nature portfolio

Peer Review File

Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery

Reviewer #1 (Remarks to the Author):

In their work “Predicting transcriptional responses to novel chemical perturbations using deep generative model,” Qi et al present a “perturbation-coded deep generative model” named PRnet, capable of predicting changes in cell type-specific bulk and single-cell gene expression under novel chemical perturbations across multiple doses. They also identify and experimentally test novel compound candidates, and generate a “large-scale integration atlas of perturbation profiles”. This work is very interesting and novel in terms of: i) experimental validation of candidate drugs; and ii) prediction of responses for out-of-sample chemicals. The manuscript is also clearly written and easy to follow.

I have some methodological concerns (designated by ** below), and some editorial suggestions.

** 1. I would suggest that the authors emphasize the main novelty of their work early on in the introduction: the fact that, of the many cell perturbation modeling tools that have been published recently, few predict responses for novel chemicals. (chemCPA is an exception [https://arxiv.org/abs/2204.13545], which the authors should cite).

** 2. It’s not clear from the introduction whether the authors are proposing a completely new neural network architecture, or are using a standard architecture like an autoencoder / GAN. “Perturbation-conditioned deep generative model” is not a standard term, and the authors need to clarify early on what they mean by this. After reading into the details of the text, I gathered that it is similar to a variational autoencoder, with an extra pre-processing step to incorporate chemical structures (the “Perturb-adaptor”). If this is correct, then it is misleading to contrast their work with “Auto-encoder-based models” in the Introduction (lines 55-61).

** 3. It is not correct to state that auto-encoder-based models are “incapable of modeling … novel cell types”. Both scGen [PMID: 31363220] and scVIDR [PMID: 37602218] accomplish precisely this task.

4. The sentence in lines 61-64 (“Linear regression based methods …”) is unclear and needs to be rephrased.

5. Likewise the sentence in lines 75-78 (“Consequently, perturbations response models are required to…”) isn’t quite clear. It may be best to split it into two sentences.

6. In lines 107-108, the authors should clarify that the “large integration atlas of perturbation profiles” they generated is virtual, not an actual experimental data set.

7. The sentence in lines 116-119 is not clear. Needs to be rephrased.

8. Line 149: What is “rFCFP”? This is not a standard term, and needs to be defined.

9. Line 169 should read: “This allows the model to directly *generalize* to…”

10. Line 177 should read: “In step 2, (we) calculate…”

11. Line 209: “…all datasets were strictly split (by) perturbation attributes..”

12. Line 225: “978 landmark genes” – it needs to be clarified what “landmark genes” are.

13. Line 243: The “Pearson of log(FC) in cov compounds” metric needs to be defined.

14. From the text in Lines 253-265, it is not clear what the difference between Fig 2C and 2D is: is it simply all the cell types vs. some of the cell types shown?

15. Lines 275-277: the text should explain the difference between “R2 in compound” and “R2 in cov compound” metrics.

16. Lines 284-287: please check if the labels to panels Fig. 2 F and Fig. 2 G are correct.

17. Line 289 should read: “… consensus trajectory, in which…”

18. Line 290: “the pseudodose trajectory…” – the concept of a “pseudo-dose” needs to be explained. How is it calculated?

19. The authors point out that PRnet demonstrates the similarity of perturbations for *similar* cell types. It would be helpful if this can be demonstrated by a quantitative metric, e.g. cosine similarity of the perturbation vectors in latent space.

** 20. The basic assumption used in this work for prediction of perturbations in gene expression was not clear to me. Is the perturbation vector assumed to be a function of chemical structure and unperturbed gene expression state, or only of chemical structure? What is the nature of this functional dependence?

21. Lines 594-595 should read: “… novel chemical perturbations that were never experimentally *studied* (or *carried out*) at bulk and single-cell levels.”

22. Line 702: “Training” is mis-spelt.

Reviewer #2 (Remarks to the Author):

In this article, Qi et al built a deep learning model named PRnet to predict chemical perturbation of transciptional perturbation in different cellular contexts. The model was trained on several data sets including bulk and single-cell RNA-seq in response to FDA and natural chemical compound libraries, using Simplified Molecular Input Line Entry System (SMILES) to represent compound structures. Using the model, the author performed in silico screens and validated some of the findings by in vitro assays and literature reviews, including various cell lines, lung cancer, colon cancer, fatty liver, and inflammatory bowel diseases.

Overall the study is well performed and comprehensive, which can be a useful resource for the broader community. Below are my comments:
1. The impact of the paper will be enhanced if the author provides a way for online search of their results and databases and testing new compound using their AI model. At least articulate a clear way of sharing all the datasets to the broader readership. I do not think the paper should be published without the authors' commitment to share the AI model and all the results in a straightforward way for broader dissemination.
2. The paper needs to explain how good SMILES is for representing the compound structure and what the tradeoffs and alterantives are.
3. For the disease examples such as NASH, Crohn's and PCOS, the author need to list the ranked drug predictions in tables in extended dataset. Currently they only referred to a figure for each disease that highlighted a couple of handpicked targets that they had literature evidence, but it is impossible (at least for me) to read the entire prediction list to assess the quality of the predictions and whether it predicts promising new compounds. Also, for the few handpicked examples for the various diseases, the authors need to include more data from the in silico screening to show how the model predicted transcriptional perturbation (what were the pathways/genes perturbed? In heatmap and GSEA), what exactly are the enrichment scores, how the predictions were ranked, etc.

Reviewer #3 (Remarks to the Author):

Qi et al. presents the development of PRnet, a perturbation-conditioned deep generative model designed to predict transcriptional responses to novel chemical perturbations. The manuscript highlights the model's outperformance of alternative methods in predicting responses across novel compounds, pathways, and cell lines, enabling gene-level response interpretation and in-silico drug screening for diseases based on gene signatures. Additionally, PRnet has been utilized to identify and experimentally test novel compound candidates against small cell lung cancer and colorectal cancer, and has generated a large-scale integration atlas of perturbation profiles, covering 88 cell lines and 52 tissues perturbed by various screening compound libraries.

The manuscript is clear and well-written. Authors present an useful and novel tool that can be used in different biological, clinical and preclinical contexts including the drug discovery field. However I have some questions and suggestions that authors should consider before publication:

Major points

1) Authors present the generation of a large-scale integration atlas of perturbation profiles and the implementation of a robust and scalable drug recommendation workflow based on the profiles atlas. This large-scale atlas of perturbations predicts 25M of transcriptional responses for different drug libraries and datasets. However, those are limited to cell lines and the GTEx dataset. I am wondering if it is possible to include bulk or single cell data from patients tissues? In my opinion, authors should include some results (ex: TCGA cohort, single-cell human cell atlas) to prove the application in patient data. In fact, there are some studies in cancer of pre-post treatment scRNA-seq data: is it possible to validate if the predicted transcriptional profile

by PRNet is similar to the post-treatment patient samples? This could be interesting for instance to predict the patient state to propose a combination or second-line treatment.

2) Authors should specify in the manuscript where the large-scale atlas of perturbations profiles is stored/included. It is not clear to me if they develop a resource that users can query to check perturbations or cell lines and find gene signatures associated with the perturbation. I think creating this resource is essential and will greatly improve the manuscript.

3) It is described the importance of capturing cell-type specific response and cellular heterogeneity which are relevant in drug response. However, the manuscript is focused on capturing the heterogeneous response between cell lines and not within the same cell-type which is a known-event (Ben-David et al. Nature). I think, it would be interesting to describe and study the inter and intra-heterogeneity between cell lines. Can be calculated the relationship between heterogeneity with the predicted perturbated transcriptional profile? Do homogeneous cells have more "sensitivity" than heterogeneous? I would expect that some responses even from different cell lines are more similar. Following the drug recommendation workflow, it would be nice to know if there are more recommended drugs for specific cancer cell types, based on heterogeneity, etc.

4) Fig 2G "MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses." It is not clear why is a homogeneous response. Please explain in more detail this point.

5) My main concern is about considering a compound sensitive which is employed to identify potential compound candidates in SCLC and CRC. Here, you are assuming that perturbated transcriptional profile are associated with sensitivity in some diseases which is not necessarily true although this idea was presented a while ago by Lamb et al. 2006 Science, resulting in projects such as Connectivity Map, L1000, etc. For the drug recommendation workflow: are you calculating the enrichment score employing a similar assumption of reverse signature paradigm established by Lamb et al?. Authors should describe in more detail these previous works and how they can compare it with previous results/studies that followed this approach. So, can the authors include information about cancer cell lines drug sensitivity screenings such as GDSC, CTRP, CCLE, etc to validate this point?

6) In my opinion discussion and results could be more detailed. The results section presents the performance evaluation of PRnet for predicting responses to unseen perturbations in human diseases (cancer, NASH, etc). However, I miss more applications in patient data as I commented before. The discussion section should describe in more detail PRNet limitations which are not properly addressed.

7) More documentation and good demos to run the tool available in Github are needed:
- The installation is simple, but for someone without basic Python knowledge it might lack details.
- There is no documentation on how to run the tool, there are several separate scripts and it is not intuitive.
- The demos cannot be tested and more details are needed: The dataset files are missing. It is not clear where they come from. Also, within the scripts they are named completely different

from the reference.
- Also specify if these files need to be preprocessed before using the demos.

8) Figure 1C and Figure 4A/ Figure 5b show similar concepts/results. In fact, Figure 1C could be considered as a graphical abstract. I suggest including some sort of summary of "the atlas of perturbated profiles" or a summary of the resource which I think it is essential to make it public.

Minor points

Line 291: You might want to mention figure 2H.
Line 313: "Fig 3b investigated in detail .." Please rephrase the sentence figures do not investigate..
Paragraphs in line 400 and line 436. Please include IC50 etc in a supplementary table to make the reading easier.

Reviewer #3 (Remarks on code availability):

More documentation and good demos to run the tool available in Github are needed:
- The installation is simple, but for someone without basic Python knowledge it might lack details.
- There is no documentation on how to run the tool, there are several separate scripts and it is not intuitive.
- The demos cannot be tested and more details are needed: The dataset files are missing. It is not clear where they come from. Also, within the scripts they are named completely different from the reference.
- Also specify if these files need to be preprocessed before using the demos.

Reviewers comments:

Reviewer #1 (Remarks to the Author):

In their work “Predicting transcriptional responses to novel chemical perturbations using deep generative model,” Qi et al present a “perturbation-coded deep generative model” named PRnet, capable of predicting changes in cell type-specific bulk and single-cell gene expression under novel chemical perturbations across multiple doses. They also identify and experimentally test novel compound candidates, and generate a “large-scale integration atlas of perturbation profiles”. This work is very interesting and novel in terms of: i) experimental validation of candidate drugs; and ii) prediction of responses for out-of-sample chemicals. The manuscript is also clearly written and easy to follow.

I have some methodological concerns (designated by ** below), and some editorial suggestions.

Reply: Thank you very much for your positive comments. We sincerely appreciate your constructive comments and valuable suggestions regarding our manuscript. We have incorporated these suggestions into the revised manuscript accordingly as listed in details below.

** 1. I would suggest that the authors emphasize the main novelty of their work early on in the introduction: the fact that, of the many cell perturbation modeling tools that have been published recently, few predict responses for novel chemicals. (chemCPA is an exception [https://arxiv.org/abs/2204.13545], which the authors should cite).

Reply 1-1: Thank you for your valuable suggestion and reminding us of the chemCPA work. We have revised the introduction to emphasize the main novelty of our work: “These recently published cell perturbation modeling tools precisely simulated chemical perturbations and predicted chemical perturbations in gene expression of

unseen cell types, but few predict responses to novel chemicals". We have also cited chemCPA, which introduced a new encoder-decoder architecture which incorporated the compounds' structure to predict the perturbational effects of unseen drugs. Furthermore, we also cited scVIDR which was mentioned in the following questions which is capable of predicting chemical perturbations in gene expression across cell types. At last, combined with the following questions to make the manuscript more understandable, we have reorganized related works in the field, and corrected the original mischaracterization (see details lines 65-77 on pages 2-3, in the revised manuscript and copied below).

Initial version:

Auto-encoder-based models [4–6] mapped chemical induce transcriptomic effect into a latent space to reconstruct perturbation response.

Auto-encoder-based and optimal-transport-based models effectively reconstructed or matched the observations experimentally perturbed, but were incapable of modeling novel perturbations, such as novel compounds or novel cell types.

Revised version:

lines 65-77 on pages 2-3

CPA [4] utilized auto-encoder-based model to map chemical induce transcriptomic effect into a latent space to reconstruct perturbation response. Biolord [5], scGen [6] and scVIDR [7] performed counterfactual prediction via deep encoder-decoder/generator based generative frameworks, which take a control cell and unseen labels as input to predict the gene expression of the unseen cellular states. These recently published cell perturbation modeling tools precisely simulated chemical perturbations and predicted chemical perturbations in gene expression of unseen cell types, but few predict responses to novel chemicals. Following CPA [4], chemCPA [8] introduced a new encoder-decoder architecture which incorporated the compounds' structure to predict the perturbational effects of unseen drugs. Deep generative frameworks with encoder-decoder and its variants effectively predicted

single-cell gene expression perturbations.

** 2. It's not clear from the introduction whether the authors are proposing a completely new neural network architecture, or are using a standard architecture like an autoencoder / GAN. "Perturbation-conditioned deep generative model" is not a standard term, and the authors need to clarify early on what they mean by this. After reading into the details of the text, I gathered that it is similar to a variational autoencoder, with an extra pre-processing step to incorporate chemical structures (the "Perturb-adaptor"). If this is correct, then it is misleading to contrast their work with "Auto-encoder-based models" in the Introduction (lines 55-61).

Reply 1-2: Thank you for your advice and we apologize for any confusion caused by the lack of clarity in describing our model. Autoencoders and it's variation VAE are a special type of unsupervised feedforward neural network which learn efficient representations of input data. These models utilized an encoder-decoder architecture to reconstruct input. However, PRnet is not an unsupervised model. The input (unperturbed transcriptional profile) and output (unperturbed transcriptional profile) of PRnet are different. Inspired by the ability to learn efficient representations of encoder-decoder architecture, we introduce a new encoder-decoder architecture based generative model which comprises three components: the Perturb-adaptor, the Perturb-encoder, and the Perturb-decoder, named PRnet. We have revised the description of our model (lines 113-118 on pages 3-4) and corrected description about auto-encoder-based models in the introduction section (lines 65-77 on pages 2-3).

Initial version:

Here we present PRnet, which is a flexible and scalable perturbation-conditioned deep generative model for predicting transcriptional responses to novel chemical perturbations that were never experimentally perturbed at bulk and single-cell levels.

Revised version:

lines 113-118 on pages 3-4

Here we present PRnet, which is a flexible and scalable perturbation-conditioned deep generative model for predicting transcriptional responses to novel chemical perturbations that were never experimentally perturbed at bulk and single-cell levels. PRnet is a new encoder-decoder architecture based generative model which comprises three components, including the Perturb-adaptor, the Perturb-encoder, and the Perturb-decoder.

Initial version:

Auto-encoder-based models [4–6] mapped chemical induce transcriptomic effect into a latent space to reconstruct perturbation response.

Auto-encoder-based and optimal-transport-based models effectively reconstructed or matched the observations experimentally perturbed, but were incapable of modeling novel perturbations, such as novel compounds or novel cell types.

Revised version:

lines 65-77 on pages 2-3

CPA [4] utilized auto-encoder-based model to map chemical induce transcriptomic effect into a latent space to reconstruct perturbation response. Biolord [5], scGen [6] and scVIDR [7] performed counterfactual prediction via deep encoder-decoder/generator based generative frameworks, which take a control cell and unseen labels as input to predict the gene expression of the unseen cellular states. These recently published cell perturbation modeling tools precisely simulated chemical perturbations and predicted chemical perturbations in gene expression of unseen cell types, but few predict responses to novel chemicals. Following CPA [4], chemCPA [8] introduced a new encoder-decoder architecture which incorporated the compounds' structure to predict the perturbational effects of unseen drugs. Deep generative frameworks with encoder-decoder and its variants effectively predicted single-cell gene expression perturbations.

** 3. It is not correct to state that auto-encoder-based models are "incapable of modeling … novel cell types". Both scGen [PMID: 31363220] and scVIDR [PMID: 37602218] accomplish precisely this task.

Reply 1-3: Thank you for pointing out our incorrect statement in the discussion of related work. We have revised the statement to accurately reflect that autoencoder-based models, including biolord, scGen and scVIDR, are capable of precisely modeling novel cell types (67-73 on pages 2-3).

Initial version:
Auto-encoder-based and optimal-transport-based models effectively reconstructed or matched the observations experimentally perturbed, but were incapable of modeling novel perturbations, such as novel compounds or novel cell types.
Revised version:
lines 67-73 on pages 2-3
Biolord [5], scGen [6] and scVIDR [7] performed counterfactual prediction via deep encoder-decoder/generator based generative frameworks, which take a control cell and unseen labels as input to predict the gene expression of the unseen cellular states. These recently published cell perturbation modeling tools have precisely simulated chemical perturbations and predicted chemical perturbations in gene expression of unseen cell types, but few predict responses to novel chemicals.

4. The sentence in lines 61-64 ("Linear regression based methods …") is unclear and needs to be rephrased.

Reply 1-4: Thank you for your advice for clarification the sentence. We have rephrased the sentence (lines 81-85 on page 3 in the revised manuscript, copied below).

Initial version:
Linear regression based methods [11, 12] linearly combined the effects of genetic perturbation cause the limitation in capturing nonlinear chemical perturbations of

diverse cell type and compound combinations.

Revised version:

lines 81-85 on page 3

Linear regression based methods [11, 12] estimate the impact of perturbations on gene expression by linearly combining the effects of genetic perturbation. However, this linear combination approach leads to limitations in accurately modeling the nonlinear chemical perturbations across diverse cell types and compound combinations.

5. Likewise the sentence in lines 75-78 (“Consequently, perturbations response models are required to…”) isn’t quite clear. It may be best to split it into two sentences.

Reply 1-5: Thank you for your advice. We have split the sentence for clarity (lines 101-105 on page 3 in the revised manuscript, copied below).

Initial version:

Consequently, perturbations response models are required to address the limited exploration power to novel perturbations of existing experimental and computational methods and discover promising therapeutic drug candidates.

Revised version:

lines 101-105 on page 3

Consequently, perturbations response models are required to address the limited exploration power to novel perturbations of existing experimental and computational method, which are also needed for predicting the response to unseen perturbations and discovering promising therapeutic drug candidates.

6. In lines 107-108, the authors should clarify that the “large integration atlas of perturbation profiles” they generated is virtual, not an actual experimental data set.

Reply 1-6: Thank you for your advice. We have made the clarification in the revised manuscript (lines 138-142 on page 4, copied below).

Initial version:

We therefore leveraged PRnet to in silico screen various compounds libraries and generated a large integration atlas of perturbation profiles, covering 88 cell lines and 52 tissues, and compounds libraries comprising 935 FDA-approved drugs, 4,158 active compounds, 30,456 natural compounds, and 29,670 drug-like compounds.

Revised version:

lines 138-142 on page 4

We therefore leveraged PRnet to *in silico* screen various compounds libraries and generated a virtual large integration atlas of perturbation profiles, covering 88 cell lines and 52 tissues, and compounds libraries comprising 935 FDA-approved drugs, 4,158 active compounds, 30,456 natural compounds, and 29,670 drug-like compounds.

7. The sentence in lines 116-119 is not clear. Needs to be rephrased.

Reply 1-7: Thank you for your advice. We have rephrased the sentence (lines 148-152 on page 4 in the revised manuscript, copied below).

Initial version:

Take three metabolic disorders as examples, including nonalcoholic steatohepatitis (NASH), polycystic ovary syndrome patients (PCOS), and inflammatory bowel disease (IBD), PRnet-recomand drugs have been verified by previous literature with human or animal studies.

Revised version:

lines 148-152 on page 4

In three cases of metabolic disorders, including non-alcoholic steatohepatitis (NASH), polycystic ovary syndrome (PCOS) patients and inflammatory bowel disease (IBD), drugs recommended by PRnet have been supported by previous literature with human or animal studies.

8. Line 149: What is “rFCFP”? This is not a standard term, and needs to be defined.

Reply 1-8: Thank you for your advice. We apologized for the missing definition of "rFCFP" in the main text. "rFCFP" refers to scaled fingerprints generated by scaling the Functional-Class Fingerprints (FCFP) with their corresponding dosages. We have rephrased sentences in sections of Results and Methods and added the definition of “rFCFP” (lines 179-186 on page 5 and lines 789-792 on page 19 in the revised manuscript, copied below).

Initial version in the Results section:

Initially, given a chemical perturbation imposed by the structure of compounds represented by Simplified Molecular Input Line Entry System (SMILES) strings and the dosages of the corresponding compounds, PRnet leverages RDKit [24] to capture the functional topology information of the structures, and scales them with their dosages to generate fingerprints $rFCFP$.

Revised version in the Results section:

lines 179-186 on page 5

Initially, given a chemical perturbation imposed by the structure of compounds represented by Simplified Molecular Input Line Entry System [30] (SMILES) strings and the dosages of the corresponding compounds, PRnet leverages RDKit [31] to capture the functional topology information of the structures, generate the Functional-Class Fingerprints ($FCFP$) of compounds. These $FCFP$s were scaled by their dosages and then summed to generate rescaled Functional-Class Fingerprints ($rFCFP$) embedding.

Initial version in the Methods section:

Then, the fingerprint embedding $rFCFP_i$ of P_i was calculated by the weighted sum of all fingerprint embeddings from all compounds applied by P_i:

$$rFCFP_i = \sum_j \phi(d_{i,j}) \times H(s_{i,j}) = \sum_j log_{10}(d_{i,j} + 1) \times H(s_{i,j}),$$

where ϕ is a log scale function using the dosage of j-th compound.

Revised version in the Methods section:

lines 789-792 on page 19

Then, the rescaled Functional-Class Fingerprints $rFCFP_i$ embedding of P_i was calculated by the weighted sum of all fingerprint embeddings of all compounds applied by P_i:

$$rFCFP_i = \sum_j \phi(d_{i,j}) \times H(s_{i,j}) = \sum_j log_{10}(d_{i,j} + 1) \times H(s_{i,j}),$$

where ϕ is a log scale function using the dosage of j-th compound.

9. Line 169 should read: "This allows the model to directly *generalize* to…"

Reply 1-9: Thank you for your advice. We have revised the sentence (see lines 207-210 on page 6 in the revised manuscript, copied below).

Initial version:

This allows the model to directly generalization to novel perturbation scenarios involving novel compounds, pathways, cell types, and cell lines that have not been previously perturbed.

Revised version:

lines 207-210 on page 6

This allows the model to directly generalize to novel perturbation scenarios involving novel compounds, pathways, cell types, and cell lines that have not been previously perturbed.

10. Line 177 should read: "In step 2, (we) calculate…"

Reply 1-10: Thank you for your advice. We have revised the sentence (see page 6, lines 219-221 in the revised manuscript, copied below).

Initial version:

In step 2, calculate the average transcriptional profile, fold-change in gene expression, and gene ranking after perturbing of each compound.

Revised version:

lines 219-221 on page 6

In step 2, we calculate the average transcriptional profile, and fold-change in gene expression of each compound, and rank genes according to their fold-change values.

11. Line 209: "…all datasets were strictly split (by) perturbation attributes.."

Reply 1-11: Thank you for your advice. We have revised the sentence (see lines 250-253 on page 7 in the revised manuscript, copied below).

Initial version:

To evaluate the performance of PRnet for predicting responses to unseen perturbations, all datasets were strictly split perturbation attributes (*compound_split*, *cell_line_split* and *pathway_split*) into 3 subsets: train, validation, and test (Supplementary Fig. 1 A).

Revised version:

lines 250-253 on page 7

To evaluate the performance of PRnet for predicting responses to unseen perturbations, all datasets were strictly split by perturbation attributes (*compound_split*, *cell_line_split* and *pathway_split*) into 3 subsets: train, validation, and test (Supplementary Fig. 1 A).

12. Line 225: "978 landmark genes" – it needs to be clarified what "landmark genes" are.

Reply 1-12: We apologize for the missing definition of the landmark genes. In the L1000 project, 978 genes were selected to represent the diversity of biological pathways

and processes in human cells. These representative genes were referred to as the landmark genes. We have added the description about the 978 landmark genes in the revised manuscript (line 266-272 on page 7, copied below).

Initial version:

After data preprocessing (as detailed in 3. Methods), we obtained 836,352 paired bulk RNA-seq observations with 978 landmark genes from 82 cell lines perturbed by 175,549 compounds [2].

Revised version:

line 266-272 on page 7:

We employed bulk high-throughput screening data from the L1000 project [1] to fit the model, in which 978 genes (hereinafter called landmark genes) were selected to represent the diversity of biological pathways and processes in human cells. We first preprocessed these data (as detailed in 4. Methods), and obtained 836,352 paired bulk RNA-seq observations (represented by the expression levels of 978 landmark genes), covering 82 cell lines and their perturbation data perturbated by 175,549 compounds.

13. Line 243: The "Pearson of log(FC) in cov compounds" metric needs to be defined.

Reply 1-13: Thank you for your advice. The "Pearson of log(FC) in cov_compounds" is the Pearson correlation coefficient between the true and predicted mean log(FC) perturbed by the same compound within the same cell line in the test set. The formula has been added. We have added the definition of "Pearson of log(FC) in cov_compounds" in the Results section (lines 286-293 on pages 7-8) and calculation formula in the Methods section (lines 865-894 on pages 21-23).

Initial version in the Results section:

In more challenging scenarios, we evaluated the performances of predicting cell line specific compound induced changes in genes, that is the log(FC) of compounds within the same cell line in the test set. PRnet achieved the best performance in the

"Pearson of log(FC) in cov_compounds" metric across three scenarios.

Revised version in the Results section:

lines 286-293 on pages 7-8

In more challenging scenarios, we evaluated the performances of predicting cell-line-specific compound-induced changes in genes by "Pearson of log(FC) in cov_compounds" metric. The "Pearson of log(FC) in cov_compounds" metric is the Pearson correlation between the true and predicted mean log(FC) perturbed by the same compound within the same cell line in the test set. PRnet achieved the best performance in the "Pearson of log(FC) in cov_compounds" metric across three scenarios.

Revised version in the Methods section:

lines 865-894 on pages 21-23

3. Pearson of log(FC) in compounds: evaluates the Pearson correlation between the true and predicted post-perturbation of the average logarithm of the fold-change in gene expression (log(FC)) perturbed by the same compound:

$$\overline{\log(fc_j)} = \frac{1}{n}\sum_{i=1}^{n} \log(fc_{i,j}),$$

$$\overline{\log(\widehat{fc_j})} = \frac{1}{n}\sum_{i=1}^{n} \log(\widehat{fc_{i,j}}),$$

$$PCC_j = PCC\left(\overline{\log(fc_j)}, \overline{\log(\widehat{fc_j})}\right),$$

$$\text{Pearson of log(FC) in compounds} = \frac{1}{m}\sum_{j=1}^{m} PCC_j,$$

where $fc_{i,j}$ and $\widehat{fc_{i,j}}$ denote as the ground truth and predicted post-perturbation fold change in gene expression x_i perturbed by compound j. The set $(fc_{i,j})_{i=1}^{n}$ represents all perturbed fold-changes for compound j. $\overline{\log(fc_j)}$ and $\overline{\log(\widehat{fc_j})}$ are the ground truth and predicted post-perturbation average logarithm of logarithm of

the fold-change in gene expression for compound j, respectively. PCC_j is the Pearson correlation coefficient of the mean log(FC) perturbed by compounds j. The average of the Pearson correlations for all m compounds in the test set is referred to as the "Pearson of log(FC) in compounds". A higher value of "Pearson of log(FC) in compounds" indicates a better performance.

4. Pearson of log(FC) in cov_compounds: evaluates the Pearson correlation between the true and predicted post-perturbation of the average logarithm of the fold-change in gene expression (log(FC)) perturbed by the same compound within the same cell line:

$$\overline{\log\left(FC_j^c\right)} = \frac{1}{n}\sum_{i=1}^{n} \log\left(fc_{i,j}^c\right),$$

$$\overline{\log\left(\widehat{FC_j^c}\right)} = \frac{1}{n}\sum_{i=1}^{n} \log\left(\widehat{fc_{i,j}^c}\right),$$

$$PCC_j^c = PCC\left(\overline{\log\left(FC_j^c\right)}, \overline{\log\left(\widehat{FC_j^c}\right)}\right),$$

$$\text{Pearson of log(FC) in cov\_compounds} = \frac{1}{z}\sum_{c,j=1}^{z} PCC_j^c,$$

where $fc_{i,j}^c$ and $\widehat{fc_{i,J}^c}$ denote the ground truth and predicted post-perturbation fold change in gene expression x_i perturbed by compound j in covariate c, respectively. The covariate c represents cell line or cell type of x_i. $\left(fc_{i,j}^c\right)_{i=1}^n$ represents all perturbed fold-changes for compound j in covariate c. $\overline{\log\left(FC_j^c\right)}$ and $\overline{\log\left(\widehat{FC_j^c}\right)}$ are the ground truth and predicted post-perturbation average logarithm of logarithm of the fold-change in gene expression for compound j in covariate c, respectively. PCC_j^c is the Pearson correlation coefficient of the mean log(FC) perturbed by compounds j in covariate c. The average of the Pearson correlations for all z "cov_compounds" conditions in the test set is referred to as the "Pearson of log(FC) in cov_compounds". A higher value of "Pearson of log(FC) in cov_compounds" indicates a better performance.

14. From the text in Lines 253-265, it is not clear what the difference between Fig 2C and 2D is: is it simply all the cell types vs. some of the cell types shown?

Reply 1-14: We apologize for the confusion between Figures 2C and 2D. Both Figure 2C and 2D provide low-dimensional (t-SNE) representations of the latent embeddings (z_i^l), which are derived from the chemical perturbation effects on the unperturbed state as computed by PRnet.

1) Figure 2C shows the t-SNE representation for all the cell lines included in our study.
2) Figure 2D focuses on a subset of these cell lines, specifically colorectal, breast, and lung cancers, providing a more detailed view of their latent embeddings.

Therefore, your understanding is correct, and the difference between Fig 2C and 2D is all the cell types vs. some of the cell types. The latent embeddings in Figure 2D are a subset of those shown in Figure 2C. We also have relabeled Figures 2C and 2D. To improve the clarity of Figure 2, we have moved Figure 2D to Supplementary Fig. 6 A-C.

Initial version:

Interestingly, we observed that the embedding learned by PRnet also represents cell line similarity in response to various perturbations. Cell lines originating from the same organ exhibit a similarity preference in the latent space, resulting in close spatial locations, such as cell lines from colon, breast and lung. In Fig. 2D, we also demonstrate in detail the low-dimensional (t-SNE) representation of the latent embeddings for cell lines from colorectal cancer, breast cancer, and lung cancer, respectively.

Revised version:

lines 308-316 on page 8

Interestingly, we observed that the embedding learned by PRnet also represents cell line similarity in response to various perturbations. Fig. 2C illustrates the t-SNE representation of the latent embeddings for all cell lines, where cell lines originating

from the same organ exhibit a similarity preference in the latent space, resulting in close spatial locations, such as cell lines from colon, breast, and lung. Supplementary Fig. 6 A-C demonstrates in detail the low-dimensional (t-SNE) representation of the latent embeddings for cell lines from colorectal cancer, breast cancer, and lung cancer, respectively.

Initial version:

Fig 2C and 2D

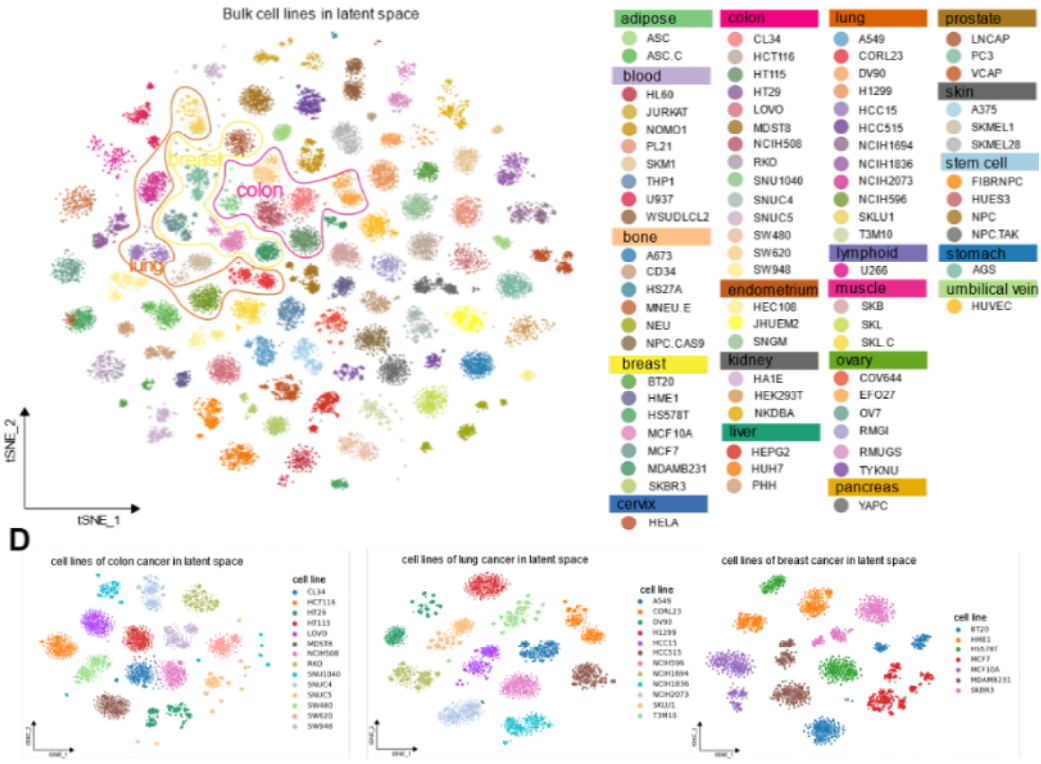


Fig. 2 (C) T-SNE representations of latent embeddings learned by PRnet. Post-perturbation transcriptional profiles are from 82 cell line from 20 different organs/tissues. (D) T-SNE representation illustrates the latent embeddings of post-perturbation transcriptional profiles for cell lines from colorectal cancer, breast cancer, and lung cancer.

Revised version:

Fig 2C

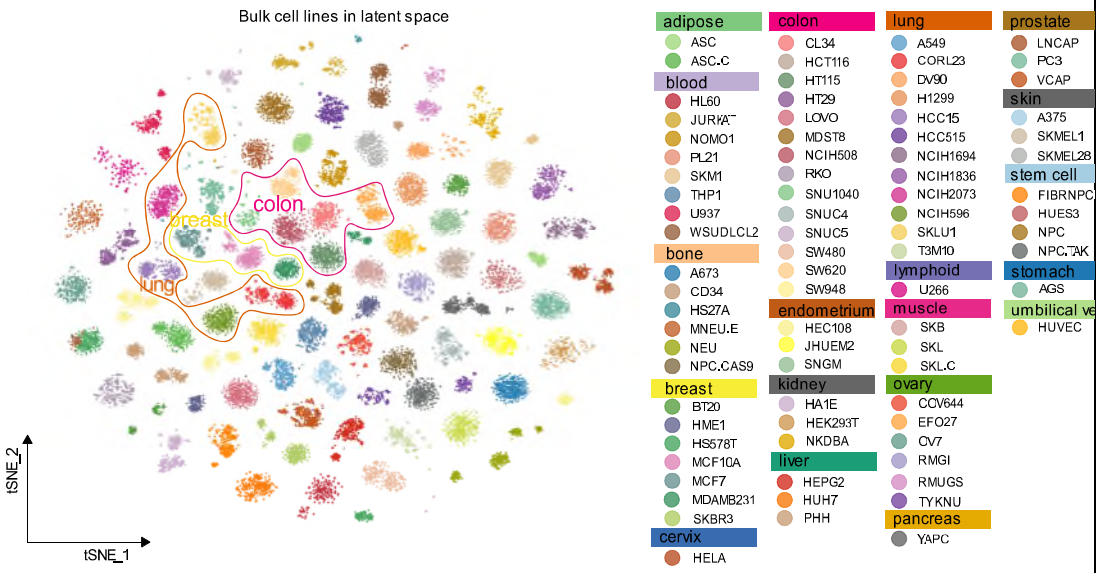
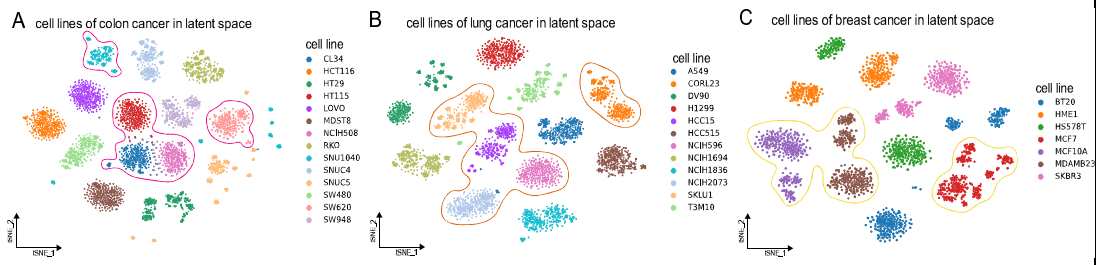


Fig. 2 (C) T-SNE representations of latent embeddings learned by PRnet. Post-perturbation transcriptional profiles are from 82 cell line from 20 different organs/tissues.

Supplementary Fig. 6 A-C



Supplementary Fig. 6 (A-C) T-SNE representation illustrates the latent embeddings of post-perturbation transcriptional profiles for cell lines from colorectal cancer, breast cancer, and lung cancer.

15. Lines 275-277: the text should explain the difference between "R2 in compound" and "R2 in cov compound" metrics.

Reply 1-15: Thank you for your advice, and we are sorry for the confusion between between "R^2 in compounds" and "R^2 in cov_ompounds" metrics. The "R^2 in compounds" metric is the R^2 score of the average post-perturbation gene expression of

the same compounds in the test set. The "R^2 in cov_compounds" metrics is the R^2 score of the average post-perturbation of the same compounds within each specific cell line. The former evaluates the performance of predicted post-perturbation gene expression cross cell line. And the latter evaluates the cell-type-specific performance. We have rephrased the sentence in the Results section (lines 345-349 on page 9) and added the definition of "R^2 in compounds" and "R^2 in cov_compounds" metrics in the Methods section (lines 895-920 on pages 23-24).

Initial version in the Results section:

The "R^2 in compound" metric evaluated the R^2 score of the average gene expression perturbed by the same compounds in the test set. We also compared the cell-type-specific performance "R^2 in cov_compound" metric which evaluated the R^2 score of the average gene expression perturbed by the same compounds within the same cell line.

Revised version in the Results section:

lines 345-349 on page 9

The "R^2 in compound" metric evaluated the R^2 score of the average post-perturbation gene expression of the same compound in the test set. We also compared the cell-type-specific performance "R^2 in cov compound" metric which evaluated the R^2 score of the average post-perturbation of the same compound within each specific cell line.

Revised version in the Methods section:

lines 895-920 on pages 23-24

5. R^2 in compounds: is the mean coefficient of determination score between the predicted and true post-perturbation gene expression of the same compound:

$$\overline{x_t} = \frac{1}{p}\sum_{i=1}^{p} x_{i,t},$$

$$\overline{\widehat{x_t}} = \frac{1}{p}\sum_{i=1}^{p} \widehat{x_{i,t}},$$

$$R_t^2 = R^2\left(\overline{x_t}, \overline{\widehat{x_t}}\right),$$

$$R^2 \text{ in compounds} = \frac{1}{T}\sum_{j=1}^{T} R_t^2,$$

where $x_{i,t}$ and $\widehat{x_{i,t}}$ denote as the ground truth and predicted post-perturbation gene expression x_i perturbed by compound t. The set $\left(x_{i,t}\right)_{i=1}^{p}$ represents all post-perturbation gene expression for compound t. $\overline{x_t}$ and $\overline{\widehat{x_t}}$ are the ground truth and predicted post-perturbation average gene expression for compound t, respectively. R_t^2 is the coefficient of determination score of the mean post-perturbation gene expression of compounds t. The average of the coefficient of determination score for all T compounds in the test set is referred to as the "R^2 in compounds". A higher value of "R^2 in compounds" indicates a better performance.

6. R^2 in cov_compounds: is the mean coefficient of determination score between the predicted and true post-perturbation gene expression of the same compounds within each specific cell line:

$$\overline{x_t^c} = \frac{1}{p}\sum_{i=1}^{p} x_{i,t}^c,$$

$$\overline{\widehat{x_t^c}} = \frac{1}{p}\sum_{i=1}^{p} \widehat{x_{i,t}^c},$$

$$R_t^2 = R^2\left(\overline{x_t^c}, \overline{\widehat{x_t^c}}\right),$$

$$R^2 \text{ in cov\_compounds} = \frac{1}{Q}\sum_{j=1}^{Q} {R^2}_t^c,$$

where $x_{i,t}^c$ and $\widehat{x_{i,t}^c}$ denote as the ground truth and predicted post-perturbation gene expression x_i perturbed by compound t in covariate c, respectively. The covariate c represents cell line or cell type of x_i. The set $\left(x_{i,t}^c\right)_{i=1}^{p}$ represents all post-perturbation gene expression for compound t in covariate c. $\overline{x_t^c}$ and $\overline{\widehat{x_t^c}}$ are the ground truth and predicted post-perturbation average gene expression for compound t in covariate c, respectively. ${R^2}_t^c$ is the coefficient of determination

score of the mean post-perturbation gene expression of compounds t in covariate c. The average of the coefficient of determination score for all Q "cov_compounds" conditions in the test set is referred to as the "R^2 in cov_compounds". A higher value of "R^2 in cov_compounds" indicates a better performance.

16. Lines 284-287: please check if the labels to panels Fig. 2 F and Fig. 2 G are correct.

Reply 1-16: Thank you for your advice. We have revised the labels to panels Fig. 2 F and Fig. 2 G. (Fig. 2 E and Fig. 2 F in page 37, copied below).

Initial version:

random split (Fig. 2 F)

single cell cell lines in latent space (Fig. 2 G)

Revised version:

Random split (Fig. 2 E)

Cell line latent representations from single-cell screens (Fig. 2 F)

17. Line 289 should read: "… consensus trajectory, in which…"

Reply 1-17: Thank you for your advice. We have revised the sentence. Inspired by your feedback (scVIDR [PMID: 37602218]) and suggestion regarding the pseudo-dose trajectory, we have revised related sentences (see lines 358-362 on page 9 in the revised manuscript, copied below).

Initial version:

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses.

Revised version:

lines 358-362 on page 9

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with AG-14361 produces a pseudo-dose trajectory, indicating that AG-14361 induced heterogeneous responses. Several other pseudo-dose trajectory examples were illustrated in Supplementary Fig. 6 D-F.

18. Line 290: "the pseudodose trajectory…" – the concept of a "pseudo-dose" needs to be explained. How is it calculated?

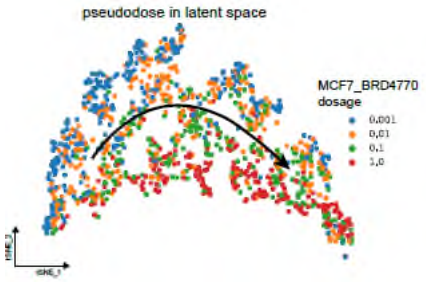
Reply 1-18: Thank you for the suggestion to clarify the "pseudo-dose". The pseudo-dose trajectory is not calculated directly but refers to the variation trend of the mean latent embeddings under the same dosage conditions in the latent space. Inspired by your feedback (scVIDR [PMID: 37602218]), we have added quantitative analysis to our study regarding the pseudo-dose trajectory:

To calculate the pseudo-dose trajectory, we take the following steps:

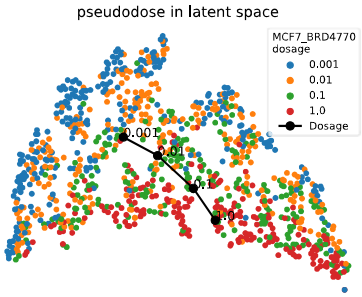
1) Embed each cell's gene expression profile into a latent space by PRnet.
2) Compute the t-SNE embedding of each cell.
3) Group the latent embeddings by drug dose and compute the mean of t-SNE embeddings for each drug dose group.
4) This mean t-SNE embeddings represents a point on the pseudo-dose trajectory.

We have redrawn the pseudo-dose trajectory of cells treated with BRD4770 from MCF7 cell line accordingly. We also added quantitative analysis to several other conditions (including MCF7_AG-14361, MCF7_BMS-911543 and MCF7_IOX2). The revised sentences in the Methods section (lines 932-938 on page 24), the Results section (lines 358-362 on page 9) were copied below, as well as figures (Fig 2G and Supplementary Fig. 6 D-F).

Initial version:



Revised version:



Revised version in the Methods section:

lines932-938 on page 24

The pseudo-dose trajectory. To calculate the pseudo-dose trajectory, we take the mean t-SNE embedding of cells with the same dosage as a point on the pseudo-dose trajectory:

$$\overline{z_d} = \frac{1}{n_d}\sum_{i=1}^{n_d} z_{i,d},$$

where $z_{i,d}$ is the t-SNE latent embedding of cell i at dose d, $n_d$is the number of cells at dose d, and $\overline{z_d}$ is the mean t-SNE embedding for dose d, representing a point on the pseudo-dose trajectory. The sequence of these mean latent embeddings $\overline{z_d}$ forms the pseudo-dose trajectory.

Initial version in the Results section:

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses.

Revised version in the Results section:

lines 358-362 on page 9

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with AG-14361 produces a pseudo-dose trajectory, indicating that AG-14361 induced heterogeneous responses. Several other pseudo-dose trajectory examples were illustrated in Supplementary Fig. 6 D-F.

Fig 2G

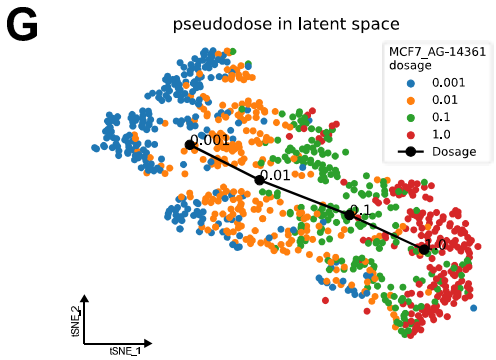
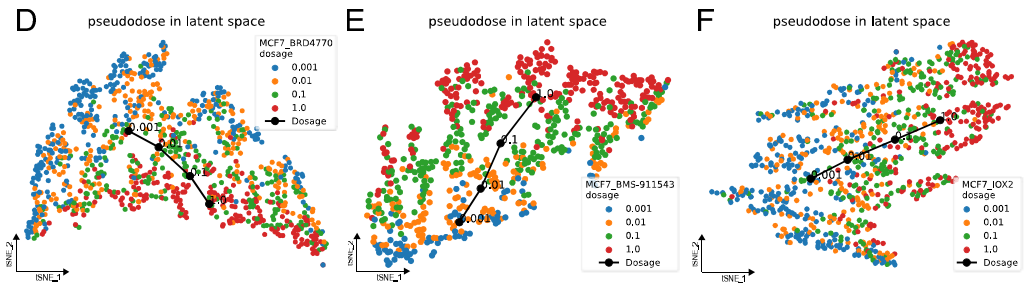


Fig. 2 (G) T-SNE representation of latent embedding of transcriptional profiles from MCF7 cell line perturbed with AG-14361. The pseudo-dose trajectory is displayed with a black line.

Revised version in the Supplementary:

Supplementary Fig. 6 D-F



Supplementary Fig. 6 (D-F) T-SNE representation of latent embedding of transcriptional profiles from MCF7 cell line perturbed with BRD4770, BMS-911543 and IOX2, respectively. The pseudo-dose trajectory is displayed with a black line.

19. The authors point out that PRnet demonstrates the similarity of perturbations for *similar* cell types. It would be helpful if this can be demonstrated by a quantitative metric, e.g. cosine similarity of the perturbation vectors in latent space.

Reply 1-19: Thank you for your insightful advice. We have added the quantitative analysis into our study to evaluate the similarities among perturbations for cell types:

1) We computed the mean embeddings in latent space of all cell lines,
2) We calculated the cosine similarity of every pair of cell lines,
3) We calculated the normalized cosine similarities and visualized them in a heatmap plot.

The heatmap in Supplementary Fig. 6 G shows that the embeddings of cell lines from the same tissue/organ exhibit higher similarity in latent space compared to those from different sources.

We greatly appreciated your advice, which contributed to enhancing the rigor of our research. The revised result (lines 316-321 on page 8) and figure (Supplementary Fig. 6 G) were copied below.

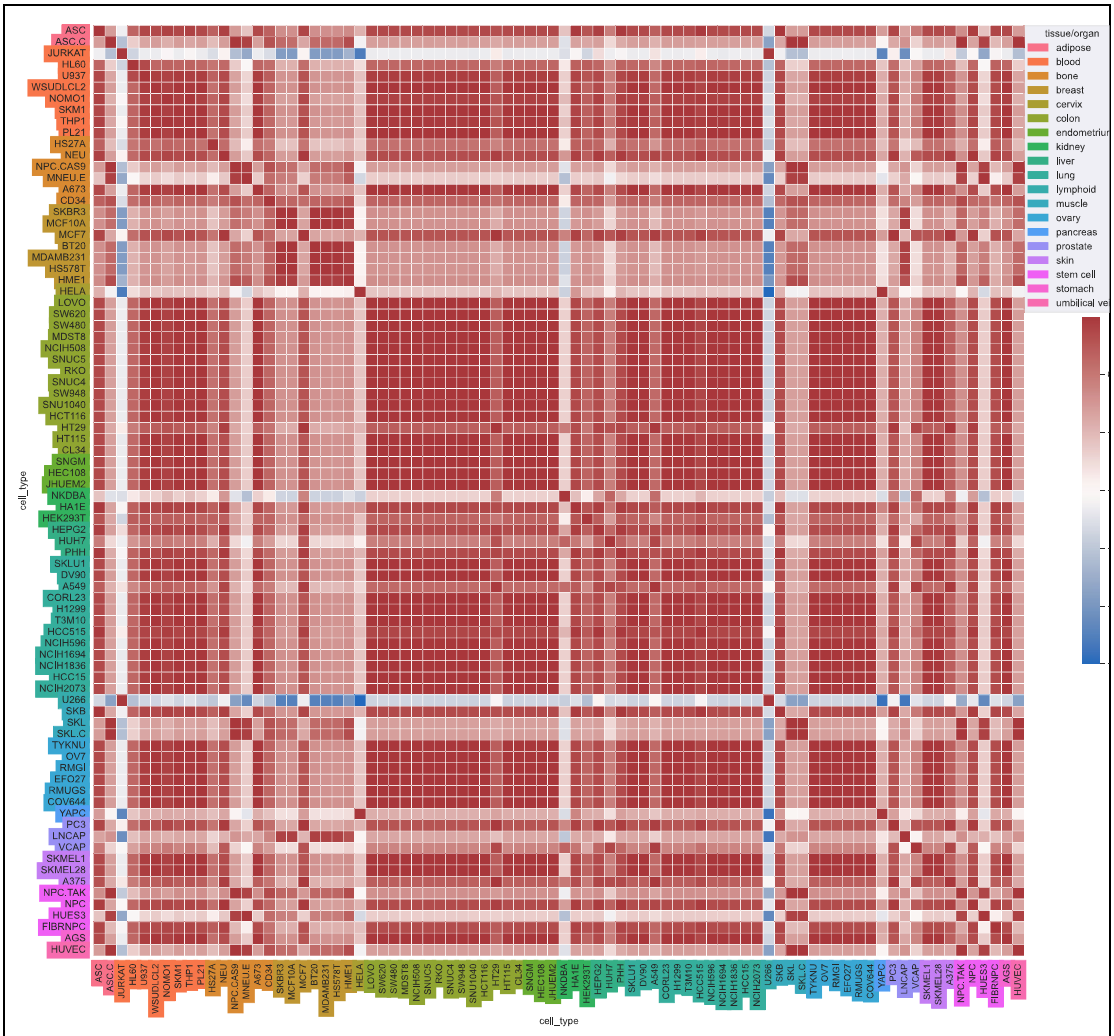
Revised version in the Results section:

lines 316-321 on page 8

To quantitatively evaluate the similarity of perturbations among cell types, we computed the normalized cosine similarity among the mean embeddings in the latent space for all cell lines. The resulting cosine similarity heatmap (Supplementary Fig. 6 G) shows that most cell lines from the same tissue exhibit higher similarity in the latent space compared to those from different tissues.

Revised version in the Supplementary section:

Supplementary Fig. 6 G



Supplementary Fig. 6 G The cosine similarity heatmap among all cell lines. Normalized cosine similarity is represented by the color of the heat map. The horizontal and vertical axes are cell lines, and cell lines from the same source have the same color.

** 20. The basic assumption used in this work for prediction of perturbations in gene expression was not clear to me. Is the perturbation vector assumed to be a function of chemical structure and unperturbed gene expression state, or only of chemical structure? What is the nature of this functional dependence?

Reply 1-20: Thank you for your insightful question. PRnet predicted transcriptional response to perturbations based on chemical structure and unperturbed gene expression. Cells inherently recognize and respond to chemical stimuli, with their responses influenced by both external stimuli and their intracellular state. Different unperturbed

state will produce different responses, so the design of PRnet takes into account both chemical structure and unperturbed gene expression. Moreover, the perturbation vector in this study refers to chemical perturbations $P_i = (s_i, d_i)$, including the chemical structures s_i and their dosages d_i. Here, we provide a detailed explanation of the basic assumption used in our work for predicting perturbations in gene expression.

The functional dependence of PRnet

Cells inherently recognize and respond to chemical stimuli, with their responses influenced by both external stimuli and their intracellular state. Given that most diseases are associated with characteristic gene expression profiles, these profiles have historically been used to reflect underlying disease mechanisms [1-3]. Connectivity Map (CMap) [1] proposed a concept that connected genes, drugs, and diseases by virtue of common gene-expression signatures, leading to projects such as CMap [1], L1000 [4], etc. In experimental projects such as CMap [1] and L1000 [4], a large number of parameters would need to be optimized for each perturbation, including cell types, compounds, compound concentrations, and etc. Based on extensive experiments and studies previously mentioned, our paper hypothesizes that it is possible to predict perturbations in gene expression from chemical structure and unperturbed gene expression state. However, conducting experiments is time-consuming and labor-intensive. Therefore, we have developed a computational method, PRnet, to efficiently predict these gene expression perturbations. We envisioned that a model capable of predicting transcriptional response to novel chemical perturbations would facilitate in silico screening of compounds to identify potential therapeutic candidates. To emulate the process of experimental high-throughput screening, we developed PRnet to predict transcriptional responses to novel chemical perturbations. Given the unperturbed gene expression, chemical structures, and dosages of the compounds applied, PRnet predicts the perturbed gene expression. Then the chemically inducible changes in transcriptional profiles, and gene ranking for compounds can be calculated. Finally, inspired by the "connectivity score" in CMap [1] and L1000 [4], given the gene signature for a

particular disease or known sensitive compounds, gene set enrichment analysis (GSEA) was employed to evaluate compound efficacy with their enrichment scores.

We have revised the description about the functional dependence of PRnet in the Introduction section (lines 118-123 on page 4 and lines 88-101 on page 3), the Results section (lines 160-165 on page 5), and the Methods section (lines 748-765 on page 18), which are copied below.

[1] Lamb J, Crawford E D, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease[J]. science, 2006, 313(5795): 1929-1935.

[2] Rudin C M, Poirier J T, Byers L A, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data[J]. Nature Reviews Cancer, 2019, 19(5): 289-297.

[3] Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer[J]. Nature medicine, 2015, 21(11): 1350-1356.

[4] Subramanian A, Narayan R, Corsello S M, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles[J]. Cell, 2017, 171(6): 1437-1452. e17.

Initial version in the Introduction section:

Taking compound structures and unperturbed transcriptional profiles as input, PRnet adapts novel compounds and diseases in various perturbation scenarios without prior knowledge and annotation.

Revised version in the Introduction section:

lines 118-123 on page 4

PRnet adapts novel compounds and diseases in various perturbation scenarios by taking compound structures and unperturbed transcriptional profiles as input to predict transcriptional responses. The Perturb-adaptor uses simplified molecular-

input line-entry system (SMILES) chemical encoding as input, enabling generalization to unseen compounds without prior knowledge and annotation.

Initial version in the Results section:

In the single cell or bulk HTS, transcriptional responses to chemical perturbations are affected by multiple conditions, such as compound structures, compound dosages, and covariates such as cell types, and cell lines.

Revised version in the Results section:

lines 160-165 on pages 5

Cells inherently recognize and respond to chemical stimuli, with their responses influenced by both external stimuli and their intracellular state. In the single cell or bulk HTS, transcriptional responses to chemical perturbations are affected by multiple conditions, such as compound structures, compound dosages, and covariates such as cell types, and cell lines.

Initial version in the Methods section:

Given a dataset $D = (x_i, P_i)_{i=1}^{N}$, where $x_i \in R^n$ denotes as the n-dimensional gene expression and P_i is the attribute set of perturbation, PRnet was trained to learn a function f that predicts transcriptional responses to novel perturbations. As a generation model, the Gaussian negative log-likelihood loss is chosen as the loss function. For HTS RNA-seq datasets $D = (x_i, P_i)_{i=1}^{N} = \left(x_i, (s_i, d_i, c_i)\right)_{i=1}^{N}$, the compounds s_i and dosages d_i attributes are usually considered, in which $s_i = \left(s_{i,1}, s_{i,2}, \dots, s_{i,M}\right)$ describes the Canonical SMILES format of M compounds of perturbation i and $d_i = \left(d_{i,1}, d_{i,2}, \dots, d_{i,M}\right)$ are the dosages of compounds. When $d_{i,j} = 0$, the compounds j was not applied in perturbation i. If $M = 0$, there is no perturbation performed, and in this case perturbation i is an unperturbed state, otherwise is in a perturbed state. c_i represents covariates such as cell lines of

perturbation i and can be extended to cells, cell lines, patients, species, and so on depending on the datasets.

Revised version in the Methods section:

lines 748-765 on page 18

Given a dataset $D = (x_i, P_i)_{i=1}^{N}$, where $x_i \in R^n$ denotes as the n-dimensional gene expression and P_i is the attribute set of perturbation, PRnet was trained to learn a function f that predicts transcriptional responses $\widehat{x_i} = f(x_i^u, P_i)$ to novel perturbations. Here, x_i^u represents the unperturbed gene expression, $\widehat{x_i}$ is the predicted perturbed gene expression, and $P_i = (s_i, d_i)$ is the attribute set of perturbation, which includes the chemical structures s_i and dosages of the compounds d_i. As a generation model, the Gaussian negative log-likelihood loss is chosen as the loss function. For HTS RNA-seq datasets $D = (x_i, P_i)_{i=1}^{N} = \left(x_i, (s_i, d_i)\right)_{i=1}^{N}$, the gene expression x_i, compounds s_i and dosages d_i are usually considered, in which $s_i = \left(s_{i,1}, s_{i,2}, \dots, s_{i,M}\right)$ describes the Canonical SMILES format of M compounds of perturbation i and $d_i = \left(d_{i,1}, d_{i,2}, \dots, d_{i,M}\right)$ are the dosages of compounds. When $d_{i,j} = 0$, the compounds j was not applied in perturbation i. If $M = 0$, there is no perturbation performed, and in this case perturbation i is an unperturbed state x_i^u, otherwise is in a perturbed state. Additionally, the covariates vector c_i contains discrete covariates such as cell-types, cell lines or species, depending on the available data. While c_i does not directly serve as input to the function f, it relates to the intermediate variable vector x_i^u of the function f.

Initial version in the Introduction section:

Gene signature matching methods for bulk HTS prediction like DLEPS [13] and OCTAD [14] screened candidates effectively. DLEPS [13] predicted chemically induced changes in transcriptional profiles directly from molecular structure, and OCTAD [14] virtually screened compounds by matching the cancer-specific

expression signature to compound-induced gene expression profiles. Although DLEPS and OCTAD screened candidates effectively, they failed to predict cell-type-specific transcriptional response to novel perturbations and model cellular heterogeneity which is highly relevant to treatments.

Revised version in the Introduction section:

lines 88-101 on page 3

Given that most diseases are associated with characteristic gene expression profiles, Connectivity Map (CMap) [15] proposed a concept that connected genes, drugs, and diseases by virtue of common gene-expression signatures, leading to projects such as CMap [15], L1000 [1], etc. Inspired by this concept, DLEPS [16], OCTAD [17] and other studies (such as [18–20], etc.) other studies utilized gene signature matching methods to screen candidates by finding drugs that reverse the disease signature. DLEPS [16] predicted chemically induced changes in transcriptional profiles directly from molecular structure, and OCTAD [17] virtually screened compounds by matching the cancer-specific expression signature to compound-induced gene expression profiles. Although DLEPS and OCTAD screened candidates effectively based on bulk HTSs, they failed to predict cell-type-specific transcriptional response to novel perturbations and model cellular heterogeneity which is highly relevant to treatments.

21. Lines 594-595 should read: "… novel chemical perturbations that were never experimentally *studied* (or *carried out*) at bulk and single-cell levels."

Reply 1-21: Thank you for your advice. We have revised the sentence (see lines 676-680 on page 16 in the revised manuscript, copied below).

Initial version:

To address the limited exploration power of existing experimental methods, we developed PRnet, which supports the prediction of transcriptional responses to novel chemical perturbations that were never experimentally perturbed at bulk and single-

cell levels.

Revised version:

lines 676-680 on page 16

To address the limited exploration power of existing experimental methods, we developed PRnet, which supports the prediction of transcriptional responses to novel chemical perturbations that were never experimentally studied at bulk and single-cell levels.

22. Line 702: “Training” is mis-spelt.

Reply 1-22: Thank you for your advice. We have corrected the misspelling of "Training" in line 825 on page 20.

Initial version:

Trianing and testing

Revised version:

lines 825 on page20

Training and testing

Reviewer #2 (Remarks to the Author):

In this article, Qi et al built a deep learning model named PRnet to predict chemical perturbation of transcriptional perturbation in different cellular contexts. The model was trained on several data sets including bulk and single-cell RNA-seq in response to FDA and natural chemical compound libraries, using Simplified Molecular Input Line Entry System (SMILES) to represent compound structures. Using the model, the author performed in silico screens and validated some of the findings by in vitro assays and literature reviews, including various cell lines, lung cancer, colon cancer, fatty liver, and inflammatory bowel diseases.

Overall the study is well performed and comprehensive, which can be a useful resource for the broader community. Below are my comments:

Reply: We greatly appreciated your positive comments. Thank you for your recognition of the efforts put into building the model and conducting comprehensive analyses. We have incorporated these suggestions into the revised manuscript accordingly as listed in details below. Your feedback encourages us to continue our efforts to contribute valuable resources to the broader scientific community.

1. The impact of the paper will be enhanced if the author provides a way for online search of their results and databases and testing new compound using their AI model. At least articulate a clear way of sharing all the datasets to the broader readership. I do not think the paper should be published without the authors' commitment to share the AI model and all the results in a straightforward way for broader dissemination.

Reply 2-1: Thank you for your valuable suggestions. In the revised manuscript, we have provided a website (prnet.drai.cn) where readers can browse and download our database and results. Additionally, the predicted signature results in this paper have been deposited in the OMIX, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix: accession no. OMIX006910) and packaged our prediction model and required environment on GitHub (https://github.com/Perturbation-Response-Prediction/PRnet).

To facilitate access to our data by the broader scientific community, we have developed a website with the following features:

1. Browse metadata of all compounds, diseases, tissues and cell lines.
2. Download predicted preturbed signatures of all compounds, cell lines, tissues and diseases mentioned in the paper.
3. Download the predicted large-scale atlas of perturbations profiles.

We will continue to predict more compounds in future research and update our website to make it accessible to a wider audience.

Revised version:

lines 1099-1104 on page 28

We provided a website (prnet.drai.cn) to browse and download our database and results. The predicted signature results in this paper have been deposited in the OMIX [64, 65], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix: accession no. OMIX006910)

2. The paper needs to explain how good SMILES is for representing the compound structure and what the tradeoffs and alterantives are.

Reply 2-2: Thank you for your thoughtful advice. Here, we explain the strengths and limitations of SMILES, as well as potential alternatives.

Use of SMILES for Compound Structure Representation

Simplified Molecular Input Line Entry System [30] (SMILES) is widely used for representing chemical structures due to its simplicity and efficiency in encoding complex molecules as strings. SMILES notation allows for easy storage, retrieval, and manipulation of chemical data, making it a popular choice for various computational tasks in chemistry and drug discovery. However, SMILES may not fully capture complex molecular features such as 3D geometry, conformational flexibility, or dynamic behavior, which can be important for understanding molecular interactions.

While SMILES is a robust and widely-used representation, there are alternative methods that can be considered. MOL/SDF formats represent chemical structures with explicit atom coordinates and bond connectivity. They are particularly useful for 3D molecular modeling and visualization, but they are more complex and require more storage space compared with SMILES. Graph-Based Representations use graph theory to represent molecules where atoms are nodes and bonds are edges. This kind of methods can provide a more intuitive and flexible representation, but require pre-training to obtain better representations. Considering the simplicity and efficiency of SMILES, we finally chose SMILES representing chemical structures.

In the revised manuscript, we have added the description about the strengths and limitations of SMILES (lines 118-123 on page 4 and lines 201-203 on page 6), as well as alternatives such as MOL/SDF files, and graph-based representations (lines 700-717 on page 17).

Initial version:

Taking compound structures and unperturbed transcriptional profiles as input, PRnet adapts novel compounds and diseases in various perturbation scenarios without prior knowledge and annotation.

Revised version:

lines 118-123 on page 4

PRnet adapts novel compounds and diseases in various perturbation scenarios by taking compound structures and unperturbed transcriptional profiles as input to predict transcriptional responses. The Perturb-adaptor uses simplified molecular-input line-entry system [30] (SMILES) chemical encoding as input, enabling generalization to unseen compounds without prior knowledge and annotation.

Initial version:

By taking SMILES of the compound as the input, the Perturb-adaptor has sufficient flexibility to screen large-scale compound libraries without any prior knowledge.

Revised version:

lines 201-203 on page 6

SMILES is widely used for representing chemical structures due to its simplicity and efficiency in encoding complex molecules as strings. By taking SMILES of the compound as the input, the Perturb-adaptor has sufficient flexibility to screen large-scale compound libraries without any prior knowledge.

Revised version:

lines 700-717 on page 17

Due to the scalability of the SMILES format [30] and the RDKit [31], PRnet is able to encode unseen complex compounds as embeddings. SMILES is widely used for representing chemical structures due to its simplicity and efficiency in encoding complex molecules as strings. However, SMILES may not fully capture complex molecular features ssuch as 3D geometry, conformational flexibility, or dynamic behavior, which can be important for understanding molecular interactions. Alternative encoding methods such as MOL/SDF Files and graph-based Representations are more flexible representations in capturing the 3D geometry structure of compounds. MOL/SDF formats represent chemical structures with explicit atom coordinates and bond connectivity. Graph-Based Representations use graph theory to represent molecules where atoms are nodes and bonds are edges. These methods can provide a more intuitive and flexible representation. But they are more complex or require pre-training to obtain better representations. In the scenario of large-scale *in silico* screening, we chose SMILES to encode chemical structures as its simplicity and scalability. Alternative encoding methods such as MOL/SDF Files and graph-based Representations will be considered in future work.

[30] Weininger, D.: Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. Journal of chemical information and computer sciences 28(1), 31–36 (1988)

3. For the disease examples such as NASH, Crohn's and PCOS, the author need to list the ranked drug predictions in tables in extended dataset. Currently they only referred to a figure for each disease that highlighted a couple of handpicked targets that they had literature evidence, but it is impossible (at least for me) to read the entire prediction list to assess the quality of the predictions and whether it predicts promising new compounds. Also, for the few handpicked examples for the various diseases, the authors need to include more data from the in silico screening to show how the model predicted transcriptional perturbation (what were the pathways/genes perturbed? In heatmap and GSEA), what exactly are the enrichment scores, how the predictions were ranked, etc.

Reply 2-3: Thank you for your valuable advice. We have listed the rankings of all 935 FDA-approved drugs for 233 diseases in the Supplementary file Table 4. For clarity, we have listed enrichment scores of the top 10 recommended drugs for three diseases in the Supplementary Tab. 4. Additionally, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of predicted up/downregulated genes of recommended drugs (Supplementary Fig. 7). The added Table (Supplementary Tab. 4) and Figure (Supplementary Fig. 7) were copied below. Thank you for bringing these important points to our attention.

Revised version:

lines 605-606 on page 15

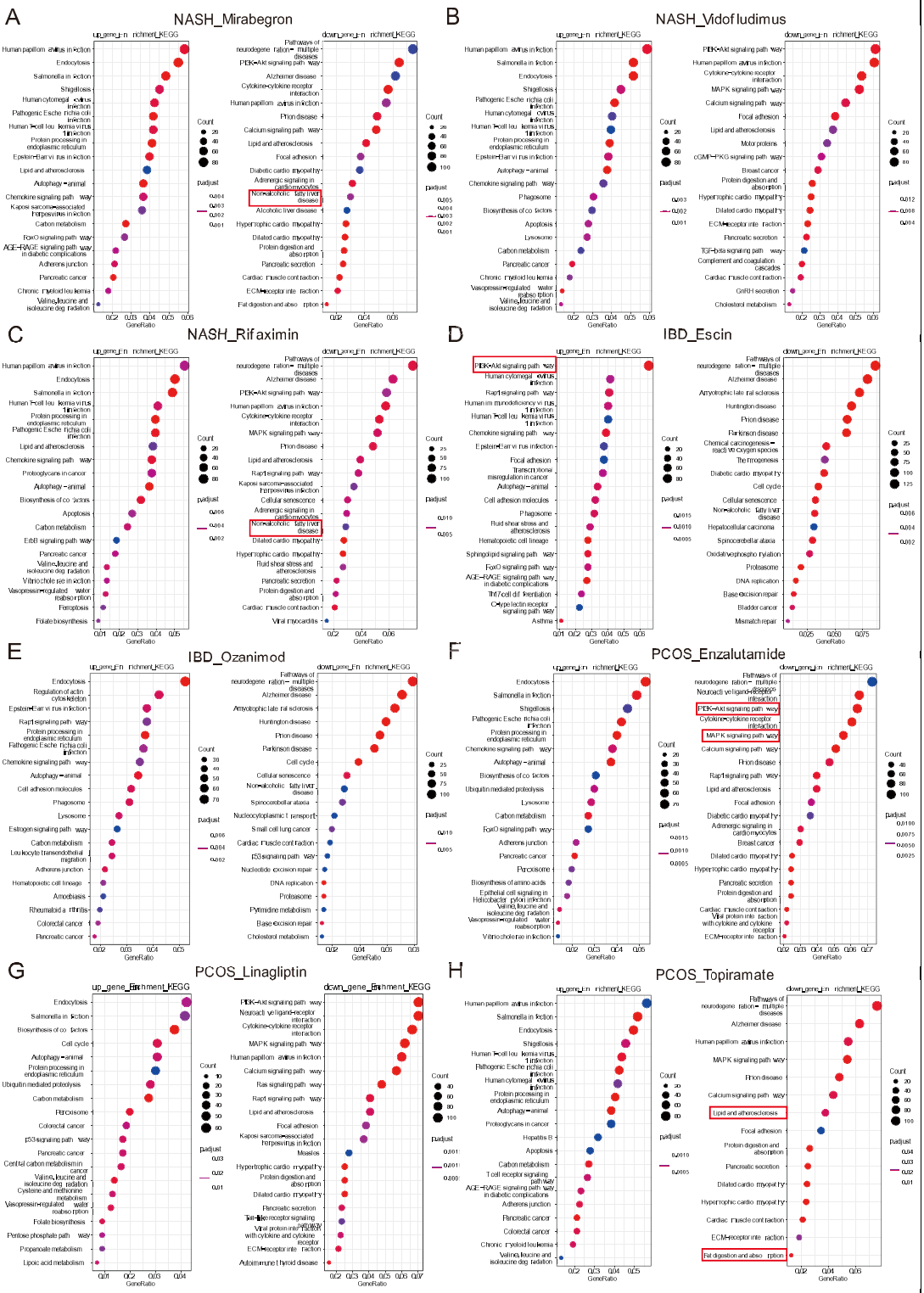
Supplementary Tab. 4 lists the Top 10 enrichment scores for drugs against NASH, Crohn's disease, and PCOS.

Supplementary Tab. 4 Top 10 enrichment scores for drugs against NASH, Crohn's disease, and PCOS

Disease	Compound	a of enrichment score	b of enrichment score	Enrichment score
NASH	Mirabegron	0.107755	-0.169943	0.277698
	Hygromycin	0.138039	-0.131809	0.269848
	Anidulafungin	0.083510	-0.173291	0.256801
	Tebipenem	0.161092	-0.090009	0.251101

	Vidofludimus	0.128646	-0.120228	0.248874
	Rifaximin	0.122784	-0.120173	0.242957
	Aprepitant	0.095854	-0.139477	0.235330
	Penicillamine	0.088275	-0.143927	0.232202
	Daclatasvir	0.091986	-0.136580	0.228566
	Temozolomide	0.058668	-0.161298	0.219966
Crohn's disease	Entrectinib	0.178993	-0.160508	0.339501
	Omarigliptin	0.198258	-0.135800	0.334058
	Escin	0.125242	-0.191233	0.316475
	Taladegib	0.170269	-0.142891	0.313160
	Pracinostat	0.183138	-0.128306	0.311445
	Enasidenib	0.169965	-0.138373	0.308338
	Carmustine	0.168898	-0.138313	0.307211
	Flavopiridol	0.115370	-0.190074	0.305444
	Ozanimod	0.167106	-0.133180	0.300285
	Omipalisib	0.166943	-0.132578	0.299521
PCOS	Tebipenem	0.213969	0.213969	0.343800
	Hygromycin	0.196443	0.196443	0.331134
	Enzalutamide	0.167405	0.167405	0.322596
	Azilsartan	0.172115	0.172115	0.319767
	Rifaximin	0.165659	0.165659	0.315044
	Mirabegron	0.179043	0.179043	0.305577
	Penciclovir	0.130578	0.130578	0.298134
	Linagliptin	0.189892	0.189892	0.296413
	Topiramate	0.143278	0.143278	0.294766
	Topiroxostat	0.168178	0.168178	0.291646

Supplementary Fig. 7



Supplementary Fig. 7 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis results of predicted up/downregulated genes of recommended drugs against NASH, Crohn's disease, and PCOS.

lines 626-630 on page 15

We also performed KEGG pathway enrichment analysis of predicted up/down-regulated genes of Mirabegron, Vidofludimus, and Rifaximin (Supplementary Fig. 7). The KEGG pathway enrichment analysis of downregulated genes for both Mirabegron and Rifaximin suggested they may downregulate pathways associated with NASH.

---

lines 646-649 on page 16

The predicted upregulated pathways (Supplementary Fig. 7) suggested Escin may upregulated the PI3K/Akt signaling pathway which regulates a broad cascade of target proteins including nuclear factor kappa B (NF-κB) and glycogen synthase kinase-3β (GSK-3β) [48].

[48] Williams, D.L., Ozment-Skelton, T., Li, C.: Modulation of the phosphoinositide 3-kinase signaling pathway alters host response to sepsis, inflammation, and ischemia/reperfusion injury. Shock 25(5), 432–439 (2006)

---

lines 663-669 on page 16

The predicted regulated pathways (Supplementary Fig. 7) of Enzalutamide showed that Enzalutamide may upregulated the PI3K/Akt signaling pathway and MAPK signaling pathway which are related to Androgen signaling [52]. The predicted downregulated pathways suggested Topiramate may downregulated the Lipid and atherosclerosis and the Fat digestion and absorption pathway, thereby contributing to weight loss.

[52] Kaarbø, M., Klokk, T.I., Saatcioglu, F.: Androgen signaling and its interactions with other signaling pathways in prostate cancer. Bioessays 29(12), 1227–1238 (2007)

Reviewer #3 (Remarks to the Author):

Qi et al. presents the development of PRnet, a perturbation-conditioned deep generative model designed to predict transcriptional responses to novel chemical perturbations. The manuscript highlights the model's outperformance of alternative methods in predicting responses across novel compounds, pathways, and cell lines, enabling gene-level response interpretation and in-silico drug screening for diseases based on gene signatures. Additionally, PRnet has been utilized to identify and experimentally test novel compound candidates against small cell lung cancer and colorectal cancer, and has generated a large-scale integration atlas of perturbation profiles, covering 88 cell lines and 52 tissues perturbed by various screening compound libraries.

The manuscript is clear and well-written. Authors present an useful and novel tool that can be used in different biological, clinical and preclinical contexts including the drug discovery field. However I have some questions and suggestions that authors should consider before publication:

Reply: Thank you for your positive comments. We are glad to hear that you found the tool useful and novel for various biological, clinical, and preclinical applications, particularly in drug discovery. We have carefully incorporated these advices into the revised manuscript accordingly as listed in details below.

Major points

1) Authors present the generation of a large-scale integration atlas of perturbation profiles and the implementation of a robust and scalable drug recommendation workflow based on the profiles atlas. This large-scale atlas of perturbations predicts 25M of transcriptional responses for different drug libraries and datasets. However, those are limited to cell lines and the GTEx dataset. I am wondering if it is possible to include bulk or single cell data from patients tissues? In my opinion, authors should include some results (ex: TCGA cohort, single-cell human cell atlas) to prove the

application in patient data. In fact, there are some studies in cancer of pre- post treatment scRNA-seq data: is it possible to validate if the predicted transcriptional profile by PRNet is similar to the post-treatment patient samples? This could be interesting for instance to predict the patient state to propose a combination or second-line treatment.

Reply 3-1: Thank you for your valuable advice. We recognize the importance of validating our model's predictions using patient data to demonstrate its clinical relevance. In the revised manuscript, we have incorporated data from a cohort of 13 pediatric acute myeloid leukemia (AML) patients who achieved disease remission after chemotherapy treatment in Supplementary Note 2. We also added related descriptions in the Results section (lines 363-370 on page 9) and the Methods section (lines 1060-1078 on pages 27-28) in the revised manuscript.

1) Data Collection:

We collected scRNA-seq data of paired pre- and post-chemotherapy whole bone marrow samples from 13 pediatric AML patients who achieved disease remission following chemotherapy, as described by Zhang et al [34]. These patients were treated with either a low-dose chemotherapy (LDC) regimen or a standard-dose chemotherapy (SDC) regimen. The LDC regimen consisted of cytarabine (10 mg/m2) and mitoxantrone or Idarubicin, administered concurrently with G-CSF (5 ug/kg). The SDC regimen consisted of cytarabine (100 mg/m2), daunorubicin, and etoposide.

We downloaded the paired pre- and post-chemotherapy expression profiles and annotations of these patients.

2) Data Processing:

We compiled the expression profiles of all 224,217 cells from the 13 patients and performed normalization and log-transformation.

Each cell was annotated with patient information, cell type, and chemotherapy regimen based on the information provided in the study. The LDC regimen was annotated with cytarabine and mitoxantrone, while the SDC regimen was annotated with cytarabine, daunorubicin, and etoposide.

Through highly variable gene analysis, we retained 33,538 highly variable genes for training.

3) Training and Testing:

We grouped the cells by patient, cell type, and chemotherapy regimen. Specifically, patient_cell_type_chemotherapy_regimen_split (PCRS) means that cells in one group have the same cell type, patient and chemotherapy_regimen. And patient_cell_type_split (PCS) means that cells in one group from the same cell type and patient. While patient_split (PS) means that cells in one group from the same patient. We divided the PCRS categories into training, validation, and test sets in a 6:2:2 ratio.

Then We trained and tested the data using PRnet.

For testing, we grouped the test cells by PCRS, PCS and PS categories. We calculated the mean expression profiles for each group and computed the Pearson correlation between the true and predicted expression profiles.

4) Results:

The test results showed that PRnet achieved a mean Pearson correlation of 0.732 in PCRS, 0.755 in PCS and 0.951 in PS. PRnet showed robustness in predicting transcriptional responses for pediatric AML patients post-chemotherapy.

We also analyzed the top 10 differentially expressed genes (DEGs) before and after chemotherapy and compared the true and predicted expression of DEGs. The predicted distribution of gene expression after chemotherapy closely aligned with the actual observed distribution.

The results demonstrate the potential of PRnet to assist clinical application. We believe that these findings will enhance the clinical relevance and applicability of our approach, particularly in predicting patient response to treatment and guiding personalized medicine strategies.

Revised version in the Results section:

lines 363-370 on page 9

To validate the clinical relevance of PRnet, we incorporated scRNA-seq data from a

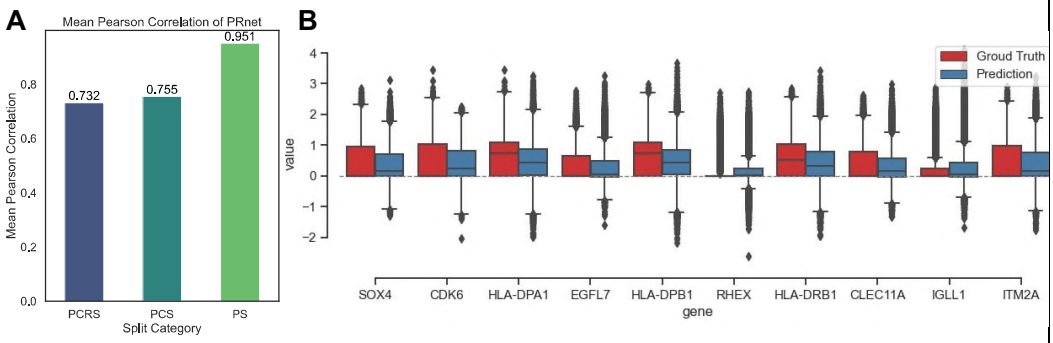
cohort of pediatric acute myeloid leukemia (AML) patients [35]. We trained PRnet to predict transcriptional responses to chemotherapy treatments and validated it's potential for clinical application. PRnet showed robustness in predicting transcriptional responses for pediatric AML patients post-chemotherapy, demonstrating PRnet's potential to assist in clinical applications. For detailed experimental results, please see Supplementary Note 2 and Supplementary Fig. 8.

Supplementary Note 2 PRnet predicted the chemotherapy response of Pediatric AML Patients

To validate the application in patient data of PRnet, we incorporated scRNA-seq data from a cohort of pediatric acute myeloid leukemia (AML) patients. Our goal was to train PRnet with patient-derived scRNA-seq data to predict transcriptional responses to chemotherapy treatments. We collected scRNA-seq data of paired pre- and post-chemotherapy whole bone marrow samples from 13 pediatric AML patients who achieved disease remission following chemotherapy (Zhang et al. [2]). These patients were treated with low-dose (LDC) or standard-dose (SDC) chemotherapy regimens. After preprocessing, a total of 224,217 cells from the 13 patients with 33,538 highly variable genes, were retained for training. We grouped the cells by patient, cell type, and chemotherapy regimen. Specifically, patient_cell_type_chemotherapy_regimen_split (PCRS) means that cells in one group have the same cell type, patient and chemotherapy regimen. And patient_cell_type_split (PCS) means that cells in one group from the same cell type and patient. While patient_split (PS) means that cells in one group from the same patient. We divided the PCRS categories into training, validation, and test sets in a 6:2:2 ratio.

PRnet was then trained on these data to fit transcriptional responses to chemotherapy. For testing, we grouped the test cells by PCRS, PCS and PS categories. We calculated the mean expression profiles for each group and computed the mean Pearson correlation between the true and predicted expression profiles. The

test results showed that PRnet achieved a mean Pearson correlation of 0.732 in PCRS, 0.755 in PCS and 0.951 in PS (see Supplementary Figure 9 A). PRnet showed robustness in predicting transcriptional responses for pediatric AML patients post-chemotherapy. Additionally, we analyzed the top 10 differentially expressed genes (DEGs) before and after chemotherapy, comparing the true and predicted expression of these DEGs. The predicted distribution of gene expression after chemotherapy closely aligned with the actual observed distribution (see Supplementary Figure 9 B). These results demonstrate PRnet's potential to assist in clinical applications, highlighting its utility in predicting transcriptional responses and aiding in the design of personalized treatment strategies.



Supplementary Figure 9 PRnet predicted chemotherapy response of patient data. (A) Mean Pearson correlations of PRnet in predicting transcriptional responses to chemotherapy under three grouping strategies. (B) The box plots illustrate true and predicted expressions of the 10 different expression genes of AML patient before and after chemotherapy.

Revised version in the Methods section:

lines 1061-1078 on pages 27-28

The scRNA-seq data of pediatric AML patients cohort. We collected scRNA-seq data of paired pre- and post-chemotherapy whole bone marrow samples from 13 pediatric AML patients who achieved disease remission following chemotherapy, as described by Zhang et al. [2]. These patients were treated with either a low-dose chemotherapy (LDC) regimen or a standard dose chemotherapy (SDC) regimen. The

LDC regimen consisted of cytarabine (10 mg/m^2) and mitoxantrone or Idarubicin, administered concurrently with G-CSF (5 μg/kg). The SDC regimen consisted of cytarabine (100 mg/m^2),1048 daunorubicin, and etoposide. We compiled the expression profiles of all 224,217 cells from the 13 patients and performed normalization and log-transformation. Each cell was annotated with patient information, cell type, and chemotherapy regimen based on the information provided in the study. The LDC regimen was annotated with cytarabine and mitoxantrone, while the SDC regimen was annotated with cytarabine, daunorubicin, and etoposide. Through highly variable gene analysis, we retained 33,538 highly variable genes for training. The scRNA-seq data of patients were downloaded from the Genome Sequence Archive for Human at the BIG data center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation under accession number (OMIX005223).

[35] Zhang, Y., Jiang, S., He, F., Tian, Y., Hu, H., Gao, L., Zhang, L., Chen, A., Hu, Y., Fan, L., et al.: Single-cell transcriptomics reveals multiple chemoresistant properties in leukemic stem and progenitor cells in pediatric aml. Genome biology 24(1), 199 (2023)

2) Authors should specify in the manuscript where the large-scale atlas of perturbations profiles is stored/included. It is not clear to me if they develop a resource that users can query to check perturbations or cell lines and find gene signatures associated with the perturbation. I think creating this resource is essential and will greatly improve the manuscript.

Reply 3-2: Thank you for your valuable suggestions. In the revised manuscript, we have provided a website (prnet.drai.cn) where readers can browse and download our database and results. Additionally, the predicted signature results in this paper have been deposited in the OMIX, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix:

accession no. OMIX006910) and packaged our prediction model and required environment on GitHub (https://github.com/Perturbation-Response-Prediction/PRnet). To facilitate access to our data by the broader scientific community, we have developed a website with the following features:

1. Browse metadata of all compounds, diseases, tissues and cell lines.
2. Download predicted preturbed signatures of all compounds, cell lines, tissues and diseases mentioned in the paper.
3. Download the predicted large-scale atlas of perturbations profiles.

We will continue to predict more compounds in future research and update our website to make it accessible to a wider audience.

Revised version:

lines 1099-1104 on page 28

We provided a website (prnet.drai.cn) to browse and download our database and results. The predicted signature results in this paper have been deposited in the OMIX [64, 65], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix: accession no. OMIX006910).

3) It is described the importance of capturing cell-type specific response and cellular heterogeneity which are relevant in drug response. However, the manuscript is focused on capturing the heterogeneous response between cell lines and not within the same cell-type which is a known-event (Ben-David et al. Nature). I think, it would be interesting to describe and study the inter and intra-heterogeneity between cell lines. Can be calculated the relationship between heterogeneity with the predicted perturbated transcriptional profile? Do homogeneous cells have more “sensitivity” than heterogeneous? I would expect that some responses even from different cell lines are more similar. Following the drug recommendation workflow, it would be nice to know

if there are more recommended drugs for specific cancer cell types, based on heterogeneity, etc.

Reply 3-3: Thank you for your insightful comments. We have conducted additional experiments to explore both inter- and intra-heterogeneity within the same cell type, focusing on A549 and MCF7 cell strains as described by Ben-David et al. (Nature) [33]. To explore the impact of inter- and intra-heterogeneity of cell lines on drug responses, we conducted additional experiments focusing on A549 and MCF7 cell strains, as described by Ben-David et al. [33] in Supplementary Note 1. These experiments were designed to explore inter- and intra-cellular heterogeneity, similar patterns of the same MOA drug across cell strains, and screening candidate compounds for heterogeneous cells.

Experiment Overview

1. Data Collection:

1) We collected expression profiles of 27 MCF7 strains and 23 A549 strains from Ben-David et al. Nature [33].
2) We included 35 compounds from the chemical screen, retaining 33 after removing those which RDKit could not generate to FCFP fingerprints.
3) Expression profiles for 978 L1000 landmark genes were used for each strain.

2. Exploring Heterogeneity Between Cell Lines:

1) We utilized our model trained on L1000 data to predict post-perturbation expression profiles for the strains at a dosage of 5 μM.
2) The similarity Clustermap of latent embeddings effectively distinguished between A549 and MCF7 cell lines.
3) T-SNE embeddings confirmed the separation between the two cell lines.

3. Exploring Heterogeneity Within Cell Lines:

1) For MCF7 strains, we calculated t-SNE embeddings using latent embeddings from the 33 compounds.
2) The results showed distinct clusters for MCF7.M, MCF7.C, MCF7.Q, and MCF7.H strains, while other strains clustered by the same drug.

3) The similarity clustermap of latent embeddings aligned closely with the hierarchical clustering in Ben-David et al.'s Figure 3b. Results showed heterogeneity within cell lines, leading to variability in drug sensitivity. These results supported the conclusion that drug response clustering is consistent with genetic or gene expression clustering.

4. Active Patterns of Compounds with the Same Mechanism:

1) We examined the activity patterns of three proteasome inhibitors (bortezomib, MG-132, and carfilzomib) using t-SNE embeddings.
2) T-SNE embeddings of MCF7 strains clustered by their type, showing similar activity patterns across the inhibitors with the same mechanism of action.

5. Compound Screening in Heterogeneous Strains:

1) We computed fold-changes of predicted perturbations for MCF7 strains.
2) We selected some of compounds with decreasing effects to screen all 33 compounds.
3) For each MCF7 strain, we used half of the provided sensitive compounds as known active drugs, calculating enrichment scores for all compounds.
4) Top 10 candidates contained 76.3% with decreasing effects and 83.3% with both decreasing and weakly active effects.

These additional experiments have deepened our understanding of intra- and inter-heterogeneity in predicting drug responses in perturbation response. Thank you again for your valuable feedback. Your suggestions have significantly contributed to the depth and rigor of our study.

Response in Manuscript

We have added a supplementary note (Supplementary Note 1) detailing the exploration of intra- and inter-heterogeneity in cell lines, which is copied below. We also added related descriptions in the Results section (lines 322-335 on pages 8-9) and the Methods section (lines 1079-1086 on page 28) in the revised manuscript.

Supplementary Note 1 Exploring the impact of inter- and intra-heterogeneity of cell lines on drug responses

It has been approved in Ben-David et al. Nature [1] that cancer cell lines are in fact highly genetically heterogeneous, and genetic heterogeneity leads to varying patterns of gene expression, which in turn result in differential drug sensitivity. To explore the impact of inter- and intra-heterogeneity of cell lines on drug responses. We collected expression profiles for 27 MCF7 strains and 23 A549 strains. These MCF7 strains include 19 strains without drug treatment or genetic manipulation, 7 strains with neutral genetic modifications, and one strain (MCF7.M) expanded in vivo following anti-oestrogen therapy. The A549 strains include 13 untreated and unmodified strains and 10 genetically manipulated strains. We also collected the drug response to 35 active compounds of MCF7 strains in the chemical screen, including decreasing (active), decreasing (weakly active), and inactive.

After preprocessing, we used expression profiles for 978 L1000 landmark genes as unperturbed profile and 33 compounds at a 5 μM dosage for prediction. PRnet predicted transcriptional responses, fold-change of perturbations and latent embeddings of MCF7 strains and A549 strains to chemical screening.

Firstly, we explore heterogeneity between cell lines. Cosine similarities among MCF7 and A549 strains were calculated, and clustermap analysis were performed on these similarities. This analysis effectively distinguished between A549 and MCF7 cell lines, demonstrating the model's capacity to capture heterogeneity between cell lines (Supplementary Figure 8 A). The low-dimensional (t-SNE) representation of the latent embeddings further confirmed the separation between these two cell lines (Supplementary Figure 8 G).

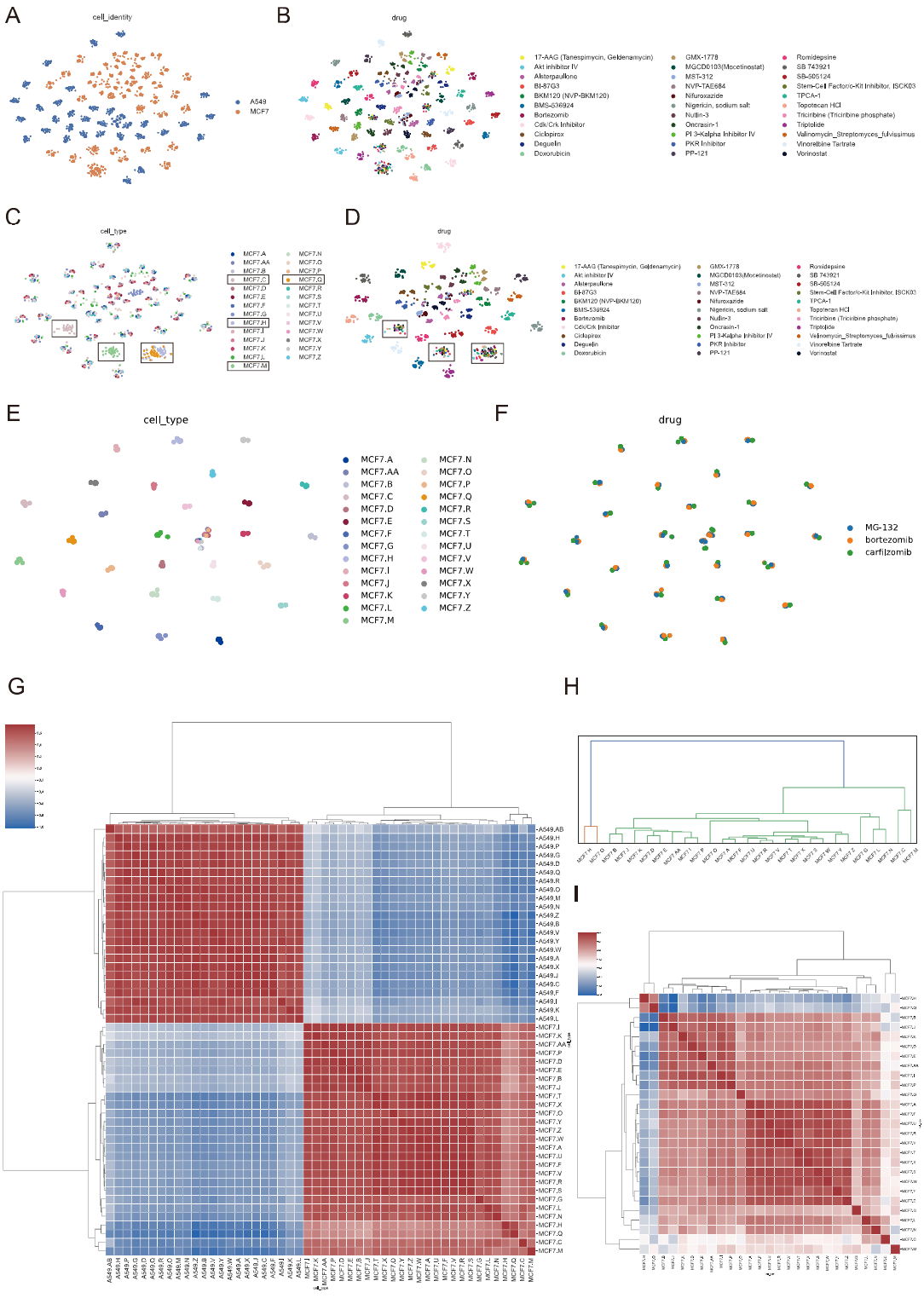
Next, we investigated intra-heterogeneity within cell lines by focusing on MCF7 strains. We created a similarity clustermap of latent embeddings (Supplementary Figure 8 I), and found it’s hierarchical cluster tree (Supplementary Figure 8 H) similar to the hierarchical clustering based on global gene expression profiles in Ben-David et al [1]. The t-SNE analysis results of latent embeddings (Supplementary

Figure 8 C-D) revealed that while most MCF7 strains clustered by drugs, certain strains like MCF7.M and MCF7.C, as well as MCF7.Q and MCF7.H, formed distinct clusters. These strains (MCF7.M and MCF7.C, as well as MCF7.Q and MCF7.H) also showed less similarity with other strains in clustermap. Results showed heterogeneity within cell lines, leading to variability in drug sensitivity. These results indicated that drug response clustering was highly similar to genetic or gene expression clustering of the MCF7 strains, supporting the conclusions drawn in the original study.

Then, we also examined the activity patterns of three proteasome inhibitors (bortezomib, MG-132, and carfilzomib) using t-SNE embeddings (Supplementary Figure 8 E-F). The MCF7 strains clustered by type, showing similar activity patterns across the inhibitors with the same mechanism of action which aligned with findings in the original study.

Finally, we assess the model's ability to screen compounds in heterogeneous cell strains. The fold-changes of perturbations were used to calculate enrichment score of compounds. The same screening method was performed as in screening candidates for SCLC and CRC. We identified compounds with decreasing effects and screened all 33 compounds. For each MCF7 strain, we randomly selected half of the known sensitive compounds for each strain from the provided list as reference sensitive compounds. We calculated enrichment scores for all compounds of each sensitive compound for every MCF7 strain. For each specific strain, mean enrichment scores for all compounds to reference sensitive compounds were calculated, and the top 10 compounds were selected as candidates. Among the top 10 candidates for all MCF7 strains, 76.3% had decreasing effects, and 83.3% had either decreasing or weakly active effects. The detailed Top10 results were shown in Supplementary Table 5. These results demonstrated the model's effectiveness in identifying potential drugs for heterogeneous cell strains.

[1] Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K., Strathdee, C.A., Dempster, J., Lyons, N.J., Burns, R., et al.: Genetic and transcriptional evolution alters cancer cell line drug response. Nature 560(7718), 325–330 (2018)



Supplementary Figure 8 Inter- and intra-heterogeneity of cell lines on drug responses. (A-B) T-SNE representations of latent embeddings for MCF7 and A549 strains. (C-D) T-SNE representations of latent embeddings for MCF7 strains. (E-F) T-SNE representations of latent embeddings for MCF7 strains treated with bortezomib, MG-132, and carfilzomib. (G) Clustermap of latent embeddings for MCF7 and A549 strains. (H) Hierarchical cluster tree of latent embeddings for MCF7 strains. (I) Clustermap of latent embeddings for MCF7 strains.

Supplementary Table 5 Hit Rate of Top10 candidates for all MCF7 strains

MCF7 strains	Hit Rate of Active	Hit Rate of Weakly Active	Hit Rate of Active and Weakly Active
MCF7.A	0.7	0.1	0.8
MCF7.AA	0.9	0.1	1.0
MCF7.B	0.8	0.2	1.0
MCF7.C	0.5	0.1	0.6
MCF7.D	0.7	0.1	0.8
MCF7.E	0.5	0.3	0.8
MCF7.F	1.0	0.0	1.0
MCF7.G	0.7	0.1	0.8
MCF7.H	0.6	0.2	0.8
MCF7.I	0.5	0.2	0.7
MCF7.J	0.7	0.2	0.9
MCF7.K	0.5	0.1	0.6
MCF7.L	0.8	0.1	0.9
MCF7.M	0.5	0.2	0.7
MCF7.N	0.8	0.1	0.9
MCF7.O	0.7	0.2	0.9
MCF7.P	1.0	0.0	1.0
MCF7.Q	0.6	0.1	0.7
MCF7.R	0.7	0.2	0.9
MCF7.S	0.5	0.3	0.8
MCF7.T	0.6	0.1	0.7
MCF7.U	0.9	0.1	1.0
MCF7.V	0.7	0.3	1.0
MCF7.W	0.3	0.2	0.5
MCF7.X	0.3	0.2	0.5
MCF7.Y	0.3	0.3	0.6
MCF7.Z	0.4	0.2	0.6
Average	0.763	0.070	0.833

Revised version in the Results section:

lines 322-335 on pages 8-9

Human cancer cell lines have facilitated drug discoveries in cancer biology, but they are neither clonal nor genetically stable. This instability can generate variability in drug sensitivity, as shown by Ben-David et al.[33]. The genomic evolution of cancer cell lines leads to a high degree of variation across cell line strains. To explore the impact of inter- and intra-heterogeneity of cell lines on drug responses, we conducted additional experiments focusing on A549 and MCF7 cell strains, as described by Ben-David et al. [33] in Supplementary Note 1. These experiments were designed to explore inter- and intra-cellular heterogeneity, similar patterns of the same MOA drug across cell strains, and screening candidate compounds for heterogeneous cells.

These results indicated that drug response was highly correlated with genetic or gene expression which aligned with findings in the original study. These findings underscore the utility of our model in capturing both inter- and intra-heterogeneity, enhancing its application in screening for specific cancer strains.

Revised version in the Methods section:

lines 1079-1086 on page 28

The data of A549 and MCF7 cell strains. We collected expression profiles for 27 MCF7 strains and 23 A549 strains from Ben-David et al. [34]. Expression profiles for 978 L1000 landmark genes were used for each strain. We included 35 compounds from the chemical screen in the study [34], retaining 33 after removing those that RDKit could not generate to FCFP fingerprints. We also collected the drug response to 35 active compounds of MCF7 strains in the chemical screen, including decreasing (active), decreasing (weakly active), and inactive.

[34] Ben-David, U., Siranosian, B., Ha, G., Tang, H., Oren, Y., Hinohara, K., Strathdee, C.A., Dempster, J., Lyons, N.J., Burns, R., et al.: Genetic and transcriptional evolution alters cancer cell line drug response. Nature 560(7718), 325–330 (2018)

4) Fig 2G "MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses." It is not clear why is a homogeneous response. Please explain in more detail this point.

Reply 3-4: We apologize for the spelling mistake regarding the pseudodose trajectory in Fig 2G. The term "homogeneous responses" should be "heterogeneous response". T-SNE embedding of MCF7 cell line treated with BRD4770 at different dosage produced a pseudo-dose trajectory , indicating a range of heterogeneous responses rather than a uniform homogeneous one. Additionally, we added quantitative analysis

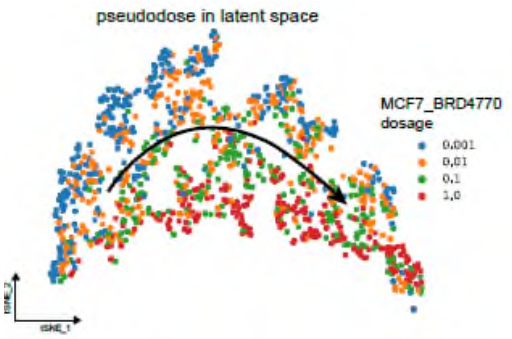
to our study regarding the pseudo-dose trajectory according to advice given by Reviewer 1:

To calculate the pseudo-dose trajectory, we take the following steps:

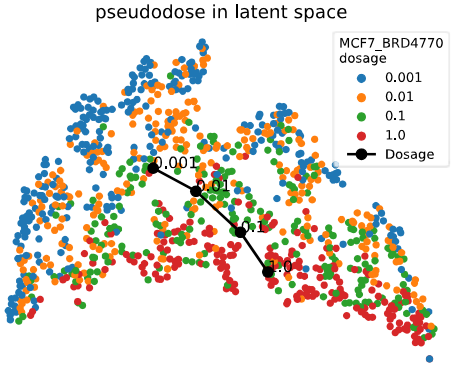
1) Embed each cell's gene expression profile into a latent space by PRnet.
2) Compute the t-SNE embedding of each cell.
3) Group the latent embeddings by drug dose and compute the mean of t-SNE embeddings for each drug dose group.
4) This mean t-SNE embeddings represents a point on the pseudo-dose trajectory.

We have redrawn the pseudo-dose trajectory of cells treated with BRD4770 from MCF7 cell line accordingly. We also added quantitative analysis to several other conditions (including MCF7_AG-14361, MCF7_BMS-911543 and MCF7_IOX2). The revised sentences in the Methods section (lines 932-938 on page 24), the Results section (lines 358-362 on page 9), and figures (Fig 2G and Supplementary Fig. 6 D-F), are also shown below.

Initial version:



Revised version:



Revised version in the Methods section:

lines 932-938 on page 24

The pseudo-dose trajectory. To calculate the pseudo-dose trajectory, we take the mean t-SNE embedding of cells with the same dosage as a point on the pseudo-dose trajectory:

$$\overline{z_d} = \frac{1}{n_d}\sum_{i=1}^{n_d} z_{i,d},$$

where $z_{i,d}$ is the t-SNE latent embedding of cell i at dose d, $n_d$is the number of cells at dose d, and $\overline{z_d}$ is the mean t-SNE embedding for dose d, representing a point on the pseudo-dose trajectory. The sequence of these mean latent embeddings $\overline{z_d}$ forms the pseudo-dose trajectory.

Initial version in the Results section:

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses.

Revised version in the Results section:

lines 358-362 on page 9

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with AG-14361 produces a pseudo-dose trajectory, indicating that AG-14361 induced heterogeneous responsses. Several other pseudo-dose trajectory examples were illustrated in Supplementary Fig. 6 D-F.

Fig 2G

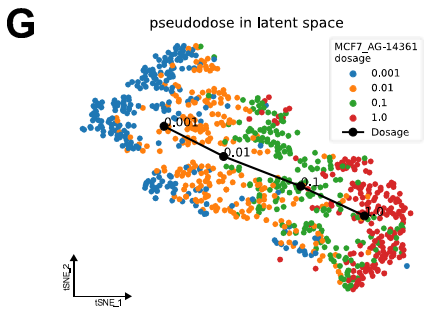
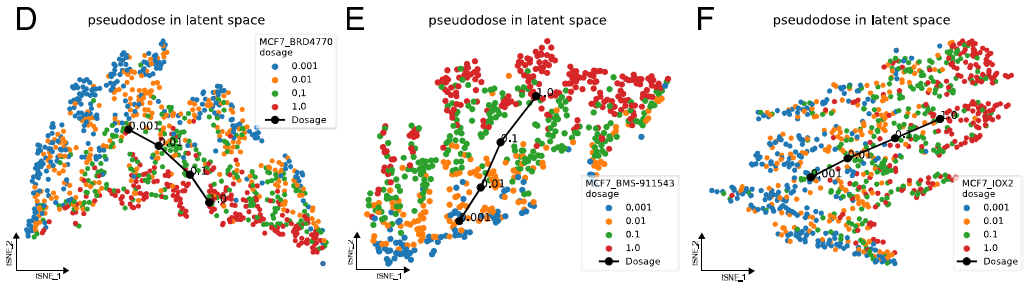


Fig. 2 (G) T-SNE representation of latent embedding of transcriptional profiles from MCF7 cell line perturbed with AG-14361. The pseudo-dose trajectory is displayed with a black line.

Revised version in the Supplementary:

Supplementary Fig. 6 D-F



Supplementary Fig. 6 (D-F) T-SNE representation of latent embedding of transcriptional profiles from MCF7 cell line perturbed with BRD4770, BMS-911543 and IOX2, respectively. The pseudo-dose trajectory is displayed with a black line.

5) My main concern is about considering a compound sensitive which is employed to identify potential compound candidates in SCLC and CRC. Here, you are assuming that perturbated transcriptional profile are associated with sensitivity in some diseases which is not necessarily true although this idea was presented a while ago by Lamb et al. 2006 Science, resulting in projects such as Connectivity Map, L1000, etc. For the drug recommendation workflow: are you calculating the enrichment score employing a similar assumption of reverse signature paradigm established by Lamb et al?. Authors should describe in more detail these previous works and how they can compare it with previous results/studies that followed this approach. So, can the authors include information about cancer cell lines drug sensitivity screenings such as GDSC, CTRP, CCLE, etc to validate this point?

Reply 3-5: Thank you for raising this concern and providing valuable feedback. You are correct that our drug recommendation workflow employs the reverse signature

paradigm as established by Lamb et al in CMap [15]. We also understand your concern about the assumption of compound sensitivity based on perturbated transcriptional profiles and its association with diseases. We revised the manuscript and added more detailed description about previous works and assumptions and methodologies employed in our work.

Drug Recommendation Workflow and Enrichment Score Calculation

You are correct that our drug recommendation workflow employs the reverse signature paradigm as established by Lamb et al in CMap [15]. We create the gene expression signatures of disease and compounds by PRnet or Experimental data and calculate enrichment scores using the Gene Set Enrichment Analysis (GSEA) methodology. Specifically, the Kolmogorov–Smirnov test in GSEA was used to evaluate the distribution of the query genes in the reference list. When calculating enrichment scores, the score to reverse disease up- and down-regulated features are calculated separately and then summed together. This involves:

1) Generating Query Signatures: Creating gene expression signatures representing the disease states or sensitive compound responses.
2) Generating Screen Compound Signatures: Predicting perturbated transcriptional profiles induced by compounds, and generate gene expression signatures.
3) Calculating Enrichment Scores: Employing the Kolmogorov–Smirnov test to calculate enrichment scores of screening compounds to query signatures, analogous to those used in CMap and L1000.

Assumption of Perturbated Transcriptional Profiles and Sensitivity

Our study assumes that perturbated transcriptional profiles are associated with drug sensitivity in diseases such as small cell lung cancer (SCLC) and colorectal cancer (CRC). This assumption builds on foundational work by Lamb et al. (2006, Science) and subsequent projects like Connectivity Map (CMap) and L1000, which have demonstrated the utility of transcriptional signatures in identifying drug-disease connections. Inspired by this concept, DLEPS [16], OCTAD [17] screened candidates by finding drugs that reverse the disease signature. This assumption has demonstrated

its effectiveness in previous studies, such as the successful identification of inhibitors for small cell lung cancer [19] and the discovery of novel drug candidates for the treatment of metastatic colorectal cancer [20].

However, we acknowledge that this assumption may not hold true for all diseases, and there are instances where transcriptional changes do not correlate directly with drug sensitivity. Several studies have identified limitations and challenges of this assumptions. Jie Cheng et al. [55] made systematic evaluations to assessed CMAP methodologies, suggesting that CMap given an effective matching algorithm especially for neoplastic diseases. But they also find the reverse signature paradigm may not be universally applicable. Some of the disease signatures are low quality even worse than random performance which lead to poor accuracy. Moreover, many diseases may not be represented accurately by the transcriptional response in the current signatures. Gh Rasool Bhat [56] discusses how cancer cells exhibit phenotypic plasticity under drug treatment, leading to drug resistance. Drug resistance mechanisms often involve complex regulatory networks that are not fully captured by changes in gene expression alone.

The reverse signature paradigm is an effective matching algorithm in many diseases. But there are also cases do not fit this algorithm, more comprehensive omics data, such as genomics, epigenetics, and proteomics should be considered in the future work to better characterize disease states and facilitate drug discovery.

We apologize for the lack of description of the assumptions and methodologies in previous related works. The revised manuscript in the Introduction section (lines 88-101 on page 3), the Results section (lines 213-216 on page 6), the Methods section (lines 939-962 on pages 24-25) and the Discussion section (lines 718-731 on pages 17-18) are copied blow:

Initial version in the Introduction section:

Gene signature matching methods for bulk HTS prediction like DLEPS [13] and OCTAD [14] screened candidates effectively. DLEPS [13] predicted chemically

induced changes in transcriptional profiles directly from molecular structure, and OCTAD [14] virtually screened compounds by matching the cancer-specific expression signature to compound-induced gene expression profiles. Although DLEPS and OCTAD screened candidates effectively, they failed to predict cell-type-specific transcriptional response to novel perturbations and model cellular heterogeneity which is highly relevant to treatments.

Revised version in the Introduction section:

lines 88-101 on page 3

Given that most diseases are associated with characteristic gene expression profiles, Connectivity Map (CMap) [15] proposed a concept that connected genes, drugs, and diseases by virtue of common gene-expression signatures, leading to projects such as CMap [15], L1000 [1], etc. Inspired by this concept, DLEPS [16], OCTAD [17] and other studies (such as [18–20], etc.) utilized gene signature matching methods to screen candidates by finding drugs that reverse the disease signature. DLEPS [16] predicted chemically induced changes in transcriptional profiles directly from molecular structure, and OCTAD [17] virtually screened compounds by matching the cancer-specific expression signature to compound-induced gene expression profiles. Although DLEPS and OCTAD screened candidates effectively based on bulk HTSs, they failed to predict cell-type-specific transcriptional response to novel perturbations and model cellular heterogeneity which is highly relevant to treatments.

Revised version in the Results section:

lines 213-216 on page 6

Inspired by the assumption embodied in the CMap [15] concept that gene signatures are used as indicators reflecting the underlying mechanisms of diseases, PRnet predicted new therapeutic candidates by finding drugs that reverse the disease signature.

Initial version in the Methods section:

Screening candidates based on gene signatures. PRnet provides a workflow for screening candidates for complex diseases without well-defined target proteins according to the reference changes in gene sets. In step 1, given the SMILES of the screening compounds library and unperturbed transcriptional profiles of specific cell lines, PRnet predicts the perturbed transcriptional profiles of all compounds across multiple concentration gradients in specific cell lines with 3 repeats or computational robustness. In step 2, transform the transcriptional profiles of 978 landmark genes into 12,328 genes through linear transformation (the linear transformation vectors was calculated from the L1000 project [1]). Subsequently, compute the fold-change of transcriptional profiles: the average fold-change of each drug across cell lines for drug recommendation, and the cell line specific fold-change of each compound for hit compounds screening. Rank genes based on their fold-change values. In step 3, compute the enrichment scores for all screening compounds:

Revised version in the Methods section:

lines 939-962 on pages 24-25

Screening candidates based on gene signatures. Inspired by the CMap connectivity score [15] and L1000 project [1], PRnet employed a similar reverse signature paradigm to screen candidates based on gene signatures. PRnet provided a workflow for screening candidates for complex diseases according to the reference changes in gene sets.

Step 1: Given the SMILES of the screening compounds library and unperturbed transcriptional profiles of specific cell lines, PRnet predicts the perturbed transcriptional profiles of all compounds across multiple concentration gradients. This prediction is performed with three repeats to ensure computational robustness.

Step 2: The transcriptional profiles of 978 landmark genes are transformed into profiles of 12,328 genes using linear transformation vectors derived from the L1000 project [1]. The fold-change of transcriptional profiles is then calculated, which includes the average fold-change of each compound across cell lines. Genes are ranked based on their fold-change values. Gene expression signatures representing

the disease states or sensitive compound responses are generated, known as query signatures. Query signatures include an up-regulated gene set and a down-regulated gene set which are the reversed signatures of the disease. Gene expression signatures of screening compounds are also generated, which are the ranked gene lists.

Step 3: The Kolmogorov-Smirnov test is employed to calculate enrichment scores, which measure the connectivity of screening compound profiles with the query signatures. The score to reverse disease up- and down-regulated features are calculated separately and then summed together:

Revised version in the Discussion section:

lines 718-731 on pages 17-18

The reverse signature paradigm established by Lamb et al in CMap [15] has demonstrated its effectiveness in previous studies [16-20]. However, this may not hold true for all diseases, and there are instances where transcriptional changes do not correlate directly with drug sensitivity. Jie Cheng et al. [55] found the reverse signature paradigm may perform poor accuracy when some disease signatures are of low quality. Rasool Bhat [56] discusses how cancer cells exhibit phenotypic plasticity under drug treatment, leading to drug resistance which often involves complex regulatory networks that are not fully captured by changes in gene expression alone. The reverse signature paradigm is an effective matching algorithm in many diseases. But there are also cases do not fit this algorithm, more comprehensive omics data, such as genomics, epigenetics, and proteomics should be considered in the future work to better characterize disease states and facilitate drug discovery.

[1] 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)

[13] Roohani, Y., Huang, K., Leskovec, J.: Predicting transcriptional outcomes of novel multigene perturbations with gears. Nature Biotechnology, 1–9 (2023)

[14] Kamimoto, K., Stringa, B., Hoffmann, C.M., Jindal, K., Solnica-Krezel, L., Morris, S.A.: Dissecting cell identity via network inference and in silico gene perturbation. Nature 614(7949), 742–751 (2023)

[15] Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., et al.: The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. science 313(5795), 1929–1935 (2006)

[16] 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)

[17] Zeng, B., Glicksberg, B.S., Newbury, P., Chekalin, E., Xing, J., Liu, K., Wen, A., Chow, C., Chen, B.: Octad: an open workspace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features. Nature protocols 16(2), 728–753 (2021)

[18] Sirota, M., Dudley, J.T., Kim, J., Chiang, A.P., Morgan, A.A., Sweet-Cordero, A., Sage, J., Butte, A.J.: Discovery and preclinical validation of drug indications using compendia of public gene expression data. Science translational medicine 3(96), 96–779677 (2011)

[55] Cheng, J., Yang, L., Kumar, V., Agarwal, P.: Systematic evaluation of connectivity map for disease indications. Genome medicine 6, 1–8 (2014)

[56] Bhat, G.R., Sethi, I., Sadida, H.Q., Rah, B., Mir, R., Algehainy, N., Albalawi, I.A., Masoodi, T., Subbaraj, G.K., Jamal, F., et al.: Cancer cell plasticity: From cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. Cancer and Metastasis Reviews 43(1), 197–228 (2024)

6) In my opinion discussion and results could be more detailed. The results section presents the performance evaluation of PRnet for predicting responses to unseen

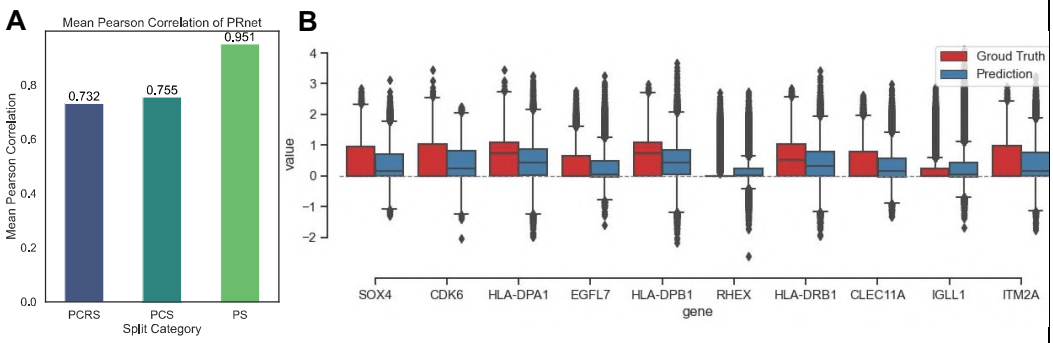
perturbations in human diseases (cancer, NASH, etc). However, I miss more applications in patient data as I commented before. The discussion section should describe in more detail PRNet limitations which are not properly addressed.

Reply 3-6: Thank you for your valuable advice. We first incorporated data a cohort of 13 pediatric acute myeloid leukemia (AML) patients to predict chemotherapy response in Supplementary Note 2 (copied below). Then, we included detailed limitations and alternative in encoding chemical structures by the SMILES format [30] and the RDKit [31] in the discussion section. We also included discussion about instances where transcriptional changes do not correlate directly with drug sensitivity. Detailed discussions about limitations are also given below (lines 700-731 on pages 17-18):

Supplementary Note 2 PRnet predicted the chemotherapy response of Pediatric AML Patients

To validate the application in patient data of PRnet, we incorporated scRNA-seq data from a cohort of pediatric acute myeloid leukemia (AML) patients. Our goal was to train PRnet with patient-derived scRNA-seq data to predict transcriptional responses to chemotherapy treatments. We collected scRNA-seq data of paired pre- and post-chemotherapy whole bone marrow samples from 13 pediatric AML patients who achieved disease remission following chemotherapy (Zhang et al. [34]). These patients were treated with low-dose (LDC) or standard-dose (SDC) chemotherapy regimens. After preprocessing, a total of 224,217 cells from the 13 patients with 33,538 highly variable genes, were retained for training. We grouped the cells by patient, cell type, and chemotherapy regimen. Specifically, patient_cell_type_chemotherapy_regimen_split (PCRS) means that cells in one group have the same cell type, patient and chemotherapy regimen. And patient_cell_type_split (PCS) means that cells in one group from the same cell type and patient. While patient_split (PS) means that cells in one group from the same patient. We divided the PCRS categories into training, validation, and test sets in a 6:2:2 ratio.

PRnet was then trained on these data to fit transcriptional responses to chemotherapy. For testing, we grouped the test cells by PCRS, PCS and PS categories. We calculated the mean expression profiles for each group and computed the mean Pearson correlation between the true and predicted expression profiles. The test results showed that PRnet achieved a mean Pearson correlation of 0.732 in PCRS, 0.755 in PCS and 0.951 in PS (see Supplementary Figure 9 A). PRnet showed robustness in predicting transcriptional responses for pediatric AML patients post-chemotherapy. Additionally, we analyzed the top 10 differentially expressed genes (DEGs) before and after chemotherapy, comparing the true and predicted expression of these DEGs. The predicted distribution of gene expression after chemotherapy closely aligned with the actual observed distribution (see Supplementary Figure 9 B). These results demonstrate PRnet's potential to assist in clinical applications, highlighting its utility in predicting transcriptional responses and aiding in the design of personalized treatment strategies.



Supplementary Figure 9 PRnet predicted chemotherapy response of patient data. (A) Mean Pearson correlations of PRnet in predicting transcriptional responses to chemotherapy under three grouping strategies. (B) The box plots illustrate true and predicted expressions of the 10 different expression genes of AML patient before and after chemotherapy.

Revised version in the Discussion section

lines 700-731 on pages 17-18

Although PRnet is effective in predicting drug candidates, there is still room for improvement. Due to the scalability of the SMILES format [30] and the RDKit [31], PRnet is able to encode unseen complex compounds as embeddings. SMILES is widely used for representing chemical structures due to its simplicity and efficiency in encoding complex molecules as strings. However, SMILES may not fully capture complex molecular features such as 3D geometry, conformational flexibility, or dynamic behavior, which can be important for understanding molecular interactions. Alternative encoding methods such as MOL/SDF Files and graph-based Representations are more flexible representations in capturing the 3D geometry structure of compounds. MOL/SDF formats represent chemical structures with explicit atom coordinates and bond connectivity. Graph-Based Representations use graph theory to represent molecules where atoms are nodes and bonds are edges. These methods can provide a more intuitive and flexible representation. But they are more complex or require pre-training to obtain better representations. In the scenario of large-scale in silico screening, we chose SMILES to encode chemical structures as its simplicity and scalability. Alternative encoding methods such as MOL/SDF Files and graph-based Representations will be considered in future work.

The reverse signature paradigm established by Lamb et al in CMap [15] has demonstrated its effectiveness in previous studies [16-20]. However, this may not hold true for all diseases, and there are instances where transcriptional changes do not correlate directly with drug sensitivity. Jie Cheng et al. [55] found the reverse signature paradigm may perform poor accuracy when some disease signatures are of low quality. Rasool Bhat [56] discusses how cancer cells exhibit phenotypic plasticity under drug treatment, leading to drug resistance which often involves complex regulatory networks that are not fully captured by changes in gene expression alone. The reverse signature paradigm is an effective matching algorithm in many diseases. But there are also cases do not fit this algorithm, more comprehensive omics data, such as genomics, epigenetics, and proteomics should be considered in the future work to better characterize disease states and facilitate drug

discovery.

7) More documentation and good demos to run the tool available in Github are needed:

- The installation is simple, but for someone without basic Python knowledge it might lack details.

- There is no documentation on how to run the tool, there are several separate scripts and it is not intuitive.

- The demos cannot be tested and more details are needed: The dataset files are missing. It is not clear where they come from. Also, within the scripts they are named completely different from the reference.

- Also specify if these files need to be preprocessed before using the demos.

Reply 3-7: Thank you for your feedback regarding the documentation and demos available on GitHub. We have added more details to the documentation and demos in Github (https://github.com/Perturbation-Response-Prediction/PRnet).

1) Installation Instructions: We have added more details to documentation to provide more comprehensive instructions, including a guide to install dependencies by Anaconda or Dockerfile.
2) Reqired Datasets: All required datasets can be accessed in the online webset (http://prnet.drai.cn/Statistics/). A demo dataset used to verify the environment configuration been uploaded to GitHub
3) Data Preprocessing: Detailed instructions and examples to guide users through the necessary data preparation processes have been added.
4) Running the Tool: A documentation including how to run the tool, train and test customized models and usage of each script is included.

We appreciate your valuable feedback and have updated the documentation and demos to enhance usability for users.

8) Figure 1C and Figure 4A/ Figure 5b show similar concepts/results. In fact, Figure 1C could be considered as a graphical abstract. I suggest including some sort of summary of "the atlas of perturbated profiles" or a summary of the resource which I think it is essential to make it public.

Reply 3-8: Thank you for your advice. Figure 4A illustrates the process of predicting and experimentally validating new compounds for treating SCLC and CRC based on known sensitive compounds. Figure 5B demonstrates how new candidate drugs are screened based on disease signatures. They have different input and reference screen signatures. And Figure 1C serves as a graphical abstract and highlights important concepts similar to those presented in Figure 4A and Figure 5B. We also have added a graphical abstract of "the atlas of perturbated profiles" in Figure 1D to provide a comprehensive summary of the resource.

Revised version:

Figure 1D

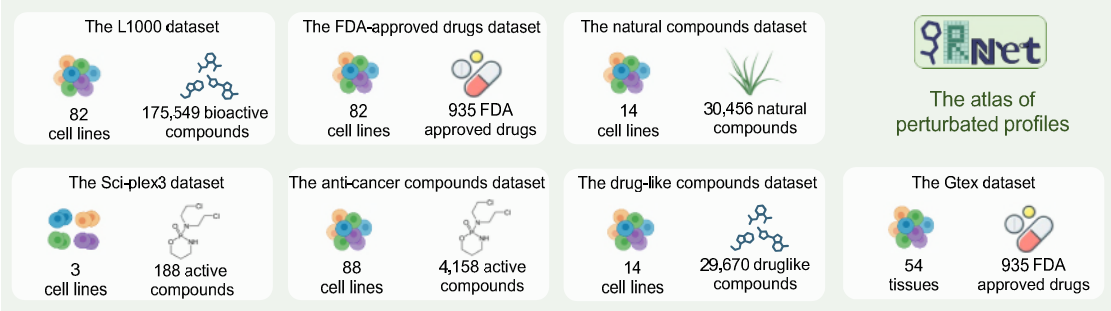


Fig. 1 (D) The large-scale integration atlas of perturbation profiles, including the L1000 dataset, the Sci-plex3 dataset, the FDA-approved drugs dataset, the anti-cancer compounds dataset, the natural compounds dataset, the drug-like compounds dataset, and the Gtex dataset.250

Minor points

Line 291: You might want to mention figure 2H.

Reply 3-9: Thank you for your advice. We indeed intended to reference Figure 2H in Line 291. However, during the revision process, we removed Figure 2D to Supplementary Fig. 6, which caused the original Figure 2H to be renumbered as Figure 2G. We apologize for any confusion this may have caused.

Initial version:

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with BRD4770 produces a consensus trajectory. In which we can observe the pseudodose trajectory, indicating that BRD4770 induced homogeneous responses.

Revised version:

lines 358-362 on page 9

The t-SNE embedding (Fig. 2 G) in the latent space of MCF7 cells treated with AG-14361 produced a consensus trajectory, in which we can observe the pseudo-dose trajectory, indicating that AG-14361 induced heterogeneous responses. Several other pseudo-dose trajectory examples were illustrated in Supplementary Fig. 6 D-F.

Line 313: "Fig 3b investigated in detail .." Please rephrase the sentence figures do not investigate.

Reply 3-10: Thank you for your advice. We have revised in the manuscript, which is copied below:

Initial version:

Fig. 3 B investigated in detail the predicted and actual distributions of fold changes in gene expression of three representative cell lines from different organs (HT29: colorectal adenocarcinoma, A549: lung adenocarcinoma and MCF7: breast adenocarcinoma) treated by vorinostat.

Revised version:

lines 392-396 on page 10

Fig. 3B shows a detailed comparison between the predicted and actual distributions of fold changes in gene expression for three representative cell lines from different organs (HT29: colorectal adenocarcinoma, A549: lung adenocarcinoma and MCF7: breast adenocarcinoma) treated with vorinostat.

Paragraphs in line 400 and line 436. Please include IC50 etc in a supplementary table to make the reading easier.

Reply 3-11: Thank you for your advice. All IC50 values have been compiled into Supplementary Table 3. Figure 3 in the manuscript also displays the IC50 values for the screened drugs. We also adjusted the description to enhance readability (lines 479-482 on page 12 and lines 516-518 on page 13).

Supplementary Table 3 Antiviability activities of compound candidates against SCLC and CRC cell lines

Cancer Type	Compound Name	Cell line	IC_{50}(μmol/L)
SCLC	(+)-Fangchinoline	NCI-H69	3.745 ± 0.184
		NCI-H526	4.402 ± 0.282
		NCI-H446	11.560 ± 0.607
		DMS114	6.406 ± 0.859
		NCI-H209	11.633 ± 0.387
		NCI-H196	6.566 ± 0.578
SCLC	SEL120-34A HCl	NCI-H69	3.701 ± 0.074
		NCI-H526	5.003 ± 0.168
		NCI-H446	5.381 ± 0.108
		DMS114	5.331 ± 0.164
		NCI-H209	4.641 ± 0.183
		NCI-H196	4.555 ± 0.331

CRC	7-Methoxyrosmanol	HT29	33.277 ± 0.760
		SW480	12.627 ± 0.475
		Caco-2	21.850 ± 4.398
		SW620	14.460 ± 2.219
		HCT116	20.323 ± 2.928
		Colo205	23.967 ± 4.248
CRC	Mulberrofuran Q	HT29	27.940 ± 2.248
		SW480	43.807 ± 5.628
		Caco-2	15.490 ± 1.143
		SW620	37.740 ± 1.189
		HCT116	37.260 ± 0.465
		Colo205	46.257 ± 2.792

Revised version:

lines 479-482 on page 12

(+)-Fangchinoline and SEL120-34A HCl moderately inhibited the viability of SCLC cell lines. Detailed antiviability activities of two candidates are shown in Supplementary Table 3.

lines 516-518 on page 13

7-Methoxyrosmanol and Mulberrofuran Q moderately inhibited the viability of CRC cell lines. Detailed antiviability activities of two candidates are shown in Supplementary Table 3.

Reviewer #3 (Remarks on code availability):

More documentation and good demos to run the tool available in Github are needed:

- The installation is simple, but for someone without basic Python knowledge it might lack details.

- There is no documentation on how to run the tool, there are several separate scripts and it is not intuitive.

- The demos cannot be tested and more details are needed: The dataset files are missing. It is not clear where they come from. Also, within the scripts they are named completely different from the reference.

- Also specify if these files need to be preprocessed before using the demos.

Reply 3-12: Thank you for your feedback regarding the documentation and demos available on GitHub. We have added more details to the documentation and demos in Github (https://github.com/Perturbation-Response-Prediction/PRnet).

1) Installation Instructions: We have added more details to documentation to provide more comprehensive instructions, including a guide to install dependencies by Anaconda or Dockerfile.
2) Reqired Datasets: All required datasets can be accessed in the online webset (http://prnet.drai.cn/Statistics/). A demo dataset used to verify the environment configuration been uploaded to GitHub
3) Data Preprocessing: Detailed instructions and examples to guide users through the necessary data preparation processes have been added.
4) Running the Tool: A documentation including how to run the tool, train and test customized models and usage of each script is included.

We appreciate your valuable feedback and have updated the documentation and demos to enhance usability for users.

Reviewer #1 (Remarks to the Author):

Thank you for addressing my concerns. This work will be a significant contribution to the new field of single-cell perturbation modeling.

Reviewer #2 (Remarks to the Author):

The revised manuscript addressed my concerns thoroughly. I especially appreciate the authors' effort to go at length to set up a website with intuitive GUI to share the model and the results, and for broad readership to use the tool. This will be a nice platform that has utility for drug discovery communities. I hope the accuracy of the prediction will stand the test of time.

Reviewer #2 (Remarks on code availability):

The code is adequately written.

Reviewer #3 (Remarks to the Author):

In my opinion, the authors have satisfactorily addressed all referees' questions and suggestions. The new results with patient data and making it publicly available have enriched the scientific value of the study. In my opinion the manuscript has also improved considerably with all the referee´s comments and is now more direct and clearer in explaining the main contribution of the paper. It has also improved the introduction and discussion. In addition, the software now is installable and has been documented in more depth including demos.

Reviewer #3 (Remarks on code availability):

The code is available and now includes demos, instructions for installing and oducmentation.

Reviewer #1 (Remarks to the Author):

Thank you for addressing my concerns. This work will be a significant contribution to the new field of single-cell perturbation modeling.

Reply: We sincerely appreciate your positive feedback and acknowledgment of the contribution our work makes to the emerging field of single-cell perturbation modeling. Your valuable comments were instrumental in reinforcing the clarity and significance of our study.

Reviewer #2 (Remarks to the Author):

The revised manuscript addressed my concerns thoroughly. I especially appreciate the authors' effort to go at length to set up a website with intuitive GUI to share the model and the results, and for broad readership to use the tool. This will be a nice platform that has utility for drug discovery communities. I hope the accuracy of the prediction will stand the test of time.

Reviewer #2 (Remarks on code availability):

The code is adequately written.

Reply: We are grateful for your positive and thoughtful review. We are especially grateful for your recognition of our efforts to make the model and its results accessible to the broader scientific community through a user-friendly platform. We believe this platform will serve as a valuable resource for drug discovery researchers, and we hope the accuracy of our predictions will continue to stand up to future scrutiny.

Reviewer #3 (Remarks to the Author):

In my opinion, the authors have satisfactorily addressed all referees' questions and suggestions. The new results with patient data and making it publicly available have enriched the scientific value of the study. In my opinion the manuscript has also

improved considerably with all the referee´s comments and is now more direct and clearer in explaining the main contribution of the paper. It has also improved the introduction and discussion. In addition, the software now is installable and has been documented in more depth including demos.

Reviewer #3 (Remarks on code availability):

The code is available and now includes demos, instructions for installing and documentation.

Reply: We sincerely appreciate your detailed and constructive feedback, which has significantly enhanced both the scientific depth and clarity of the manuscript. Your insightful suggestions particularly regarding the addition of patient data, exploration of inter- and intra-heterogeneity, and improvements to the software documentation, have substantially enriched the study and increased its accessibility to a broader audience. Thank you again for your advice.