

<https://doi.org/10.1038/s41698-024-00583-0>

Overcoming limitations in current measures of drug response may enable AI-driven precision oncology

Check for updates

Katja Ovchinnikova¹, Jannis Born², Panagiotis Chouvardas^{1,3}, Marianna Rapsomaniki²✉ & Marianna Kruithof-de Julio^{1,3}✉

Machine learning (ML) models of drug sensitivity prediction are becoming increasingly popular in precision oncology. Here, we identify a fundamental limitation in standard measures of drug sensitivity that hinders the development of personalized prediction models – they focus on absolute effects but do not capture relative differences between cancer subtypes. Our work suggests that using z-scored drug response measures mitigates these limitations and leads to meaningful predictions, opening the door for sophisticated ML precision oncology models.

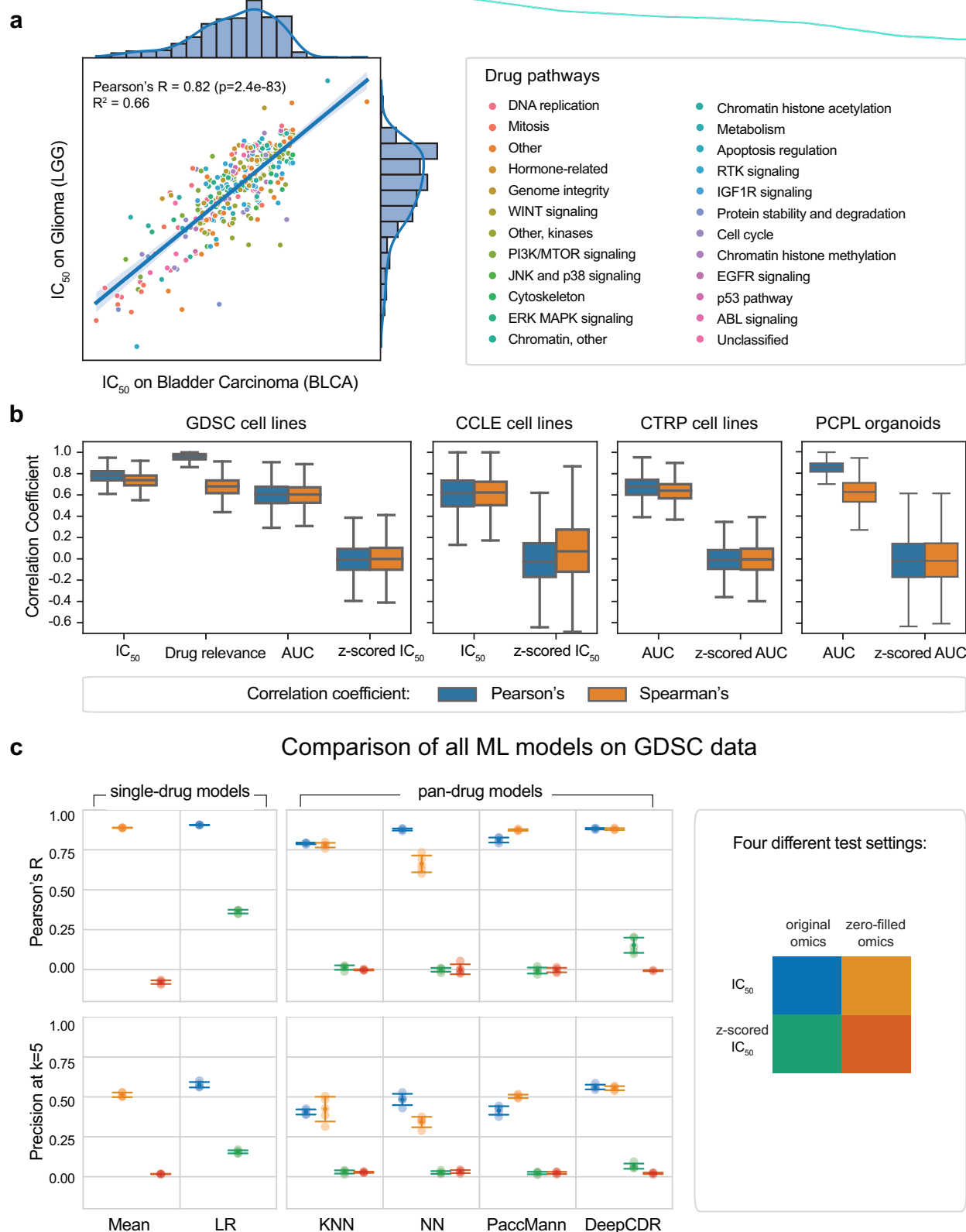
Precision oncology aims at data-driven identification of personalized treatments for cancer patients. Drug sensitivity tested on cell lines or organoids has shown great potential for predicting therapy success^{1,2}. In this approach, cancer cell lines or organoids are exposed to a large collection of anti-cancer compounds at different concentrations in vitro. The cancer cell survival rate is then used to derive measures of drug efficacy, such as the half maximal inhibitory concentration (IC₅₀) or the area under the dose-response curve (AUC)³. With the ever-increasing availability of large data resources containing drug sensitivity measurements and paired omic profiles across hundreds of cell lines^{4–7}, machine learning (ML) models have emerged as a promising approach towards predicting drug response^{8–13}. ML models for drug response prediction typically integrate omic data from cancer cell lines with drug profiles to predict drug sensitivity, as measured by IC₅₀ or AUC^{13,14}.

Several studies have so far addressed open questions on how to train ML models for drug response prediction. Notably, Sharifi-Noghabi et al.¹⁴ carried out a systematic study on the comparative performance of several ML models when trained and tested on the most popular cell line datasets to predict different measures of drug response. In agreement with previously reported striking discordances between two large pharmacogenomic datasets¹⁵, namely CGP⁶ and CCLE⁵, cross-domain generalization issues that question the application of ML models in clinically relevant tasks have been reported¹⁴. The use of IC₅₀ as a proxy of therapeutic efficacy has also been considerably debated^{14–17}, as IC₅₀ indicates potency and not necessarily clinical outcomes. Additionally, IC₅₀ is highly dependent on the cell division rate¹⁷ and the overall drug toxicity¹⁸. The latter suggests that when comparing drugs by IC₅₀ values alone, more toxic drugs would be unnecessarily

prioritized, hindering personalized predictions. To address these issues, alternative scores have been proposed. The AUC, a metric that is independent of the dose and captures the cumulative effect of the drug, is less related to mere potency and was reported to better explain systematic variation in cancer drug response¹⁶. Other alternative scores include the drug relevance score (ratio of the drug's IC₅₀ to its maximum therapeutic dosage) that approximates the therapeutic index used in drug development¹⁸, the normalized version of AUC (i.e., the ratio of the drug's AUC to the maximum area for the concentration range)¹⁹, the normalized growth rate inhibition (GR)¹⁷ that compares growth rates in the presence and absence of drug, accounting for confounding effects of division rate, the activity area (AA)⁵ that reflects both drug efficacy and potency, and the drug sensitivity scoring (DSS)²⁰ that integrates multiple dose-response relationships in cancer and control cells. Still, the vast majority of ML models for drug response prediction are trained to predict IC₅₀ or AUC^{13,14}.

Motivated by the above, here we investigate how standard measures of drug response impact ML models of precision oncology. We used GDSC, the largest resource that forms the basis of most current ML models, containing IC₅₀ values of hundreds of anticancer drugs in thousands of cancer cell lines⁴. We first observed that IC₅₀ drug response profiles were very highly correlated even between cell lines of distinct origins (example of a bladder carcinoma and a glioma cell line shown in Fig. 1a). We extended this analysis to all pairs of cell lines in GDSC and found that this effect was omnipresent (i.e., consistent across all cancer subtypes; Fig. 1b, Supplementary Fig. 1), local (i.e., pronounced within cell lines of the same subtype; Supplementary Fig. 1), and robust (i.e., consistent across all drug pathways, see Supplementary Fig. 2). Following the same process for two distinct and

¹Urology Research Laboratory, Department for BioMedical Research, University of Bern, Bern, Switzerland. ²IBM Research Europe, Zurich, Switzerland. ³Department of Urology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ✉e-mail: marianna.rapsomaniki@unil.ch; marianna.kruithofdejulio@unibe.ch



widely used cancer cell line datasets, namely the Cancer Cell Line Encyclopedia (CCLE)⁵ and the Cancer Therapeutics Response Portal (CTRP)^{21–23}, corroborated our observations (Fig. 1b). This surprising finding indicates that drug response, as measured by IC₅₀, is largely dependent on the drug and not the cell line it was tested on, suggesting that predicting drug response based on IC₅₀ is a trivial task. Viewing GDSC through the lens

of the *drug*, previous work has observed high correlation across drugs reported as a “general level of drug sensitivity” by Geeleher et al.²⁴ and “general response across drugs” by White et al.²⁵, in part explained by multi-drug resistance. Conversely, our analysis views drug sensitivity prediction through the lens of the *cell* by using drug response to compute pairwise cell line similarity. When repeating our analysis using the drug relevance score¹⁸

Fig. 1 | Different measures of drug response on cell line datasets. a A scatterplot of IC₅₀ values of all drugs tested on two cell lines (Bladder Carcinoma and Glioma) indicates a high correlation between the corresponding IC₅₀ drug response profiles. **b** Boxplots of pairwise Pearson and Spearman correlation coefficients for different measures of drug response (e.g., IC₅₀, AUC), computed for each pair of GDSC, CCLE and CTRP cell lines and PCPL organoids. For the GDSC boxplots, the correlation coefficient was computed across all tested drugs shared across a cell line pair, which made up on average 81% ± 19% of all drugs. In all box plots, center line corresponds to the median, box limits to the upper and lower quartiles and whiskers to 1.5x interquartile range. **c** Performance of different ML models of drug sensitivity

prediction using the Pearson correlation coefficient between observed and predicted drug response measures (top) and the Precision at $k = 5$ score (bottom) across all cell lines. The dots and errorbars represent five independent cross-validation folds, and corresponding standard deviation, respectively. KNN k-nearest neighbor, NN neural network, Mean mean baseline, LR linear regression. The pan-drug models were tested using both original and z-scored IC₅₀ values, with original or zero-filled omics feature vectors, resulting in a total of four test setting combinations, as indicated in different colors in the legend. The mean baseline model was evaluated without omics data and the linear regression model was evaluated with omics data only.

(IC₅₀/max concentration) and the AUC, we confirmed our observation, even if pairwise correlation coefficients were slightly decreased (Fig. 1b, left). Finally, we followed the same process using PCPL, a pancreatic cancer patient derived organoid (PDO) library with associated genomic, transcriptomic, and therapeutic profiling¹⁹. Using the normalized AUC still leads to the same problem of elevated correlation coefficients as observed in GDSC (Fig. 1b, right). Together, our findings point to a fundamental issue when using these measures for personalized drug response prediction from cell line or organoid data: drug response is heavily affected by the inherent potency or toxicity of each drug independently of the cell line it was tested on.

We next explored how this issue affects the performance of ML drug sensitivity prediction models. We first implemented single-drug models, including a mean baseline model that assigns to each test sample the mean drug response of that drug from the train set, and a linear regression (LR) model on selected genes (Methods). We compared these baseline models with four pan-drug ML models: (i) a bimodal k-Nearest Neighbor (kNN) model²⁶, (ii) a fully connected neural network (baseline NN), (iii) PaccMann, a multi-modal attention-based neural-network model²⁷ and (iv) DeepCDR, a hybrid graph convolutional network²⁸ (Methods). To mitigate the reported limitations, we applied a z-score normalization for all IC₅₀/AUC values separately for each drug, thus removing the drug-specific bias and dominance of toxic compounds. Indeed, as expected, correlation coefficients computed for z-scored IC₅₀/AUC values are now reduced and centered around zero (Fig. 1b). To test the contribution of omics in the prediction, we tested all models with normal or zero-filled omics feature vectors, resulting in a total of four test settings (Methods and Fig. 1c, legend). All settings were tested in a cross-validation fashion using both actual and z-scored IC₅₀ values. Our results clearly demonstrate that even the least sophisticated models can accurately predict IC₅₀ values (Fig. 1c, top). Model performance is not affected in the zero-filled omic setting, suggesting that indeed, the predictions solely rely on drug profiles and are indifferent to molecular properties of the cell lines (Fig. 1c, top). Strikingly, the mean baseline model that simply outputs the average IC₅₀ without considering the omic profiles achieves comparable results to more sophisticated ML models. Predicting the z-scored IC₅₀ values appears to be a much more challenging task, with all pan-drug ML models largely failing in all settings (Supplementary Data 1–3). Repeating the same process using the precision at k score (Methods, Fig. 1c, bottom) and a rank-based loss function¹⁸ (Supplementary Fig. 3) is insufficient to mitigate the issue, indicating that our results are consistent across performance assessment scores. This can be explained by considering that z-scoring removes the per-drug systematic variation from the IC₅₀ measurements, and thus predicting the z-scored IC₅₀ values implies learning to predict how cell lines respond to drugs relative to an average cell line. Overall, we conclude that, although ML models can accurately predict IC₅₀, these predictions are not in reality personalized but rather merely driven by drug features that are universal across all cancer types. In other words, established drug response metrics promote learning absolute effects of drugs while neglecting relative differences between cell lines thus counteracting a vision of precision oncology where subtleties in biological signatures drive treatment decisions. This can be mitigated through the usage of response metrics that emphasizes relative differences such as the proposed z-scored IC₅₀.

To better understand this effect, we looked deeper into the PCPL pancreatic organoid dataset and visualized drug response of the top three drugs in terms of AUC (Fig. 2a, top) and z-scored AUC (Fig. 2a, bottom) across all organoids (results on all tested drugs are given in Supplementary Fig. 4). We also predicted drug rankings across all tested organoids based on its AUC and z-scored AUC value (Fig. 2b top and bottom, respectively). A first striking observation is that Bortezomib has the lowest AUC and is thus the top-ranking drug in almost all tested organoids. Disulfuram follows closely and ranks second in almost 80% of all tested organoids, followed by SN-38. Conversely, z-scored AUC values appear to be organoid-specific, with different drugs being highly effective for different organoids (Fig. 2a, b, bottom), suggesting they can potentially be used for personalized drug recommendations. To assess this, we evaluated all previous models for predicting both the AUC and z-scored AUC in a full and zero-filled omics setting (Methods). We first employed a zero-shot inference setting, i.e., we trained pan-drug models on the cell line data (GDSC) and tested them on the PCPL data (Fig. 2c, left). All pan-drug models reached satisfactory performance when predicting the AUC without additional re-training. However, the models again relied only on drug profiles: zero-filling the omics data did not diminish performance, the simplistic mean baseline model outperformed all ML models, and all models failed in the z-scored setting (Supplementary Data 4–6). The single-drug linear regression model performs slightly worse than the pan-drug models, suggesting that it is highly dependent on the gene selection. We then trained and tested all models on the PCPL dataset in a 10-fold cross validation setting (Fig. 2c, right). As expected, pan-drug models reach higher overall performance, but are again dependent only on the drug profiles. Interestingly, this time the linear regression model achieved accurate predictions in both the AUC and z-scored AUC setting, both in terms of Pearson's correlation coefficient (Fig. 2c, right) and Precision at k (Supplementary Fig. 5). This encouraging finding suggests that, under appropriate considerations, predicting drug sensitivity from omics data is feasible even with simplistic models.

Motivated by this, we further investigated genes important for the linear regression predictions. We first removed multicollinearity in RNAseq data, resulting in a total of 6000 linearly independent genes (Methods). By computing the prediction performance in terms of average Pearson's correlation coefficient for varying number of selected genes (Fig. 2d), we see that the prediction quality is stable when the number of relevant genes varies between 100 and 1000 and reaches a maximum value of around 0.93 for approximately 350 genes on average, a pattern consistent across all tested drugs (Supplementary Fig. 6). Using the genes selected for each drug as features, we created a 2D embedding map using UMAP²⁹ and visualized the AUC of a selected drug (Paclitaxel) on that embedding (Fig. 2e). We observe that dataset and drug-specific gene selection allows for a straightforward linear separation of the organoids in terms of their AUC, whereas genes selected based on the literature do not have the same effect. This observation is true for all tested drugs (Supplementary Fig. 7) and explains the high performance of the linear regression model: unlike pan-drug models trained on GDSC, the linear model is specific to the PCPL dataset, which could be an advantage in a disease-specific clinical setting. To further assess the performance of the linear regression model, we created scatterplots of predicted vs. observed z-scored AUC values per drug (Supplementary Fig. 8) and per organoid (Supplementary Fig. 9), which indicated consistent results at an

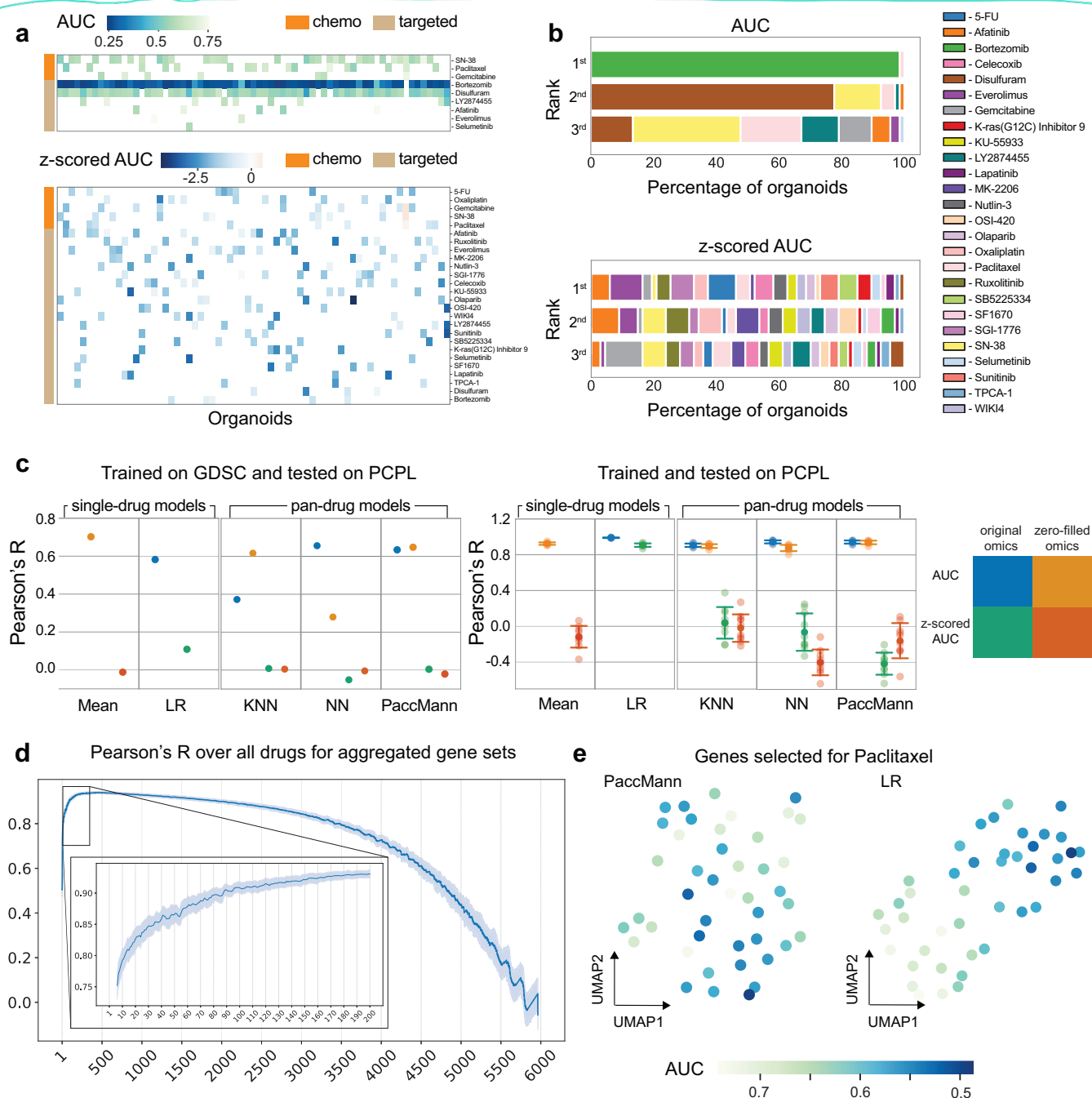


Fig. 2 | Results on pancreatic organoid data. a Heatmap of top three drugs as per AUC score (top) and z-scored AUC score (bottom) for each PCPL organoid (column). Smaller/higher AUC values indicate higher/lower efficacy of the drug in the tested organoid; smaller/higher z-scored values indicate higher/smaller efficacy related to the average efficacy of that drug across all organoids. **b** Stacked bar plots indicating the percentage of PCPL organoids in which each drug ranked 1st, 2nd, or 3rd, as per AUC (top) or z-scored AUC (bottom). **c** Pearson's R correlation coefficient between predicted and observed drug response values computed per organoid for each of the evaluated models. All models were either trained on GDSC

and tested on the PCPL dataset (left) or trained and tested on the PCPL dataset only (right). In the latter, error bars indicate the standard deviation across 10 cross validation folds. **d** Pearson's R correlation coefficient between predicted and observed AUC values for varying number of considered genes aggregated over all drugs. **e** UMAP embedding of the RNAseq organoid data using genes selected by PaccMann (left) and the drug-specific gene selection scheme of the linear regression model (right) for Paclitaxel. Here, each dot corresponds to an organoid, and the color of the dot reveals the AUC values of Paclitaxel on that organoid.

individual drug or organoid level. Together, our results suggest that personalized drug response prediction from omics measurements is indeed feasible and opens the door for more sophisticated ML models to be applied on the same single-drug, z-scored setting.

In conclusion, in this work we expose fundamental issues with the use of common drug response measures that hinder the application of ML

models for pan-cancer personalized drug response prediction. Specifically, we show that popular ML pan-drug sensitivity prediction models do not base their predictions on the molecular features and fail to make truly personalized predictions. Although using z-scored IC₅₀/AUC values mitigates this issue by reformulating the drug response prediction into prediction of drugs to which a patient responds better or worse than an average

patient, training models in a z-scored setting is a challenging task for all pan-cancer ML models. Conversely, we show that even simple models may learn from omics data, performing exceptionally well when trained in a drug-specific manner on a well-defined disease setting. Future efforts in developing personalized drug sensitivity prediction models need to address limitations associated with IC_{50} /AUC and consider adopting response metrics with drug-specific transformation schemes, such as z-scoring. A key challenge on the road toward developing improved pan-cancer models is to equip them with inductive biases to promote omics features. This enhancement can facilitate predictions for previously unseen drugs while maintaining the ability to provide personalized treatment recommendations. Finally, the use of organoids to predict responses to cancer drugs represents a significant stride forward in precision medicine, as supported by recent studies^{30–33}. The potential to tailor treatments based on individual organoid responses holds considerable promise for improving the efficacy of cancer therapies while minimizing adverse effects. As progress unfolds in this field, predictions of drug responses derived from organoids may become an integral component of personalized cancer treatment strategies.

Methods

Datasets

Cell line data were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database⁴ (we used the GDSC1 data), the Cancer Cell Line Encyclopedia (CCLE) dataset⁵ and the Cancer Therapeutics Response Portal (CTRP) dataset^{21–23} (details in Data Availability). For GDSC, the drug screen results are available in terms of viability per dosage and were used to obtain IC_{50} and AUC values. All ML models except DeepCDR were evaluated on 953 GDSC cell lines and 283 drugs. Since DeepCDR was evaluated using both transcriptomic and genomic data, we trained and tested it with the original GDSC subset provided in the DeepCDR repository (<https://github.com/kimmo1019/DeepCDR>) which contained 561 cell lines and 238 drugs. The pancreatic cancer dataset was obtained from the Pancreatic Cancer PDO Library¹⁹ (PCPL) containing RNA expression data for 43 organoids that were screened against 26 drugs. The drug efficacy was measured and reported in terms of AUC values.

Computation of correlation coefficients

Let $(x_{i,j}, x_{i,k})$ denote the drug responses of drug $i \in [1, n]$ in cell lines $j, k \in [1, m]$ respectively, reported as e.g., the IC_{50} , AUC or drug relevance score. Then, the Pearson correlation coefficient $r_{j,k}$ between cell lines j, k and across all drugs n is computed as:

$$r_{j,k} = \frac{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)(x_{i,k} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)^2 \cdot \sum_{i=1}^n (x_{i,k} - \bar{x}_k)^2}}$$

where \bar{x}_i and \bar{x}_j correspond to the mean drug response of cell line j and k respectively, across all drugs. We note that, as it is often the case that not all drugs were tested in all cell lines, only the common drugs between the two cell lines are used for the above computation. Similarly, the Spearman correlation coefficient is computed for each pair of cell lines across all drugs by considering the rank values of x .

Z-score transformations

The z-score transformations of the drug responses $x_{i,j}$ were computed for each drug across all cell lines it was tested on, or specifically:

$$Z_{i,j} = \frac{x_{i,j} - \bar{x}_i}{\sigma_i}$$

where \bar{x}_i , σ_i correspond to the mean and standard deviation of the drug response of drug i across all cell lines it was tested on, respectively. For GDSC, we used the precomputed z-scored values as provided in the GDSC

data, but we independently validated that indeed they were performed across all cell lines.

Drug sensitivity prediction from GDSC

Pan-drug models. We first tested several pan-drug prediction models that integrate molecular features (genomic or transcriptomic data of cell lines/organoids) with chemical structural properties of drugs and can thus predict drug response for both seen and unseen drugs. In this approach, the whole dataset is employed to train one prediction model. The following models were considered, in increasing complexity: (i) a baseline bimodal **k-Nearest Neighbors (kNN)** model proposed by Born et al.²⁶ which predicts drug response of a test sample (cell line or organoid) as the average of the drug responses of the k nearest samples in a bimodal similarity space defined using an inverted Tanimoto similarity of Morgan fingerprints as drug-drug distances³⁴ and the Euclidean distance for gene expression features. (ii) a **baseline neural network (NN)** consisting of 3 fully connected hidden layers and a linear output unit as used by Prasse et al.¹⁸ The network takes as input a concatenation of encoded SMILES and gene expression features, and outputs a drug response value. The network is trained using MSE of the real vs. the predicted IC_{50} as a loss function. (iii) **PaccMann**, a multi-modal attention-based neural-network model with state-of-the-art performance in IC_{50} prediction from transcriptomics data²⁷. PaccMann relies on SMILES, and uses contextual attention layers to merge information across both modalities. PaccMann incorporates prior knowledge about targets of drugs present in GDSC and protein-protein interactions and reduces the dimensionality of the gene expression data to 2128 genes. Similarly to the baseline NN, PaccMann employs an **MSE loss function**. We used the PaccMann implementation from https://github.com/PaccMann/paccmann_predictor. In a recent study¹⁸, PaccMann has been modified to predict ranking of drugs rather than IC_{50} values by replacing the MSE loss with a normalized discounted cumulative gain (NDCG). This modification was shown to outperform the reference models with respect to the drug ranking task. We tested this model with the implementation available from https://github.com/Pascalversen/mlmed_ranking. (iv) **DeepCDR**, a hybrid graph convolutional network consisting of a uniform graph convolutional network (UGCN) representing chemical drug features and multiple subnetworks representing different types of omics profiles²⁸. UGCN relies on the adjacent information of atoms in a drug and aggregates the features of neighboring atoms together. The subnetworks generalize over features of cell line omics profiles. The high-level drug and omics features are then concatenated and used as input for a final prediction. DeepCDR utilizes gene expression, genomic mutation, and DNA methylation data and only considers 697 genes from COSMIC Cancer Gene Census (<https://cancer.sanger.ac.uk/census>). The network is trained using MSE as a loss function. For a fairer comparison of DeepCDR with other models, we initially tested only with transcriptomics data and found that it predicted the same drug response for each cell line. We thus decided to use DeepCDR with all omics data available in GDSC. We used the DeepCDR implementation available at: <https://github.com/kimmo1019/DeepCDR>.

The pan-drug models were evaluated using as input drug features and omics data (transcriptomics for KNN, baseline NN, and PaccMann; genomics, transcriptomics, and epigenomics for DeepCDR) or drug features with zero-filled omics feature vectors. For both above cases, the models predicted either actual or z-scored IC_{50} values, obtained by subtracting from each observed value the mean and dividing by the standard deviation, as computed across all cell lines/organoids on that drug. Overall, this resulted in a total of 4 test settings (matrix legend in Fig. 1c), namely IC_{50} with original omics (blue), IC_{50} with zero-filled omics (light orange), z-scored IC_{50} with original omics (green), z-scored IC_{50} with zero-filled omics (dark orange).

Single-drug models. Single-drug models were trained separately for each drug; each single-drug model was blind to drug response of cell lines

or organoids to other drugs. As a baseline for single-drug models, we tested a simplistic mean baseline model that computes, for each drug, a mean value of drug responses in the train set and uses this mean value as a prediction for all test samples. The mean baseline model only uses as input drug information and is blind to omics data. We then tested a total of 8 standard regression models, namely: k-Nearest Neighbors (KNN), Linear Regression (Ridge with regularization parameter $\alpha=10$), Support Vector Regression (with a linear, radial basis function, and polynomial kernel), Decision Tree, Random Forest, Multi-layer Perceptron. For linear models (Linear Regression, linear Support Vector Regression), linearly independent genes were selected, so that no two genes are linearly correlated with Pearson's R correlation coefficient smaller than 0.6 and a p-value smaller than 0.05. This was done by iteratively performing agglomerative clustering based on 1-Pearson-distances (distance threshold between clusters = 0.4, complete linkage), selecting one representative gene from each cluster, and repeating while the clusters contain more than one element. The best results were obtained with a linear regression model with genes selected using Pearson's correlation relevance score (on average 770 genes selected per drug), which is what we report in the manuscript (Fig. 1c). The linear regression model was trained and tested with omics features only. We used a standard 5-fold cross validation scheme by splitting the dataset into 5 equal sized folds so that the folds had no overlapping cell lines. All models were implemented using scikit-learn (<https://scikit-learn.org>).

Drug sensitivity prediction from PCPL

For the prediction of AUC per drug for the PCPL organoids, we used: (i) single-drug models (linear regression with selected features using Pearson's correlation relevance score) trained on PCPL only, (ii) pan-drug models trained on GDSC and tested on PCPL, and (iii) pan-drug models trained and tested on PCPL only. DeepCDR was excluded from the evaluation, as it requires genomic features that are not available for PCPL. Pan-drug models were applied to predict response to 14 drugs present both in GDSC and PCPL, whereas single-drug models predicted response to all 26 PCPL drugs. We used a 10-fold cross validation to evaluate the PCPL models. For the last set of experiments, where the pan-drug models were trained on GDSC cell lines and tested on PCPL organoids, no cross-validation was performed; instead, the full GDSC dataset was used solely for training, while the PCPL dataset was used for testing. To make RNAseq data comparable across the two datasets (GDSC and PCPL), we z-scored each dataset separately and then z-scored the merged dataset one more time.

Gene selection

To quantify the dependence of each gene in the gene expression data to each drug, we computed a set of the following 5 scores between gene expression level and drug response across all cell lines or organoids screened with that drug: (i) Pearson's correlation coefficient and (ii) Spearman's R correlation coefficient, (iii) mutual information score³⁵, (iv) sum of squared residuals of the least squares polynomial fit of the second degree, (v) coefficients of linear SVM regression predicting drug response based on gene expression level. We then used 100 thresholds on each relevance score for filtering out irrelevant genes with the goal to determine which combination of score, threshold, and model provides the best performance for each drug based on average Pearson's correlations between observed and predicted drug response per cell line/organoid. The experiments were conducted in a cross-validation setting.

Performance assessment scores

To evaluate the performance of all models we used four following scores:

- (i) the Pearson correlation coefficient between observed and predicted drug response measures, as computed separately per patient and per drug;
- (ii) the mean squared error (MSE) of the prediction as computed separately per patient and per drug.
- (iii) Precision at k , i.e., the percentage of correct predictions of top- k drugs being ranked according to the selected scores (IC_{50} , IC_{50} z-score, AUC, AUC z-score).

- (iv) the normalized discounted cumulative gain at k , i.e. the sum of relevance values of items occupying the first k ranks normalized by the measure of the ideal ground truth ranking, for more details see (Prasse et al.)¹⁸.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

In this study we used the following datasets: (i) GDSC1, available from the Genomics of Drug Sensitivity in Cancer portal (https://www.cancerrxgene.org/downloads/drug_data?screening_set=GDSC1); drug pathways were downloaded from the same link under Preview: drugs included in download (.csv), (ii) CCLE, available from DepMap (<https://depmap.org/portal/download/all/>), (iii) CTRP (v2), available from Rees et al.²³ as Supplementary files (Supplementary Datasets 1–3), and (iv) PCPL, available from Tiriach et al.¹⁹ (Supplementary Table S4).

Code availability

The code related to this study, including figure-generating scripts, is available under an open-source license at: https://github.com/Urogenus/GDSC_Pancreatic_study.

Received: 16 August 2023; Accepted: 22 March 2024;

Published online: 24 April 2024

References

- Geeleher, P., Cox, N. J. & Huang, R. S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* **15**, R47 (2014).
- Karkampouna, S. et al. Patient-derived xenografts and organoids model therapy response in prostate cancer. *Nat. Commun.* **12**, 1117 (2021).
- Partin, A. et al. Deep learning methods for drug response prediction in cancer: Predominant and emerging trends. *Front. Med.* **10**, 1086097 (2023).
- Yang, W. et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2012).
- Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
- Iorio, F. et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740–754 (2016).
- Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
- Azuaje, F. Artificial intelligence for precision oncology: beyond patient stratification. *NPJ Precis. Oncol.* **3**, 1–5 (2019).
- Ballester, P. J. & Carmona, J. Artificial intelligence for the next generation of precision oncology. *NPJ Precis. Oncol.* **5**, 1–3 (2021).
- Rafique, R., Islam, S. M. R. & Kazi, J. U. Machine learning in the prediction of cancer therapy. *Comput Struct. Biotechnol. J.* **19**, 4003–4017 (2021).
- Baptista, D., Ferreira, P. G. & Rocha, M. Deep learning for drug response prediction in cancer. *Brief. Bioinform.* **22**, 360–379 (2021).
- Firoozbakht, F., Yousefi, B. & Schwikowski, B. An overview of machine learning methods for monotherapy drug response prediction. *Brief. Bioinform.* **23**, bbab40 (2021).
- Sharifi-Noghabi, H. et al. Drug sensitivity prediction from cell line-based pharmacogenomics data: guidelines for developing machine learning models. *Brief. Bioinform.* **22**, bbab294 (2021).

15. Haibe-Kains, B. et al. Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
16. Fallahi-Sichani, M., Honarnejad, S., Heiser, L. M., Gray, J. W. & Sorger, P. K. Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat. Chem. Biol.* **9**, 708–714 (2013).
17. Hafner, M., Niepel, M., Chung, M. & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* **13**, 521–527 (2016).
18. Prasse, P. et al. Matching anticancer compounds and tumor cell lines by neural networks with ranking loss. *NAR Genom. Bioinform.* **4**, lqab128 (2022).
19. Tiriak, H. et al. Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discov.* **8**, 1112–1129 (2018).
20. Yadav, B. et al. Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci. Rep.* **4**, 5193 (2014).
21. Basu, A. et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
22. Seashore-Ludlow, B. et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
23. Rees, M. G. et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).
24. Geleher, P., Cox, N. J. & Huang, R. S. Cancer biomarker discovery is improved by accounting for variability in general levels of drug sensitivity in pre-clinical models. *Genome Biol.* **17**, 190 (2016).
25. White, B. S. et al. Bayesian multi-source regression and monocyte-associated gene expression predict BCL-2 inhibitor resistance in acute myeloid leukemia. *NPJ Precis. Oncol.* **5**, 71 (2021).
26. Born, J., Huynh, T., Stroobants, A., Cornell, W. D. & Manica, M. Active site sequence representations of human kinases outperform full sequence representations for affinity prediction and inhibitor generation: 3D Effects in a 1D Model. *J. Chem. Inf. Model.* **62**, 240–257 (2022).
27. Manica, M. et al. Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. *Mol. Pharmaceutics* **16**, 4797–4806 (2019).
28. Liu, Q., Hu, Z., Jiang, R. & Zhou, M. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* **36**, i911–i918 (2020).
29. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
30. Ooft, S. N. et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Transl. Med.* **11**, eaay2574 (2019).
31. Cho, Y.-W. et al. Patient-derived organoids as a preclinical platform for precision medicine in colorectal cancer. *Mol. Oncol.* **16**, 2396–2412 (2022).
32. Wang, T. et al. Patient-derived tumor organoids can predict the progression-free survival of patients with stage IV colorectal cancer after surgery. *Dis. Colon Rectum* **66**, 733 (2023).
33. Narasimhan, V. et al. Medium-throughput drug screening of patient-derived organoids from colorectal peritoneal metastases to direct personalized therapy. *Clin. Cancer Res.* **26**, 3662–3670 (2020).
34. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
35. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, 066138 (2004).

Acknowledgements

This work was supported by the Swiss National Science Foundation (SNSF) Grants 202297 to M.R. and M.K.dJ and 3RCC OC-2019-003 to MKdJ.

Author contributions

Conceptualization: JB, KO, PC. Data curation and formal analysis: KO. Investigation: KO. Methodology: JB, KO, PC. Software, validation and visualization: JB, KO. Funding acquisition: MR, MKdJ. Project administration: MKdJ. Supervision: MR, MKdJ. Writing – original draft: MR, KO. Writing – review & editing: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-024-00583-0>.

Correspondence and requests for materials should be addressed to Marianna Rapsomaniki or Marianna Kruihof-de Julio.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024