. During training, our model utilizes Mean-Squared Error (MSE) loss on the perturbed gene expression profile, aiming to minimize the gap between predicted and actual values.

A key highlight of the LINCS Pretraining phase is the training of the Condition-Specific Gene-Gene Attention (CSG^2A) network. This network plays a pivotal role in learning condition-specific gene-gene interactions, contributing to the accurate prediction of perturbed gene expression profiles. By leveraging the diverse and extensive gene-level data available in LINCS, our model adapts to variations in gene expression induced by different chemical conditions. This acquired knowledge plays a critical role in the subsequent fine-tuning process, ensuring the model's adaptability and effectiveness in predicting in vitro drug responses during the later stages of our trans-level transfer learning strategy.

Fine-Tuning on GDSC In the GDSC Fine-tuning phase (Fig. 1d), our model undergoes further refinement to seamlessly adapt to the in vitro drug response dynamics observed in the GDSC dataset. The inputs for the fine-tuning task is consistent as the pretraining stage, including the basal gene expression profile and the corresponding chemical condition. However, the target value shifts to predicting the $\log$IC50 value, a key metric in quantifying the drug sensitivity of cancer cells.

To facilitate a smooth trans-level transfer, accounting for potential batch effects across different datasets, a trainable scaling layer is strategically introduced. This layer serves as a crucial bridge between the pretrained CSG^2A Network and the GDSC dataset. Notably, we intentionally eliminate all bias terms within the layers of our neural network, enhancing the adaptability of the transfer process between distinct transcriptomic spaces.

The introduced scalable layer incorporates two learnable latent variables: mean (μ) and standard deviation (σ). By passing through the scaling equation $g_0' = (g_0 - \mu)/\sigma$, the input basal gene expression profile (g_0) is aligned with the pretrained LINCS space. It is important to note that both the mean and standard deviation values function as latent variables, not having a preassigned target values. This design choice allows the model to autonomously learn and adapt its scaling parameters, contributing to its transferability to higher-level drug response tasks. The training target is the $\log$IC50 value, and the model is optimized to minimize the disparity between predicted $\log$IC50 values and actual values using the MSE loss.

An empirical observation emphasizes the critical role of the LINCS pretraining phase in learning gene-gene attention, achieving optimal prediction performance with fully frozen pre-trained parameters. Further details and insights into this observation are discussed in the Results section.

Lastly, it is important to note that the cell viability data from GDSC does not include information on the specific drug-treated conditions, such as dosage and treatment time. Therefore, to maintain consistency and align with the GDSC dataset's data generation process, we utilized the widely accepted treatment time of 72 hours for IC50 measurement and the dosage at 10uM, commonly used in LINCS L1000 dataset and also cell viability measurements, for all experiments.

A.3. Dataset and Metrics

A.3.1. LINCS L1000 Dataset

We downloaded the LINCS phase 1 data from GEO with accession number GSE92742 (Subramanian et al., 2017). Since DMSO is used as a control corresponding to the compound, transcriptome data were obtained from the sample treated with DMSO as basal gene expression. A total of 649 batches containing samples treated with DMSO were obtained, and from these batches, 202,962 transcriptomic profile of samples treated with compounds were obtained. We utilized the 978 landmark genes that were measured when the LINCS data was produced.

A.3.2. GDSC Dataset

We obtained basal gene expression profile for cell lines from Cell Model Passports (Garcia-Alonso et al., 2018). The $\log_2$FPKM values were transformed into robust z-scores to be used as input at the same level as the LINCS data. The robust z-score is computed using the following equation:

$$z_i = \frac{x_i - \text{median(X)}}{1.4826 \cdot \text{MAD}(X)},$$

where MAD indicates the median absolute deviation, X represents the expression values for a gene for all samples in the data, x_i is the expression level of a sample i, and z_i is a robust z-score for the gene of sample i. Additionally, the drug response values (i.e., $\log$IC50) for each cell line were obtained from GDSC (Garnett et al., 2012).

A.3.3. TCGA Dataset

Using the TCGA classification information of cell lines provided by GDSC, transcriptome data of patients corresponding to tumor samples for 21 cancer types were obtained from UCSC Xena (Goldman et al., 2020). For each cancer type, GDSC data and TCGA data were batch-corrected using Combat (Johnson et al., 2007) at the $\log_2$FPKM level and then converted to a robust z-score. Curated data on drug treatment and responsiveness in TCGA patients were obtained from the supplementary data provided in (Ding et al., 2016).

A.3.4. PPI Network

STRING v12.0 (Szklarczyk et al., 2023) was employed as the biological network for prior knowledge to guide the gene-gene interaction learning process. To ensure the inclusion of confident edges, edges with a combined score greater than 900 were selected.

A.4. Details on the four data partitioning schemes

Aligning with the comprehensive investigation by Partin et al. (2023), we conducted 10-fold validation in four distinct data partitioning schemes to evaluate each models' generalizability to diverse scenarios. The mixed-set scenario (known cancer and drugs) is commonly employed for its simplicity in analyzing and implementing drug response prediction models. The cell line-blind scenario (unknown cell lines and known drugs) is for simulating personalized cancer treatment utility, and the drug-blind scenario (known cancers and unknown drugs) presents challenges in developing novel drugs for cancer treatment. The disjoint-set scenario (unknown cancers and drugs) is used to assess each models' capacity to generalize in the challenging scenarios and its potential application in more clinically relevant settings. The training set comprised 80% of the data, with 10% allocated to the validation set and 10% to the test set.

A.5. Prediction of patient drug responsiveness in TCGA data

Using the TCGA(Weinstein et al., 2013) classification information of cell lines provided by GDSC (Garnett et al., 2012), transcriptome data corresponding to tumor samples for 21 cancer types were obtained from UCSC Xena (Goldman et al., 2020). The cancer types include BLCA, BRCA, CESC, COAD, ESCA, GBM, HNSC, KIRC, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PRAD, READ, SKCM, STAD, THCA, and UCEC, with COAD and READ merged into COREAD.

For each cancer type, GDSC and TCGA data were underwent batch correction at the $\log_2$FPKM level using Combat (Johnson et al., 2007). Then, the values were normalized into robust z-scores using the following equation to be used as input at the same level:

$$z_i = \frac{x_i - \text{median}(X)}{1.4826 \cdot \text{MAD}(X)}, \tag{1}$$

where MAD denotes the median absolute deviation, X represents the expression values for a gene across all samples, x_i is the expression level of a sample i, and z_i is the robust z-score for the gene of sample i. To predict IC50 values for TCGA data, models were trained for each cancer type using GDSC data.

Curated data on drug treatment and responsiveness in TCGA patients were obtained from the supplementary data provided in (Ding et al., 2016). Patients were divided into two groups: responder (including complete response and partial response) and non-responder (including stable disease and clinical progressive disease). Within both the responder and the non-responder groups, 13 drugs with more than 10 samples were identified: cisplatin, capecitabine, 5-fluorouracil, oxaliplatin, leucovorin, temozolomide, carboplatin, pemetrexed, gemcitabine, docetaxel, dacarbazine, doxorubicin, and paclitaxel.

The TCGA transcriptome data were inputted into the trained models for each cancer type, inferring IC50 values for each patient and drug pair. Then, one sided t-test was conducted to test the hypothesis that IC50 values in the responder group are lower than those in the non-responder group.

A.6. Gene set enrichment analysis on condition-specific gene attention scores

The proposed Condition-Specific Gene-Gene Attention (CSG^2A) module is designed for capturing the gene-gene interactions induced by the basal gene expression and chemical treatment condition. The resulting gene-gene attention scores are directly utilized as neural network parameters for predicting downstream target values, specifically, perturbed gene expression for LINCS L1000 and IC50 values for GDSC. Here, we describe the procedure adpoted to assess the association between known drug mechanisms and the gene-gene attention values.

Starting from the test set samples in the GDSC, we extracted the gene-gene attention map of the drugs-of-interest (DOI), fulvestrant and 5-fluorouracil (5-FU). The test set contained a total of 151 fulvestrant-treated samples and 99 5-FU treated samples. The attention maps were aggregated by averaging, resulting in a representative attention map for each drug.

After performing absolute operation on the attention scores, we identified the top 1,000 gene-gene attention values for each drugs as 'most-perturbed gene interactions'. Then all the genes appearing in the 1,000 pairs are selected as 'most-perturbed gene set', prepared for downstream Gene Set Enrichment Analysis (GSEA). Using the widely-used GSEA tool Enrichr (Xie et al., 2021), we identified enriched pathways using the pathway annotations from WikiPathways (Agrawal et al., 2024). For the two DOIs, top 10 pathways ranked by adjusted p-value are identified.

B. Cell-level drug response prediction models

B.1. AutoEncoder-based Models

DeepDR (Chiu et al., 2019) is a AutoEncoder-based deep learning model designed for predicting drug responses in cancer cells based on mutation and expression profiles. The model, incorporating pre-trained encoders and a drug response predictor network, demonstrated its effectiveness in predicting drug responses across various cancer types. However, the model does not use any drug information, making it difficult to use from a new drug development perspective.

Dr.VAE (Rampášek et al., 2019) is the first work to attempt the integration of knowledge from transcriptomic level drug response for cell line drug response prediction. Utilizing the Variational AutoEncoder framework, the authors pre-trained models on LINCS L1000 data and applied additional classifiers to predict cell line level drug responses. However, this model lacks consideration for various chemical treatment conditions. Notably, Dr.VAE trains separate models for each drug with fixed dosage and time point, limiting its ability to predict responses for drugs and conditions absent from the LINCS

L1000 dataset.

B.2. Graph Neural-Network Based Models

GraphDRP (Nguyen et al., 2021) introduced a method based on GNNs for drug response prediction by representing drugs as molecular graphs and cell lines as binary vectors of genomic aberrations. It demonstrated the efficacy of graph-based representations in enhancing drug response prediction. However, since 1D convolutional layers are employed to learn cell line representation, interactions between genes cannot be taken into account.

DRPreter (Shin et al., 2022) is an interpretable model utilizing graph neural networks to predict anticancer drug response. DRPreter incorporates domain knowledge on biological pathways, employs a transformer to detect relationships between pathways and drugs, while also providing insights into the mechanism of action. However, bias can occur when only information about specific pathways and their associated genes is utilized. Also, the fixed structure of pathways limit the model in learning chemical-induced perturbation in gene-gene interaction level.

B.3. Other Approaches

PathDNN (Deng et al., 2020) proposed a pathway-guided deep neural network (DNN) model to predict drug sensitivity in cancer cells. The model, incorporating biological pathway information, demonstrated improved interpretability, highlighting its potential for enhancing understanding and prediction of drug sensitivity in cancer treatment. However, due to the model's reliance on the drug's target information as inputs, it becomes challenging to apply it to drugs lacking such target information.

Precily (Chawla et al., 2022) utilizes a simple DNN that takes pathway scores processed from gene expression profiles and integrates them with drug descriptors, providing insights into the biological mechanisms influencing drug resistance.

DeepTTA (Jiang et al., 2022) integrates transformer-based drug representation learning with a feed-forward network for predicting anti-cancer drug responses using transcriptomic data and drug chemical substructures. However, the model structure mainly focuses on drugs and does not account for interactions between genes.

DeepCoVDR (Huang et al., 2023) utilizes a graph transformer to encode chemical compounds and feed-forward layers to encode cell-lines. Then, a cross-attention module integrates compound embedding and cell-line embedding by considering their interaction. DeepCoVDR then predicts IC50 value using the representations of the compound, cell-line, and interaction features.

C. Additional Experimental Results

C.1. Performance on Cell Line Drug Response Prediction

First, we applied our framework and compared with other state-of-the-art models in predicting cell line drug responses of the GDSC dataset. Aligning with the comprehensive investigation by Partin et al. (Partin et al., 2023), our evaluation encompassed a 10-fold cross-validation, employing four distinct data partitioning schemes further detailed in Section A.4

Among the eight comparison models in our evaluation, two were machine learning algorithms—Random Forest (RF) and Support Vector Machines (SVM). These models utilized Morgan molecular fingerprints and gene expression values as input features. Additionally, six deep learning methods (GraphDRP, PathDNN, Precily, DRPreter, DeepCoVDR, and DeepTTA) were included in the comparative analysis.

Performance metrics were assessed in terms of root mean square error (RMSE) and Pearson correlation coefficient (PCC), measuring the distance and correlation between predicted and true $\log$IC50 values, respectively.

Remarkably, our proposed model demonstrated state-of-the-art performance across all four partitioning schemes, exhibiting significant improvements, particularly in challenging drug-blind and disjoint-set scenarios (Table 1). While existing deep learning approaches have shown improved performances in mixed-set settings, they often exhibit a lack of generalizability, leading to noticeable decreases in performance in harsh split settings. This phenomenon may stem from the widely-observed overfitting tendency of deep learning models, allowing conventional machine learning algorithms to outperform them. However, our CSG^{2}A network consistently achieved the best performances in all splits, indicating the robust generalizability of our condition-specific pretraining approach to challenging validation settings.

Additional experiments on the NCI-60 dataset (Shoemaker, 2006) with cell-level 50% growth inhibitory concentration (GI50) prediction also demonstrated the outperformance of the CSG^{2}A network compared to baseline models (Table C.1.1). These results underscore the adaptability of our model in diverse application contexts, especially in the challenging grounds of drug discovery.

C.1.1. Performances on Cell Line Drug Response Prediction with NCI-60 Dataset

Additional validation on NCI-60 dataset (Shoemaker, 2006) retrieved from the supplementary data by Chen & Zhang (2021)[1]. Unlike the GDSC dataset, the processed NCI-60 provides drug sensitivity data in the form of the 50% growth inhibitory concentration (GI50). We evaluated the performance of our method alongside existing drug response prediction methods, and our method showed superior performance (Table 2).

Table 2: Drug Response Performances on the NCI-60 dataset. The best performance is highlighted in bold, and the second-best performance is underlined. (RMSE: Root Mean Square Error; PCC: Pearson Correlation Coefficient)

Models	RMSE (↓)	PCC (↑)
RF	0.833 ± 0.032	0.403 ± 0.067
SVM	0.835 ± 0.034	0.403 ± 0.073
DRPreter	0.860 ± 0.029	0.327 ± 0.123
GraphDRP	0.838 ± 0.021	0.390 ± 0.051
DeepCoVDR	0.812 ± 0.021	0.482 ± 0.044
DeepTTA	<u>0.773 ± 0.028</u>	<u>0.548 ± 0.043</u>
CSG^{2}A	**0.759 ± 0.022**	**0.564 ± 0.031**

C.1.2. Ablation studies

Our additional experiments reveal the significant impact of transfer learning from LINCS using our CSG^{2}A network on enhancing drug response prediction performances. Fig. 2 illustrates the model performances of our model on GDSC mixed-set in RMSE, based on the variations on the inclusion of pretraining and freezing.

An intriguing finding is that our model achieved its best performance when pretrained on LINCS, and the model parameters were frozen during fine-tuning. Importantly, our analysis exposes a statistically significant drop in performance (paired t-test p-value < 0.05) on 10CV rounds when layers were fine-tuned (mean RMSE 0.951) compared to the frozen CSG^{2}A network (mean RMSE 0.942). This is contrary to the conventional trend in transfer learning for deep models, where fine-tuning the parameters typically improves performance. This suggests that the gene-gene interactions learned from the transcriptomic landscape hold substantial value, and losing this context during fine-tuning does not contribute to performance improvement.

Moreover, pretrained models consistently exhibited superior performances compared to non-pretrained models, surpassing even the fine-tuned unfrozen model (mean RMSE 0.985). This indicates the importance of LINCS pretraining in capturing the dynamics of gene-gene network perturbation, emphasizing its role in boosting predictive capabilities. We also observed a statistically significant decrease in performance when the PPI network information was not integrated into the attention score matrix (mean RMSE 0.945), emphasizing the importance of informative prior knowledge.

Lastly, the performance comparison between the CSG^{2}A network and a plain linear prediction model, which does not perform attention but just adds the chemical condition representation and the basal gene expression, empirically demonstrates the importance of utilizing the attention module for best performance.

Overall, these observations underscore the success of our trans-level transfer learning strategy with attention module, highlighting the effective integration of knowledge from transcriptomic landscape to enhance predictions for cell line-level drug responses.

C.2. Zero-shot Prediction of Drug Response from Inferred Perturbed Gene Expression

To evaluate the zero-shot prediction capabilities of the LINCS-pretrained model on cell-level data, we utilized the GDSC dataset without any additional fine-tuning. Our central hypothesis was that drug-sensitive cells, undergoing phenotypic changes from uncontrolled proliferation to cell cycle arrest or death (correlating with lower IC50 values), would exhibit

[1] https://github.com/Jinyu2019/Suppl-data-BBpaper

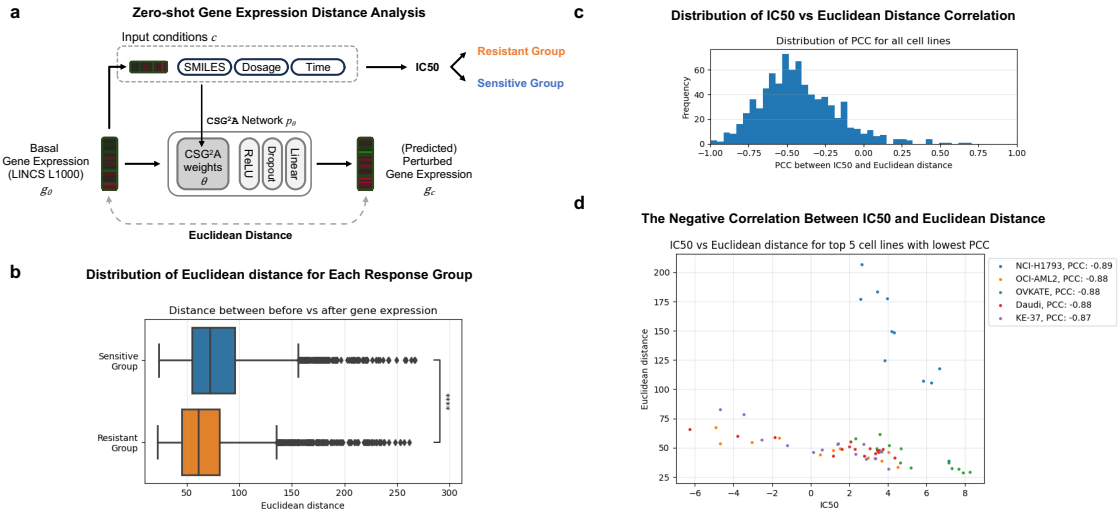

Figure 5: Zero-shot prediction of drug response from perturbed gene expression. (**a**) The zero-shot prediction framework based on inferred perturbed gene expression using the LINCS-pretrained CSG^{2}A network. (**b**) Distribution of Euclidean distance for each response group. Statistical tests have verified that the resistant group exhibited significantly closer proximity to the basal gene expression, as evidenced by a one-sided t-test yielding a p-value of 5.1E-97. (**c**) Distribution of IC50 vs Euclidean distance correlations. The distribution of Among the 803 cell lines, 756 cell lines (94.1%) showed negative correlations between the Euclidean distance and the logIC50 value of the samples from the corresponding cell line. (**d**) The negative correlation between IC50 and Euclidean distance. The results suggest that lower the IC50 value, indicating higher sensitivity to the input compound, the greater the distance observed in Euclidean gene expression space created the LINCS-pretrained CSG^{2}A network. (PCC: Pearson Correlation Coefficient; ****: t-test p-value $<$ 1E-4)

greater changes in gene expression upon treatment compared to resistant cells. This would manifest as a larger Euclidean distance between basal and perturbed gene expressions.

For each of the 803 cell lines in the GDSC dataset, we first calculated the average logIC50 value. This allowed us to categorize all samples within a cell line as either 'sensitive' or 'resistant' based on whether their individual logIC50 value fell below or above the cell-line average, respectively.

Next, we computed the Euclidean distance between the basal gene expression (original GDSC value) and the predicted perturbed gene expression. The perturbed gene expression was inferred by the LINCS-pretrained model under treatment conditions set to 10 μM for 72 hours, a common setting for IC50 measurements.

We then compared the distribution of Euclidean distances for the sensitive and resistant groups. A one-sided t-test, was performed to assess the difference in proximity to basal gene expression between the two groups. Furthermore, for each of the 803 cell lines, we calculated the correlation between the Euclidean distance and the logIC50 value of the samples. This allowed us to investigate the relationship between drug sensitivity and the magnitude of gene expression perturbation.

Table 3: Distinguishing drug responders from non-responders in TCGA dataset. The table presents the results of evaluating the GDSC-finetuned CSG^{2}A model's ability to predict drug responses in patient data from TCGA. Our model outperformed DeepTTA and GraphDRP in distinguishing responders from non-responders, demonstrating enhanced discriminative power with significant p-values for five drugs below 0.05 and seven drugs below 0.1. These findings underscore the efficacy of our model in transferring knowledge for predicting patient responses.

Model	Drugs with p <0.05		Drugs with p <0.1	
	Drugs	Count	Drugs	Count
CSG^{2}A	Dacarbazine, Gemcitabine, Paclitaxel, Capecitabine, Carboplatin	5 (38.5%)	Dacarbazine, Gemcitabine, Paclitaxel, Capecitabine, Carboplatin, Docetaxel, Leucovorin	7 (53.8%)
GraphDRP	Gemcitabine, Docetaxel, Cisplatin	3 (23.1%)	Gemcitabine, Docetaxel, Cisplatin, Leucovorin	4 (30.8%)
DeepTTA	Cisplatin, Dacarbazine	2 (15.4%)	Cisplatin, Dacarbazine, Docetaxel	3 (23.1%)

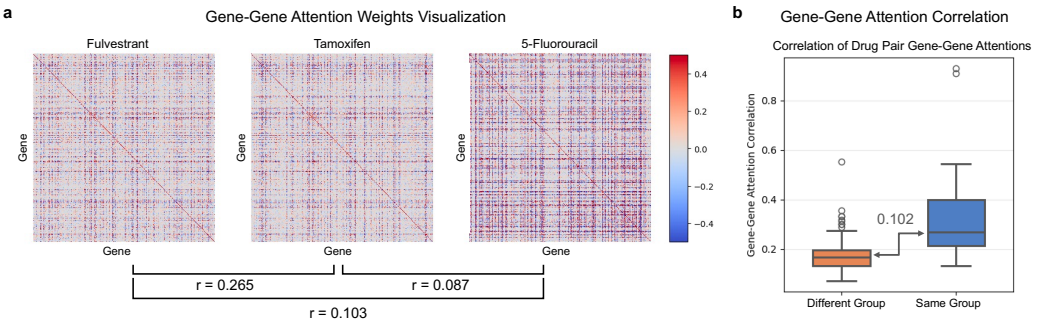

Figure 6: Analysis of Gene-Gene Attention Scores. (**a**) Visualization of gene-gene attention values for three drugs: fulvestrant, tamoxifen, and 5-fluorouracil. The pairwise correlation of the attention matrix is presented below. (**b**) Correlation of gene-gene attentions among drugs within the same group and across different groups. The correlation within drug pairs from the same group is significantly higher than those from different groups. (r: Pearson Correlation Coefficient)

C.3. Additional Results on Gene-Gene Attention Score and MoA Investigation

In this section, we showcase a case study that underscores the effectiveness of our model in capturing perturbed gene-gene interactions aligned with the Mechanism of Action (MoA) of drugs.

C.3.1. Systemic Investigation of Attention Map Similarities

In the Results Section 4.3, we have compared the attention maps of estrogen inhibitors (fulvestrant, tamoxifen) to DNA replication inhibitor (5-FU) and saw the higher correlation of attention scores within estrogen inhibitors compared to inter-class pairs (Fig. 6a). To provide a more systematic demonstration, we explored the similarity in gene-gene attention maps concerning the MoA of drugs. Categorizing anti-cancer drugs within the GDSC dataset into groups, such as DNA cleaving drugs, Cross-linking drugs, Intercalating drugs, Topoisomerase inhibitors, Antimetabolites, Antitubulin drugs, and Tyrosine kinase inhibitors, we observed that drugs sharing MoA exhibited significantly higher correlation in gene-gene attention patterns compared to drugs with differing MoAs (Fig. 6b). The average correlation between drugs in different categories was 0.169, while drug pairs within the same category showed an higher average correlation of 0.338 resulting in median difference of 0.102, analogous to the fulvestrant–tamoxifen–5-FU case.

C.3.2. GSEA Results of Fulvestrant and 5-FU

We further performed gene set enrichment analysis (GSEA) using test set data treated with either fulvestrant or 5-fluorouracil (5-FU). The procedures for gene set selection from attention maps and the subsequent enrichment analysis are detailed in Appendix A.6.

In the case of Fulvestrant, an FDA-approved selective estrogen receptor degrader (SERD), our model highlighted pathways such as Integrated breast cancer pathway (WP1984) and Apoptosis (WP254) (Alves et al., 2016). These findings intricately align with Fulvestrant's MoA, targeting estrogen receptor signaling and exhibiting anti-cancer effects, particularly in breast cancer (Carlson, 2005). Similarly, 5-FU, recognized as an anti-metabolite regulating nucleotide synthesis essential for DNA replication (Longley et al., 2003), manifested perturbations in pathways such as DNA damage response (WP707) (De Angelis et al., 2006) and Apoptosis Modulation by HSP70 (WP384) (Grivicich et al., 2007). The full list of enriched pathways are listed in Table 4, demonstrating on how the PPI network guides the attention map with biological prior knowledge.

Overall, these comprehensive analysis reinforces the model's ability to discern condition-specific gene interactions aligned with the diverse MoAs of anti-cancer drugs.

C.4. Additional Results on Predicted Gene Expression and MoA Investigation

To further validate the effectiveness of our CSG^{2}A network in translating knowledge from the transcriptome level to the cell level, we conducted a comprehensive analysis on the predicted chemical-induced gene expression profiles. To achieve this, we employed two key analyses: Differentially Expressed Gene (DEG) analysis and Gene Set Enrichment Analysis (GSEA).

Table 4: Attention analysis enrichment results The top 10 enriched pathways on highly perturbed gene sets for fulvestrant and 5-fluorouracil, ranked by adjusted p-value. The table presents the enrichment results using the whole GDSC testset samples. Results demonstrate the alignment of the enriched pathways with the Mode of Action of the two drugs. (adj. p.: adjusted p-value)

Fulvestrant (151 samples)			**5-Fluorouracil (99 samples)**		
Rank	Attention-enriched pathway	adj. p.	Rank	Attention-enriched pathway	adj. p.
1	Integrated Cancer Pathway WP1971	2.3e-5	1	DNA damage response WP707	1.5e-3
2	Integrated breast cancer pathway WP1984	2.3e-5	2	Gastrin signaling pathway WP4659	1.5e-3
3	DNA damage response WP707	2.7e-4	3	miRNA regulation of DNA damage response WP1530	1.5e-3
4	ErbB signaling pathway WP673	2.7e-4	4	Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	2.2e-3
5	RAC1/PAK1/p38/MMP2 Pathway WP3303	2.7e-4	5	Alzheimer's disease WP2059	2.5e-3
6	miRNA regulation of DNA damage response WP1530	3.0e-4	6	Apoptosis Modulation by HSP70 WP384	2.8e-3
7	Apoptosis WP254	9.0e-4	7	Unfolded protein response WP4925	5.5e-3
8	Endometrial cancer WP4155	9.0e-4	8	miRNA regulation of p53 pathway in prostate cancer WP3982	5.5e-3
9	Retinoblastoma gene in cancer WP2446	9.6e-4	9	Apoptosis WP254	1.3e-2
10	Pancreatic adenocarcinoma pathway WP4263	1.5e-3	10	Retinoblastoma gene in cancer WP2446	1.4e-2

Starting with the test set predictions from the GDSC dataset, we utilized our model to predict gene expressions based on the basal gene expression profile and input compound. Subsequently, we categorized the dataset into two groups based on IC50 values: high-IC50 and low-IC50. We then performed DEG analysis on the resulting gene expression profiles. Focusing on two most-frequently occurring drugs in the test set, oxaliplatin and fulvestrant, we conducted a gene set enrichment analysis using the Enrichr (Xie et al., 2021) with Wikipathways (Agrawal et al., 2024) gene sets. The top 5 enriched pathways for the 'sensitive' low-IC50 group, based on adjusted p-values, are summarized in Table 5.

Oxaliplatin, a platinum-based chemotherapeutic drug, forms DNA adducts and induces DNA damage for its anti-cancer effects (Alcindor & Beauger, 2011). The enriched over-expressed pathways from our model's predicted transcriptomic profiles include DNA Mismatch Repair (WP531), DNA Replication (WP466), Nucleotide Excision Repair (WP4753), and G1 to S cell cycle control (WP45), directly align with the known MoA of oxaliplatin. Additionally, the down-regulation of pathways of Chromosomal and microsatellite instability in colorectal cancer (WP4216) and Pancreatic adenocarcinoma pathway (WP4263) suggests its potential therapeutic efficacy in these cancers, consistent with the drug's clinical indications (Comella et al., 2009; Conroy et al., 2023). The results highlight the capability of CSG^{2}A in capturing and interpreting complex cellular responses at both genomic and functional pathway levels.

A similar pattern emerges in the case of fulvestrant, a selective estrogen receptor degrader (SERD). First approved by the FDA in 2002, fulvestrant's MoA is reflected in the most enriched pathway for the predicted suppressed genes in the low-IC50 group: Estrogen signaling pathway (WP712). The down-regulation of the Endometrial cancer pathway, associated with estrogen receptor regulation, further emphasizes the drug's impact on estrogen-related mechanisms. The observed over-expression of growth factor pathways including PDFG (WP2526) and VEGF (WP3888), coupled with the down-regulation of the EGFR inhibitor resistance pathway (WP4806), suggests cellular adaptations in the absence of estrogen-mediated growth signals. The well-established understanding of the intricate cross-talk between the estrogen signaling pathway and the PDGF, VEGF, and EGFR pathways (Skandalis et al., 2014; Lee et al., 2001) provides insights into these observations.

In summary, our analysis of oxaliplatin and fulvestrant showcases that the predicted gene expressions by CSG^{2}A consistently align with the known MoA of these drugs, demonstrating the model's capability to capture and interpret complex cellular responses across both cell-line and transcriptomic levels.

C.5. Additional Results on Translation to Patient-level Transcriptome Data

In order to assess the model's ability to translate its knowledge gained from the GDSC dataset to predict drug responses in the context of patient-specific data from TCGA, we leveraged the GDSC-finetuned CSG^{2}A model to predict IC50 values for various anti-cancer drugs given patients' basal tumor tissue gene expression profiles (Dataset details in Section A.5). The assessment relies on the well-established notion that lower IC50 values indicate drug sensitivity, with responders exhibiting a lower IC50 distribution than non-responders.

As part of the performance assessment, one-sided t-test was employed to compare the lower distribution of predicted IC50

Table 5: Predicted gene expression enrichment results. Top-5 Enriched pathways on over-expressed and suppressed gene sets for oxaliplatin and fulvestrant, ranked by adjusted p-value. (adj. p.: adjusted p-value)

Oxaliplatin					
Rank	Over expressed gene-enriched pathway	adj. p.	Rank	Suppressed gene-enriched pathway	adj. p.
1	DNA Mismatch Repair WP531	4.5e-5	1	Chromosomal and microsatellite instability	1.2e-3
2	Retinoblastoma gene in cancer WP2446	9.0e-5	2	Pancreatic adenocarcinoma pathway WP4263	2.2e-3
3	DNA Replication WP466	3.3e-4	3	Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	6.6e-3
4	Nucleotide Excision Repair WP4753	3.3e-4	4	EGF/EGFR signaling pathway WP437	6.6e-3
5	G1 to S cell cycle control WP45	1.7e-3	5	Gastrin signaling pathway WP4659	6.6e-3
Fulvestrant					
Rank	Over expressed gene-enriched pathway	adj. p.	Rank	Suppressed gene-enriched pathway	adj. p.
1	PDGF Pathway WP2526	3.1e-3	1	DNA damage response (only ATM dependent) WP710	3.5e-3
2	Mammary gland development pathway - Puberty (Stage 2 of 4) WP2814	8.9e-3	2	Estrogen signaling pathway WP712	3.5e-3
3	Photodynamic therapy-induced HIF-1 survival signaling WP3614	8.9e-3	3	Hepatitis B infection WP4666	3.5e-3
4	RAC1/PAK1/p38/MMP2 Pathway WP3303	8.9e-3	4	Endometrial cancer WP4155	1.2e-2
5	VEGFA-VEGFR2 Signaling Pathway WP3888	8.9e-3	5	EGFR Tyrosine Kinase Inhibitor Resistance WP4806	1.9e-2

values between responders and non-responders. The model's ability to distinguish responders from non-responders was further evaluated through a comparative analysis with DeepTTA and GraphDRP, also trained on the GDSC dataset.

Among 13 drugs with over 10 responders and non-responders each, our model demonstrated enhanced discriminative power (Fig. 4 and Table 3). Specifically, it distinguished responses for 5 drugs with p-values below 0.05 and 7 drugs below 0.1. In contrast, DeepTTA and GraphDRP achieved fewer significant differentiations, with 2,3 and 3,4 drugs below p-values 0.05 and 0.1, respectively. These results underscore the efficacy of our model in transferring knowledge for predicting patient responses and its outperformance in comparison to existing methods.