FUSED: Cross-Domain Integration of Foundation Models for Cancer Drug Response Prediction

Anonymous Author(s)
Affiliation
Address
email

Abstract

AI-driven methods for predicting drug responses hold promise for advancing personalized cancer therapy, but cancer heterogeneity and the high cost of data generation pose substantial challenges. Here we explore the transfer learning capability and introduce FUSED (**Fu**sion of Foundation Model **E**mbeddings for **D**rug Response Prediction), a novel architecture for cross-domain foundation model (FM) integration. By systematically benchmark FMs across two domains – molecular FM for drugs and single-cell FM for cell lines, we demonstrate that integrating single-cell FMs substantially reduces the number of input features required for cell line representation. Among FMs, Molformer significantly outperforms ChemBERTa, and scGPT surpasses scFoundation in predictive accuracy and training stability. Moreover, integrating single-cell FMs improves performance in both drug-known and leave-one-drug-out scenarios. These findings highlight the potential of cross-domain FM integration for more efficient and robust drug response prediction.

1 Introduction

By leveraging large amounts of pre-training data, Foundation Models (FMs) can represent input in a context beyond the available dataset, making them effective feature extractors for various downstream tasks. In the biomedical domain, pre-trained FMs for medical images [1, 2], drug molecules [3, 4], single-cells [5, 6] and many other modalities are increasingly applied and systematically explored.

Being able to predict drug response is crucial for discovery and design of cancer treatment. The inherently multi-modal nature of the task requires the integration of properties from both drug molecules and cancer cells. Previous published work such as DeepCDR, GraphDRP and XGDP etc. mainly explored various CNN- and GNN- based approaches for extracting and integrating embeddings [7, 8, 9].

In this study, we investigate the potential of FMs for Cancer Drug Response Prediction (CDRP) and present, to our knowledge, the first attempt to integrate FMs across distinct domains. We introduce FUSED (**Fu**sion of Foundation Model **E**mbeddings for **D**rug Response Prediction), and systematically benchmark four representative FMs – scGPT and scFoundation for transcriptomic profiling [5, 6], and ChemBERTa and MolFormer [4, 3] for molecular structure representation. Our work provides a principled path for leveraging cross-domain FMs in robust drug-response prediction.

2 Related Work

Single-Cell Foundation Models scFMs are pre-trained on large corpora of single-cell gene expression data and have demonstrated capability of representing "virtual cells" and enabling diverse

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

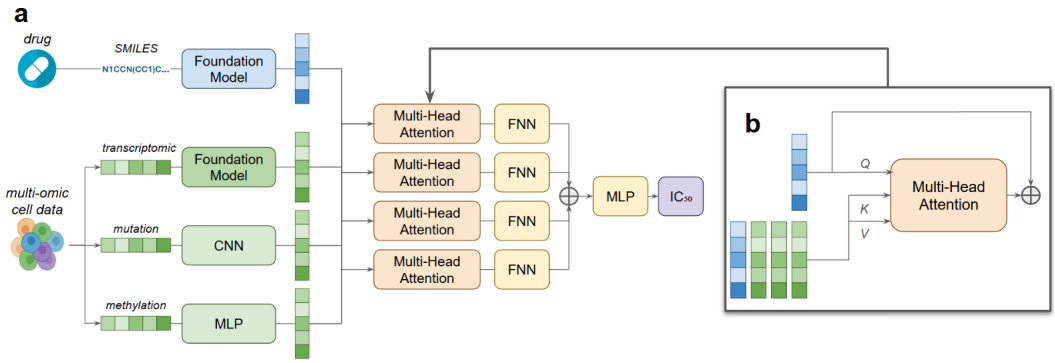

Figure 1: FUSED approach with FMs extracting drug features and transcriptomic features. **a)** We create 4 embeddings for each feature type. While transcriptomics features and drug features are transformed by a FM, mutation and methylation features are trained from scratch. Each embedding is contextualized with respect to the other three embeddings using multi-head attention. All representations are projected to equal dimensions with a linear layer, and concatenated and applied to a MLP to estimate IC_{50}. **b)** Multi-head attention mechanism: we create a Q-vector from the original embedding, and a K-, V-vector from all embeddings in order to compute attention score with respect to given Q-vector. Each multi-head attention block learn its own parameters to create Q-,K- and V-vectors.

downstream tasks. Among the recent state-of-the-art scFMs, scGPT [5] is a generative pre-trained transformer model built on ~33 million cells, and can effectively distill critical biological insights and achieve superb performance across multiple applications. scFoundation [6], another transformer-based scFM, adopts an asymmetric architecture with a pretraining objective designed to capture complex gene relations across a wide range of cell types and states.

Molecular Foundation Models Recent progress in molecular representation learning of chemical drugs has produced FMs that leverage different input modalities [10], including sequence-based models such as ChemBERTa and MolFormer [11, 3], 2D graph-based models such as MolE and KANO [12, 13], and 3D structure-based representations such as Uni-Mol and SubGDiff [14]. In this work, we focus on sequence-based models where drug molecules are represented as SMILES strings. ChemBERTa [11] is pre-trained on 77 million SMILES using masked language modeling to chemical tokens. MolFormer [3] extends this with rotary positional encoding and linear attention.

Drug Response Prediction A classical measure of cancer drug response is to estimate the drug efficacy by predicting the half-maximal inhibitory concentration (IC_{50}) for drug–cell line pairs. [15]. Most prior work employ hybrid neural network architectures where drug and cell features are encoded separately before their embeddings are integrated. While GraphDRP [8] systematically evaluates different graph neural network (GNN) architectures to encode drug structures, DeepCDR [7] focuses on integrating multi-omics information for cell line representation. More recently, XGDP [9] replaces simple embedding concatenation with a multi-head attention mechanism, yielding more expressive feature interactions than earlier CNN-based fusion.

3 Approach

We formulate cancer drug response (CDR) prediction as a regression task, where the model learns a mapping

$$f : (D, C) \rightarrow IC_{50},$$

with D representing drug features and C denoting cell features. Drug features are derived from SMILES representations, while C is decomposed into three distinct modalities: transcriptomic, methylation, and mutation profiles. The core idea is to generate embeddings for each of these four input types and integrate them to predict IC_{50} values. Our study proceeds in two steps: (1) baseline extension – incorporating FMs into the existing DeepCDR framework to generate drug and cell line embeddings; (2) develop novel integration architecture FUSED (**Fu**sion of Foundation Model **E**mbeddings for **D**rug Response Prediction), which leverages multi-head attention to integrate molecular and single-cell embeddings generated from domain-specific FMs (Fig. 1).

Datasets We use a combination of two common datasets: Cancer Cell Line Encyclopedia (CCLE) [16] and Genomics of Drug Sensitivity in Cancer (GDSC) [17]. CCLE stores multi-omics data (transcriptomics, methylation, mutation) for cancer cell lines. GDSC stores drug responses in form of IC_{50} values for cell line drug pairs, matching cell lines from CCLE database. In total, we have 561 cell lines paired with 223 drugs leading to 107446 drug-cell line pairs after filter out missing combinations.

Architecture Development Step 1: **Baseline Extension** Our baseline DeepCDR framework contains a graph convolutional network (GCN) and three subnetworks for drug structure and cancer cell profiles (genomic mutation, transcriptomics and methylation) respectively. In our extension, we replace either the GCN-based drug encoder with the molecular FM, or transcriptomic subnetworks with scFM. For the single-cell FM, we prepend it to the transcriptomic-specific subnetwork, while the molecular FM replace the GCN architecture. Step 2: **FUSED** We develop a novel architecture, FUSED, designed to enhance the integration of multiple FMs. Rather than relying solely on concatenation, FUSED employ multi-head attention to contextualize and align embeddings from different modalities before regression (Fig. 1). To optimize hyperparameters (*number of heads, number of MLP layers, learning rate, and embedding dimension*), we performed a grid search in preliminary experiments.

Experimental Configurations We used ChemBERTa or Molformer as FMs to extract drug embeddings, and scGPT or scFoundation to extract transcriptomic embeddings. In all settings, FM parameters were kept frozen to assess zero-shot capabilities. All experiments were conducted on an NVIDIA GTX 1080Ti GPU. The training time for the combined multi-omics and GCN model was ~150 minutes (100 epochs). When the GCN component was removed and frozen molecular FMs were used instead, the training time was reduced to ~70 minutes. FUSED required ~10 minutes more time than CNNs.

As primary evaluation metrics, we employed Pearson Correlation Coefficient (PCC, the higher the better) and Root Mean Squared Error (RMSE, the smaller the better). We explored two ways of splitting training/test data. In drug-known split, all drugs and cell lines were present in the training set, and the results are averaged across three runs with different seeds to avoid bias in test set. In leave-one-drug-out split, all samples associated with a given drug were placed exclusively in the test set, ensuring that the model had no prior exposure to that drug during training. This process was repeated for each of the 223 drugs, and the results were averaged across all iterations.

4 Results and Discussion

Table 1: Performance comparison in drug-known split. Top 3 models from each column are highlighted

Drug Encoder + Cell Encoder	**Transcriptomics**		**Multi-Omics**	
	PCC	**RMSE**	**PCC**	**RMSE**
GCN + Subnetworks (DeepCDR)	0.841 ± 0.003	1.483 ± 0.009	$\mathbf{0.920 \pm 0.002}$	$\mathbf{1.076 \pm 0.007}$
GCN + scFoundation	0.882 ± 0.005	1.299 ± 0.027	$\mathbf{0.920 \pm 0.003}$	1.076 ± 0.019
GCN + scGPT	$\mathbf{0.914 \pm 0.003}$	$\mathbf{1.115 \pm 0.015}$	$\mathbf{0.920 \pm 0.000}$	1.076 ± 0.004
ChemBERTa + Subnetworks	0.821 ± 0.002	1.598 ± 0.032	0.878 ± 0.003	1.326 ± 0.015
ChemBERTa + scFoundation	0.858 ± 0.006	1.440 ± 0.017	0.877 ± 0.008	1.322 ± 0.027
ChemBERTa + scGPT	0.876 ± 0.008	1.321 ± 0.034	0.881 ± 0.006	1.332 ± 0.037
MolFormer + Subnetworks	0.848 ± 0.009	1.454 ± 0.037	$\mathbf{0.919 \pm 0.002}$	1.081 ± 0.011
MolFormer + scFoundation	0.884 ± 0.003	1.308 ± 0.004	$\mathbf{0.919 \pm 0.002}$	1.089 ± 0.012
MolFormer + scGPT	$\mathbf{0.912 \pm 0.002}$	$\mathbf{1.125 \pm 0.013}$	$\mathbf{0.919 \pm 0.003}$	1.084 ± 0.018
FUSED - MolFormer + scGPT	$\mathbf{0.923 \pm 0.001}$	$\mathbf{1.103 \pm 0.094}$	$\mathbf{0.932 \pm 0.000}$	$\mathbf{0.993 \pm 0.006}$
FUSED - MolFormer + scFoundation	0.899 ± 0.005	1.141 ± 0.098	$\mathbf{0.932 \pm 0.001}$	$\mathbf{0.992 \pm 0.006}$

Drug-Known Split As shown in Table 1, when all drugs and cell lines appear in the training set, we observe following trends from both PCC and RMSE:

- **Multi-omics adds modest gains.** Incorporating multi-omics features yields improvements, particularly for models that are weaker with transcriptomics alone. This effect narrows the

performance gap among good and poor models, therefore most of our following comparisons will focus on the transcriptomics-only results.

- **scFM helps in transcriptomics only setting.** Both scGPT and scFoundation improve over the baseline, with scGPT consistently outperforming scFoundation in all experiments.
- **Molecular FMs differ.** ChemBERTa underperforms the GCN baseline, whereas MolFormer provides a slight improvement.
- **FUSED is the best.** Our FUSED approach achieves the best performance both in transcriptomic and multi-omics setting.

These findings underscore the value of FMs when comprehensive datasets (e.g., multi-omics) are unavailable. By leveraging knowledge acquired during pre-training, a FM-based approach can compensate for missing modalities. At the same time, FM choice matters: not all models confer benefits (e.g., ChemBERTa), and scGPT is more effective than scFoundation in our setting. Importantly, our FUSED approach demonstrates that integrating molecular and transcriptomic representations through an appropriate fusion mechanism can yield further performance gains, highlighting the promise of cross-domain FM integration for cancer drug response prediction.

Table 2: Leave-one-drug-out performance comparison of the different approaches

Drug Encoder + Cell Encoder	Transcriptomics		Multi-Omics	
	PCC	RMSE	PCC	RMSE
GCN + Subnetworks (DeepCDR)	0.076 ± 0.164	2.447 ± 1.352	0.449 ± 0.137	2.409 ± 1.443
GCN + scFoundation	0.433 ± 0.135	2.396 ± 1.440	0.443 ± 0.154	2.425 ± 1.359
GCN + scGPT	0.442 ± 0.145	2.447 ± 1.352	0.439 ± 0.147	2.402 ± 1.322
MolFormer + Subnetworks	0.107 ± 0.168	2.400 ± 1.455	$\mathbf{0.452 \pm 0.129}$	$\mathbf{2.349 \pm 1.406}$
MolFormer + scFoundation	0.429 ± 0.137	$\mathbf{2.323 \pm 1.396}$	0.447 ± 0.127	2.373 ± 1.395
MolFormer + scGPT	$\mathbf{0.447 \pm 0.125}$	2.375 ± 1.396	0.448 ± 0.130	2.355 ± 1.418
FUSED - MolFormer + scGPT	0.431 ± 0.148	2.444 ± 1.411	0.434 ± 0.146	2.432 ± 1.358

Leave-One-Drug-Out In the challenging drug-blind setting (Table 2), baseline without scFM performs extremely poorly when only transcriptomic features are used: PCC = 0.076 with GCN and 0.107 with MolFormer as the drug encoder. Interestingly this weakness can be mitigated by adding other omics or using a scFM, which raises PCC into the 0.41-0.45 range and reduces RMSE. Our FUSED approach does not yield additional gains in this setting, indicating the possibility of overfitting to the training drugs.

Different than the drug-known setting, PCC and RMSE are not always aligned, and standard deviations are relatively large because the summary statistics are averaged over 223 held-out drugs. This variance indicates that performance depends strongly on which drug is withheld, suggesting that drug-specific properties substantially affect model generalization.

Overall, the weak zero-shot generalization observed in drug-blind setting highlights a key limitation of current AI models for drug response prediction and motivates future work on stronger drug encoders, regularization, and augmentation strategies.

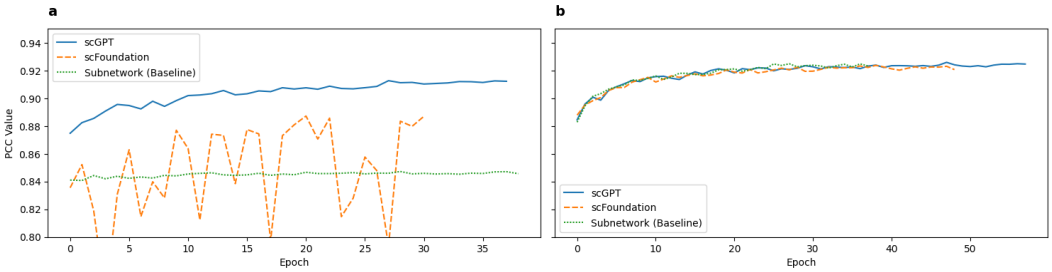

Figure 2: **Representative learning curves** PCC validation performance in drug-known setting with MolFormer as drug encoder. Baseline is transcriptomic subnetwork. a) transcriptomics only; b) multi-omics.

Training Stability We examine the validation PCC over training epochs in the drug-known setting (Figure 2). In the transcriptomics-only condition (Figure 2a), scFoundation exhibits pronounced instability—large oscillations and occasional collapses—across runs, whereas scGPT converges quickly and monotonically and the non-FM baseline improves only marginally. Adding additional omics (Figure 2b) markedly stabilizes training: the oscillations diminish and scFoundation's final performance approaches that of scGPT. These results indicate that scGPT-derived embeddings are robust in low-feature transcriptomic settings, while multi-omics integration mitigates scFoundation's instability and enables competitive accuracy.

5 Conclusion

We introduce FUSED, a compact fusion architecture that aligns molecular and single-cell foundation-model (FM) embeddings, and provide the first systematic cross-domain FM benchmark for cancer drug-response prediction. Across settings, scGPT and MolFormer emerge as the strongest single-modality encoders, and FUSED attains the best accuracy in drug-known splits while reducing reliance on extensive multi-omics features. Limitations include weak generalization to unseen drugs, instability for some scFMs in transcriptomics-only training, and evaluation restricted to frozen encoders on CCLE/GDSC cell-line data. Future work could focus on stronger drug encoders, fusion schemes explicitly regularized for zero-shot drugs, and validation beyond cell lines (e.g. patient-derived models).

Code Availability

The source code for models and experiments will be made publicly available after acceptance of the paper. We share an anonymized github repository at `https://github.com/anon-neurips2025/FUSED`.

References

[1] Richard J. Chen, Tong Ding, Ming Y. Lu, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30:850–862, 2024.

[2] Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30:863–874, 2024.

[3] Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, et al. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4:1256–1264, 2022.

[4] Walid Ahmad, Elana Simon, Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta-2: Towards chemical foundation models, 2022.

[5] Haotian Cui, Chloe Wang, Hassaan Maan, et al. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21:1470–1480, 2024.

[6] Minsheng Hao, Jing Gong, Xin Zeng, et al. Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, 21:1481–1491, 2024.

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

[8] Tuan Nguyen, Giang T. T. Nguyen, Thin Nguyen, and Duc-Hau Le. Graph convolutional networks for drug response prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(1):146–154, 2022.

[9] Conghao Wang, Gaurav A. Kumar, and Jagath C. Rajapakse. Drug discovery and mechanism prediction with explainable graph neural networks. *Scientific Reports*, 15:179, 2025.

[10] Jianbo Qiao, Junru Jin, Ding Wang, et al. A self-conformation-aware pre-training framework for molecular property prediction with substructure interpretability. *Nature Communications*, 16:4382, 2025.

[11] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta: Large-scale self-supervised pretraining for molecular property prediction, 2020.

[12] Oscar Méndez-Lucio, Christos A. Nicolaou, and Berton Earnshaw. Mole: a foundation model for molecular graphs using disentangled attention. *Nature Communications*, 15:9431, 2024.

[13] Yin Fang, Qiang Zhang, Ningyu Zhang, et al. Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nature Machine Intelligence*, 5:542–553, 2023.

[14] Jiying Zhang, Zijing Liu, Yu Wang, and Yu Li. Subgdiff: A subgraph diffusion model to improve molecular representation learning, 2024.

[15] Senem Aykul and Erik Martinez-Hackert. Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. *Analytical Biochemistry*, 508:97–103, 2016.

[16] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483:603–607, 2012.

[17] Francesco Iorio, Theo A. Knijnenburg, Daniel J. Vis, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*, 166(3):740–754, 2016.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims outlined in the abstract and introduction accurately reflect the paper's contributions and scope, particularly regarding the introduction of FUSED, the systematic evaluation of foundation models, and the insights gained from cross-domain integration efficiency and performance.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper explicitly discusses limitations in the "Results and Discussion" and "Conclusion" sections . These include weak generalization to unseen drugs in the "leave-one-drug-out" scenario, the instability of scFoundation in transcriptomics-only settings, and the restriction of evaluation to cell line data (CCLE/GDSC) using frozen encoders.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This work does not include theoretical results or formal proofs, focusing instead on empirical and methodological contributions to drug response prediction .

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides all necessary information to reproduce the main results. The FUSED architecture is detailed in the "Approach" chapter, and experimental configurations are described in the "Experimental Configurations" section. Data preprocessing steps were adapted from established scGPT/scFoundation work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide an anonymized repository containing the full code base, which is linked in our paper. All utilized datasets (CCLE and GDSC) are publicly accessible and properly cited. Minor parameter adjustments may be required for scripts.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The "Experimental Configurations" chapter provides a comprehensive overview of the experimental settings, including data splits (drug-known, leave-one-drug-out) and metrics (PCC, RMSE). All detailed hyperparameters are available in the codebase.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Tables 1 and 2 present the standard deviations of the performance metrics (PCC and RMSE). The computational setup is described in detail in the Experimental Configurations chapter.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper states in the "experimental Configurations" chapter that all experiments were performed on an NVIDIA GTX 1080Ti GPU. Training times for various configurations are also given, for instance, 150 minutes for the GCN model.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research transparently conforms to ethical guidelines by explicitly stating its preclinical nature and non-clinical applicability, addressing potential biases from limited data diversity, and acknowledging interpretability challenges inherent to the methodology. It is presented as a methodological contribution requiring further validation before any clinical use.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper discusses positive societal impacts, such as advancing personalized cancer therapy and reducing animal experiments. It also addresses negative impacts, including the risk of misapplication of preclinical results without further validation, potential biases from limited data diversity, and interpretability challenges that could affect trustworthiness in real-world deployment.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This work does not release trained models or self-generated datasets with a high risk of misuse. All utilized datasets (CCLE, GDSC) are already publicly available.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: Both datasets used, GDSC and CCLE, are properly cited and publicly available. The DeepCDR framework, used as a baseline and for parts of the code, is also mentioned multiple times.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The newly developed FUSED framework as well as the conducted experiments and their associated code are provided in an anonymized GitHub repository, which includes documentation (Readme) for these assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This work does not involve crowdsourcing experiments or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: As this work does not involve research with human subjects, IRB approval or equivalent review is not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: LLMs were exclusively used for improving text quality and editorial purposes, and did not influence the core methodology, scientific rigorousness, or originality of the research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.