



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## **Establishing predictive machine learning models for drug responses in patient derived cell culture**

Downloaded from: <https://research.chalmers.se>, 2026-03-06 12:37 UTC

Citation for the original published paper (version of record):

Abdel-Rehim, A., Orhobor, O., Griffiths, G. et al (2025). Establishing predictive machine learning models for drug responses in patient derived cell culture. *npj Precision Oncology*, 9(1). <http://dx.doi.org/10.1038/s41698-025-00937-2>

N.B. When citing this work, cite the original published paper.

<https://doi.org/10.1038/s41698-025-00937-2>

# Establishing predictive machine learning models for drug responses in patient derived cell culture

Check for updates

Abbi Abdel-Rehim<sup>1</sup> ✉, Oghenejokpeme Orhobor<sup>1</sup>, Gareth Griffiths<sup>2</sup>, Larisa Soldatova<sup>3</sup> & Ross D. King<sup>1,4,5,6</sup>

The concept of personalised medicine in cancer therapy is becoming increasingly important. There already exist drugs administered specifically for patients with tumours presenting well-defined genetic mutations. However, the field is still in its infancy, and personalised treatments are far from being standard of care. Personalised medicine is often associated with the utilisation of omics data. Yet, implementation of multi-omics data has proven difficult, due to the variety and scale of the information within the data, as well as the complexity behind the myriad of interactions taking place within the cell. An alternative approach to precision medicine is to employ a function-based profile of cells. This involves screening a range of drugs against patient-derived cells (or derivative organoids and xenograft models). Here we demonstrate a proof-of-concept, where a collection of drug screens against a highly diverse set of patient-derived cell lines, are leveraged to identify putative treatment options for a 'new patient'. We show that this methodology is highly efficient in ranking the drugs according to their activity towards the target cells. We argue that this approach offers great potential, as activities can be efficiently imputed from various subsets of the drug-treated cell lines that do not necessarily originate from the same tissue type.

A profound shift towards personalised cancer therapies is underway. This transformation is driven by the widespread recognition of the heterogeneity that characterises cancer, which is now considered to be a myriad of diseases united by uncontrolled growth. The diverse array of treatment strategies under development to treat cancer is a testament to the complexity of the disease.

Precision medicine is a transformative approach, tailoring treatments to the individual patients, while considering their unique genetic, environmental and lifestyle factors. This approach offers the potential for enhanced patient care and a reduction in the healthcare burden associated with these diseases. Nevertheless, a significant challenge still lies in the need for improved predictive models that can anticipate patient responses to specific treatments.

Recent breakthroughs in cultivating patient-derived cells in laboratory settings enable the establishment of *in vitro* models as valuable tools for advancing precision medicine. This methodology has the potential to inform on patient responses to treatment, deepening our understanding of underlying disease mechanisms, uncovering novel drug

targets and ultimately enabling the development of more efficient treatment strategies.

Obtaining valuable insights from such models remains a challenge. While omics data has been used to varying degrees of success in this area, it often falls short of providing the versatility required for the diverse spectrum of available drugs and their mechanism of action.

An alternative approach to omics is the utilisation of functional assays<sup>1</sup>. Studies have demonstrated that patient-derived xenografts (PDX) can yield drug responses that correlate with the outcomes of patient responses in clinic<sup>2</sup>. However, the direct application of PDX models for precision medicine is often impractical due to the significant cost, time, and resources necessary to generate results<sup>3</sup>.

Addressing the relationship between *ex vivo* responses and those observed in clinical settings remains a central and ongoing challenge in the realm of precision medicine. While studies correlating patient responses with those observed in patient-derived cell culture (PDC) are limited, they are steadily increasing<sup>4-6</sup>. Notably, certain studies have demonstrated improved outcomes for blood cancer patients treated with drugs selected

<sup>1</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK. <sup>2</sup>ValiRx Plc, Nottingham, UK. <sup>3</sup>Department of Computing, University of London, London, UK. <sup>4</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. <sup>5</sup>Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden. <sup>6</sup>The Alan Turing Institute, London, UK. ✉e-mail: [aar52@cam.ac.uk](mailto:aar52@cam.ac.uk)

based on their ex vivo performance, as opposed to standard treatments<sup>7–9</sup>. Addressing this challenge with vigour is paramount if we are to realise the full potential of this approach.

An essential step involves establishing an efficient methodology for precision medicine grounded in patient-derived cancer cell cultures. Growing cells in organoids provides a closer resemblance to the patient's tumour than conventional tissue culture<sup>10–12</sup>. However, they often lack cells from the tumour microenvironment (TME) and can be costly to grow<sup>4,13</sup>. Whole-tumour cell culture is proposed as another alternative, which includes representation of cells from the TME at a lower cost than conventional organoids<sup>4</sup>. The large array of approved cancer drugs, coupled with numerous FDA-approved drugs available for consideration, presents a complex therapeutic landscape. While the straightforward approach of testing all drugs for all patients may seem ideal, it is prohibitively expensive and resource-intensive, particularly given the substantial number of cancer patients requiring screening.

An alternative, more practical approach is to harness the predictive capabilities of machine learning (ML). Machine learning has been successfully applied in drug selection for a multitude of targets, as evidenced by a large number of publications in the field<sup>14,15</sup>.

Various methods have been developed for the prediction of bioactivities in cell lines and other cellular and molecular targets. The Quantitative Structure Activity Relationship approach, rooted in machine learning, is the most prevalent. It leverages molecular fingerprints based on chemical substructures to establish correlations with bioactivity in specific targets.

A different approach geared towards precision medicine makes use of omics data to predict the activity of drugs in individual patients. This method has demonstrated success for some drugs with clear correlations to specific dysregulated or dysfunctional genes<sup>16,17</sup>. However, a notable challenge remains, as no such method has yet successfully predicted the outcome of a diverse range of drugs across a varied patient population<sup>18,19</sup>.

Here we introduce a promising approach that is rooted in bioactivity fingerprints and high-throughput screening fingerprints, closely related to transformational machine learning (TML)<sup>20–22</sup>. This method relies on historical screening data as descriptors. These descriptors, much like bioactivity fingerprints, consist of bioactivities towards various targets. The key distinction is that a significant portion of these activities are actual

measurements rather than mere predictions, enhancing the reliability and applicability of the methodology.

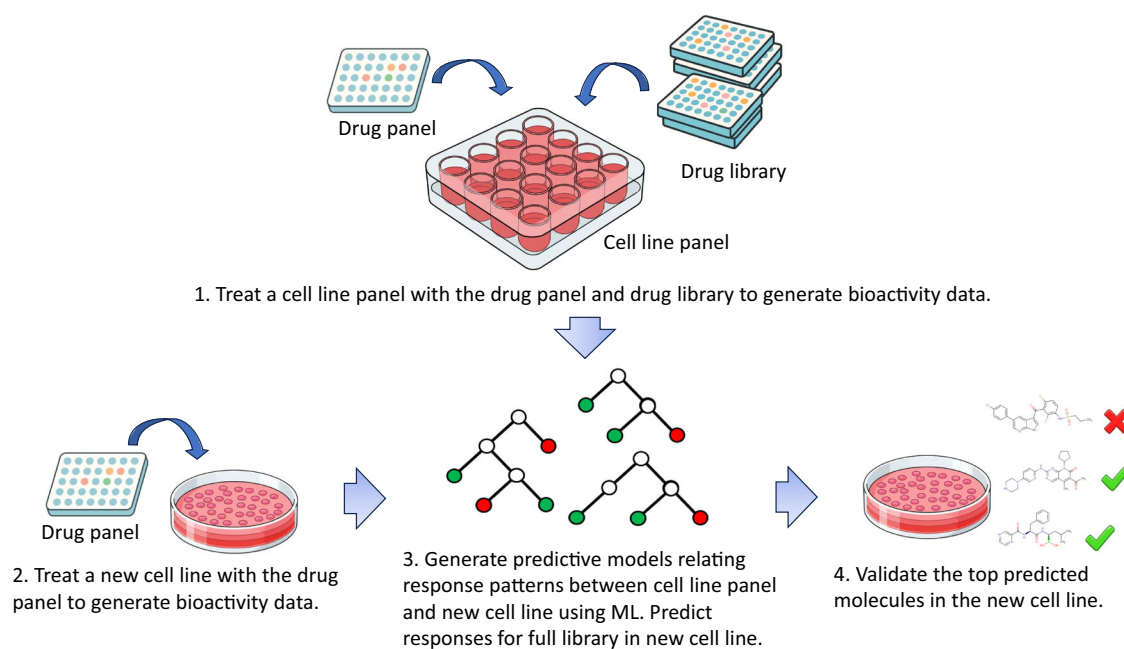
In our methodology, patient samples first undergo a comprehensive screening of a drug library. A subset of this library is then selected as a probing panel. Subsequently, a new patient-derived cell line is screened only against this smaller panel. Machine learning is employed to learn relationships between drug responses found in historical samples and those in the new sample. This model, trained on the responses from the drug panel, is then applied to predict drug responses across the entire library for the new cell line. Experimental validation can then be conducted on the top hits, and any confirmed hits represent potent candidates for a targeted drug cocktail tailored to the patient's cancer (Fig. 1).

In this scenario, we predict drug sensitivities for patient-derived cell lines through analysis of 'historical' profiles of cell lines derived from other patients. Our findings demonstrate the significant efficiency of this approach in predicting drugs that exert a substantial impact on cancer cells. Moreover, we also establish the method's applicability in addressing the considerably more challenging task of identifying highly selective drugs for specific cancer cell lines. We provide additional evidence of the method's utility by applying it to a library of FDA-approved drugs. Finally, this methodology is also applied to a unique dataset with 24 patient-derived tissue cultures screened against a limited panel of 35 FDA-approved drugs shortly after the biopsies were performed.

## Results

To develop an efficient methodology for predicting drug responses in patient-derived cell lines we performed a few initial probing experiments using a dedicated validation set from the GDSC1 dataset (Fig. 2). These provided the foundation for establishing parameters necessary to construct a viable prototype 'recommender system', capable of predicting drug responses in unseen patient cell lines.

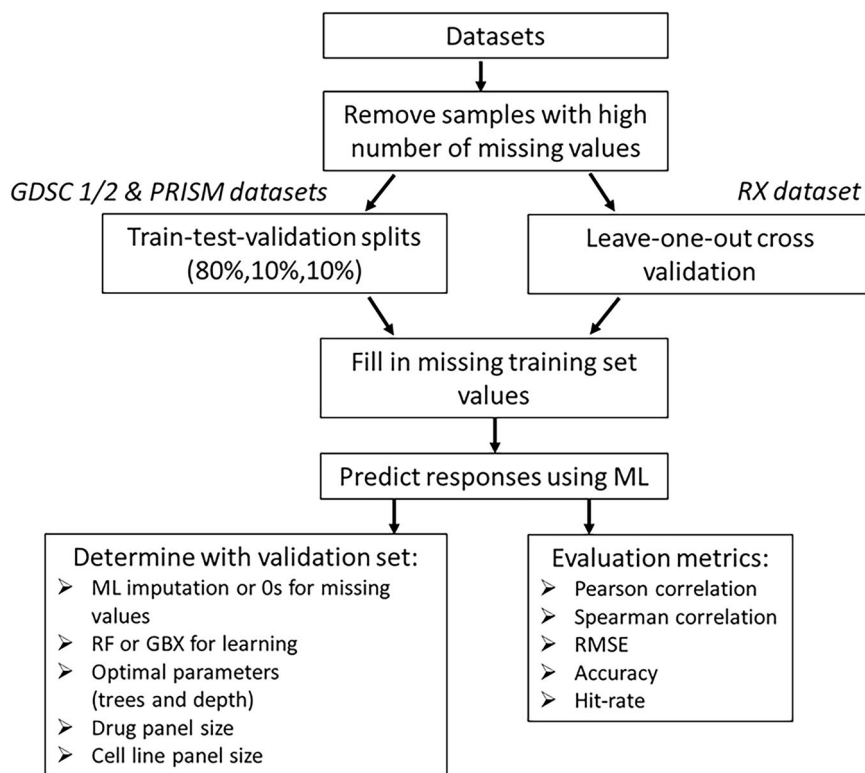
Based on the outcomes of these experiments outlined in supplementary materials, we used TML to fill in missing values in the training datasets (Table S2 and Fig. S1), we employed a random forest with 50 trees (default parameters) (Table S3), we used an initial 'probing' drug panel of 30 selected drugs for all 'unseen' cell lines (Tables S4 and S5), and we included 100



**Fig. 1 | Illustration of the machine learning-based drug prediction and validation process.** A small drug panel is first tested on a patient-derived cell line. Using historical screening data (comprising a panel of cell lines screened against both a

drug panel and a larger drug library), a machine learning model predicts drug responses for the full library. The top-ranked drug candidates are then experimentally validated.

**Fig. 2 | Overview of the study workflow, from data collection to model evaluation.** Four datasets (GDSC1, GDSC2, PRISM, RX) were pre-processed to remove missing values and imputed where necessary. The data were randomly split into training (80%), validation (10%), and test (10%) sets, with RX using leave-one-out cross-validation. A machine learning model was trained on historical drug response data to predict drug sensitivities for new patient-derived cell lines. Model performance was evaluated using Pearson correlation, Spearman correlation, RMSE, accuracy, and hit rate metrics.



**Table 1 | Predictive performance of the prototype drug recommender system on the GDSC1 dataset**

	$R_{\text{pearson}}$	$R_{\text{spearman}}$	RMSE	Top 10	Top 20	Top 30	Hit rate
Selective drugs	0.781/0.021	0.791/0.020	0.469/0.018	0.36/0.086	0.524/0.060	0.588/0.052	0.426/0.075
All drugs	0.885/0.008	0.865/0.010	0.507/0.017	0.66/0.063	0.763/0.027	0.755/0.028	0.978/0.011

Metrics reported are  $R_{\text{pearson}}$ ,  $R_{\text{spearman}}$ , RMSE, and accuracy of predictions within the top 10, 20 and 30 drugs, as well as hit rates among the top 10 recommendations. Results are presented as mean/stdev based on five experiments.

randomly selected patients (cell lines) when training the models (Table S6). These choices constitute the foundation of our prototype system.

### Predictive performance in the GDSC1 dataset

The prototype outlined above was applied to a dedicated test set, containing 81 patient-derived cell lines. The performance was evaluated through various metrics. In addition to  $R_{\text{pearson}}$ ,  $R_{\text{spearman}}$  and RMSE, we also included four accuracy metrics.

We reported the fraction of accurate predictions within the top 10, 20 and 30 drugs, providing an indication of the predictive accuracy. For instance, if 7 out of the 10 predictions matched the actual top 10 drug responses, we reported a fraction of 0.7. Additionally, we reported the hit rate among the top 10 predictions, offering a direct insight into the number of these recommendations that accurately identified hits.

All the results are reported as a mean of five experiments, accompanied by the resultant standard deviations (Table 1). For the sake of comparison, the average hit rate among the entire set of test set drugs was 17.8% for all drugs and 5.0% for ‘selective’ drugs (those active in less than 20% of the cell lines).

The prototype recommender system demonstrated excellent performance, with high correlations between predicted and actual drug activities for both the entire drug library and selective drugs alike (Table 1). When considering the entire library, we found that the recommender system performed very well. On average, 6.6 out of the top 10 predictions were correctly identified, with 15.26 and 22.65 accurate predictions for the top 20

and 30, respectively. Even when agnostic towards the exact ranking of the drugs and aiming for 10 recommended hits, the system consistently predicted almost exclusively hits.

In the more challenging task of predicting selective drugs, the results remained strong in terms of overall bioactivity ranking ( $R_{\text{pearson}} = 0.781$ ,  $R_{\text{spearman}} = 0.791$ ). When specifically aiming to identify the top 10, 20 and 30 drugs, the system provided an average of 3.6, 10.5 and 17.6 accurate predictions, respectively. The hit rate amongst the top 10 drugs was slightly higher, averaging at 4.3. It is important to note that, for the selective drugs, 50% of all cell lines had 12 or less hits in total out of the 236 available for prediction, demanding a nearly perfect system to pick them out. If one were to only consider the 41 cell lines with more than 12 hits present, the hit rate would increase to 6.1.

To further assess the accuracy of the top predictions, we conducted two additional experiments (Tables S9 and S10). First, we evaluated whether at least one of the top-3 drugs appeared among the system’s top-3 predictions, and also assessed how many of the top-3 predicted drugs were actual hits, regardless of their true rank. In both the all-drug and selective-drug scenarios, the system successfully captured top-performing drugs with high reliability (Table S9). Second, we extended the evaluation to the top-15 predictions, a realistic number for practical applications, examining how often the actual top-3 drugs were identified and whether the best-performing drug was correctly predicted. Again, the system demonstrated strong predictive performance, even when considering the more challenging selective-drug subset (Table S10).

**Table 2 | Predictive performance of the prototype drug recommender system on the GDSC2 dataset**

	$R_{pearson}$	$R_{spearman}$	RMSE	Top 10	Top 20	Top 30	Hit rate
Selective drugs	0.775/0.026	0.767/0.025	0.460/0.021	0.543/0.084	0.680/0.046	0.744/0.035	0.193/0.042 <sup>a</sup>
All drugs	0.892/0.010	0.841/0.016	0.526/0.020	0.737/0.051	0.824/0.026	0.804/0.024	0.890/0.019

The reported metrics encompass  $R_{pearson}$ ,  $R_{spearman}$ , RMSE, accuracy for the top 10, 20 and 30 predicted drugs, and hit rates among the top 10 recommendations. Results are presented as mean/stdev derived from five experiments.

<sup>a</sup>Maximum possible hit rate is 3.38.

**Table 3 | Impact of the different drug panel sizes on predictive performance on the PRISM dataset using a test set of 52 PDCs**

Panel size	$R_{pearson}$	$R_{spearman}$	RMSE	Top 0.5%	Top 1%	Top 2%	Hit rate (top 45)
23	0.530/0.045	0.394/0.035	0.776/0.012	0.067/0.036	0.169/0.059	0.368/0.087	0.559/0.103
45	0.609/0.024	0.444/0.022	0.742/0.009	0.083/0.046	0.215/0.070	0.437/0.078	0.662/0.099
90	0.687/0.019	0.473/0.015	0.684/0.010	0.159/0.064	0.345/0.074	0.567/0.054	0.845/0.063

The reported metrics include  $R_{pearson}$ ,  $R_{spearman}$ , RMSE, accuracy for the top 0.5%, 1%, and 2% of predicted drugs, and hit rates among the top 45 recommendations. Results are presented as mean/stdev derived from five experiments.

Looking closer at the performances across individual cell lines in terms of Spearman  $R$  coefficients, which indicates how well drugs are ranked in terms of their activity, we observed a minimum score of 0.76 for all drugs and 0.39 for selective drugs (Fig. S2). Notably, among the selective drugs, which are more difficult to predict, only 8 cell lines performed below 0.7 and only two fell below 0.65. These results illustrate the overall strong performance of the prototype recommender system while highlighting the increased complexity of predicting responses for selective drugs. An additional experiment comparing our approach to the use of standard molecular fingerprints was also performed and comparative results show that this approach is far superior on this dataset (Table S11).

**Predictive performance in the GDSC2 dataset**

The parameters selected for the GDSC1 dataset were applied to the GDSC2 dataset, demonstrating consistent high performance across all drugs. In the GDSC2 test set, the hit rate for all drugs was 13.2%, whereas for selective drugs, it was 2.5%. These values are notably lower than those of the GDSC1 dataset. When considering all drugs, the hit rate amongst the top 10 recommendations was high, with ~9/10 drugs being active on average (Table 2). However, the hit rate for selective drugs in the top 10 predictions was 0.193, significantly lower than in GDSC1 (Table 1). This decline can again be attributed to the small number of hits available in the dataset, averaging 3.38 across all cell lines. Consequently, the system's performance is constrained by this limit, achieving a maximum average of 3.38 hits per 10 recommendations. With this limit in mind, the system's performance can be adjusted to 57% in accuracy when compared to the theoretical maximum. Furthermore, 21 cell lines lacked any hits at all; if these are removed, the average hit rate goes up to 2.72, still standing at 57% of the maximum available hit rate (now averaging 4.77). There were only 8 cell lines that had 10 or more hits, averaging at 13.25, and for these cell lines, the average hit rate was 6.6/10.

**Predictive performance for an FDA-approved drug library**

The PRISM dataset differs significantly from previous datasets as it contains a larger library of FDA-approved drugs. Due to the unique and diverse nature of this dataset, we again investigated the optimal size for the probing drug panel, as well as the impact of the number of cell lines included in the training on model performance. We excluded the two compounds 'mg-132' and 'bortezomib' from our studies, as the authors of the PRISM study indicated their use as positive controls.

First, we explored the effect of the number of patients used in training the models. The experiments were performed using a drug panel containing 90 drugs (~2% of the dataset). Surprisingly, we found that reducing the number

of patients from 418 (all patients) to 30 patients still retained a significant amount of information necessary efficient predictions (Table S7). Intriguingly, even when working with 10 patients, correlations remained relatively close to those of the larger patient subsets. However, for this minimal cohort, there was a marked drop in accuracy amongst the top-performing drugs, as well as hit rates among the top 45 selections. Nonetheless, our study demonstrates that reducing the number of patients to ~10% of the entire dataset still yields strong predictions and a substantial number of hits in the recommendations. Based on our findings, we chose to proceed with 100 patients, as the performance was nearly identical to that of 200 and 418 patients.

Next, we investigated the impact of drug panel size on performance. We conducted experiments with drug panels consisting of 23, 45 and 90 drugs, representing ~0.5%, 1% and 2% of the dataset, respectively. We observed that there was a marked improvement with each size increase. However, considering that testing 90 drugs for every new patient could be quite substantial, we recommend using a panel of 45 drugs (or less). This smaller panel still yielded competitive results and a substantial number of compounds active against the cell lines, ~30 hits present among 45 recommendations (Table S8).

The calibration work above was performed using a validation set. Subsequently, we tested the same panel sizes using an independent test set of 52 patients and observed very similar performance (Table 3). It became evident that the hit rates (amongst the top 45 predicted drugs) corresponded to increases of ~1000–1600% compared to screening the entire library (5.2% hit rate). This gives rise to 45 recommendations containing ~25–38 active compounds on average depending on the chosen panel size.

**Predictive performance for RX dataset**

The RX dataset was acquired from a range of tumour types. The drug screens were conducted on cultured biopsies of cancer patients shortly after their arrival at a laboratory. Due to the limited number of drugs and cells in the dataset resulting from the stringent inclusion criteria (see methods), we employed leave-one-out cross-validation for this dataset. This involved predicting the viability of patient-derived cells when treated with a drug, based on their resultant viabilities when treated with the remaining drugs in the dataset as well as those of other patients. In this scenario, the unknown activities illustrated in Fig. 1 would pertain to a single drug, while the remaining drugs would be part of the drug panel.

Approximately 30 different drugs were tested against each cell line; the average resultant viability across cell lines for this library was 87.55% (Table 4). In contrast, when selecting the five drugs predicted to be most effective towards each cell line, the mean viability of the cancerous cells dropped to 33.45%. Using the threshold of <30% viability as an indication for a hit, the average number of hits available per cell line was 4.36. On

**Table 4 | Performance on RX dataset**

$R_{\text{spearman}}$	Library (viability)	Top 5 ranked (viability)	$n$ hits available	$n$ hits predicted	$n$ hits identified	$n$ drugs available
0.62/0.02	87.55	33.45/2.83	4.36	4.87/0.27	3.60/0.11	33.00

For the 24 cell lines, the mean Spearman rank is given for the entire library of 30+ drugs. The average viability resulting from the entire drug library is reported, along with that found when selecting the top 5 predicted drugs. Defined as reducing cell viability to less than thirty percent, the average number of hits available per cell line is shown, as well as the average number of drugs predicted to be hits, and the average number of these correctly identified as such. The average number of drugs tested per cell line is found in the last column 'n drugs available'. The results from five independent experiments are shown as mean/standard deviation.

average, the methodology recommended 4.87 drugs for testing and 3.60 of these were identified as hits. Three of the cell lines never experienced a viability decrease of <30% and were excluded from the hit analysis. The resultant average hit rate across all cell lines was 69.40%, and our approach managed to identify an average of 68.78% of all available hits across all cell lines. For three cell lines, the resultant hit rates were 0%, and in each case, the total hits amongst the 30+ compounds were low, at 1 or 2 hits. However, the methodology always managed to rank drugs in the library so that the top 5 predicted drugs resulted in a significantly lower viability than that of the library. Complete results can be found in Table S12.

## Discussion

The future of cancer therapy is moving towards a personalised approach, aiming to align treatment strategies with the needs of the individual patient. Current efforts in this area are focused on linking specific treatment options to genetic markers, but the choices of such personalised drugs are relatively few, and this approach only benefits ~10% of all patients<sup>23</sup>. The challenge only becomes greater when selecting medications that lack a direct link to specific gene defects.

Efforts to connect therapy choices to genetic makeup have encountered limitations in their success<sup>18,19</sup>. Most promising studies have concentrated on a select few drugs rather than a larger library or a panel of currently available drugs. These limitations are rooted in the complexity of the cell and its dynamic molecular regulation at multiple levels, rendering it challenging to predict drug outcomes based on such complex and interconnected information.

In this context, the methodology presented in this work set out to circumvent this complexity by shifting the focus away from a static depiction of a cell's state and its potential response. Instead, it explores the cell's ability to respond to a diverse drug panel, aiming to understand its interaction with drugs outside this panel.

This study delves into drug response profiles as predictors in precision medicine, seeking to establish fundamental principles for constructing robust machine learning models for recommending drugs effective against patient-derived cancer cells. The correlations between the number of patient-derived cells, the drug panel size and model performance provide valuable insights into the key parameters influencing model performance. The experiments indicate that the optimal range for achieving efficient predictions falls within 100–200 patients in the training data, with incremental gains in performance beyond this (Tables S5 and S6). As for the drug panel, depending on the type of library one wants to extract recommendations from, one will have to use varying numbers of drugs in the probing panel. For instance, the experiments demonstrate that even a probing panel as small as 20 drugs can yield high performance in predicting activities in the GDSC1 dataset consisting of ~300 anti-cancer compounds (Table S4). The experiments also showed that a small probing panel of 23 drugs resulted in an average of 25 hits amongst 45 selected from an FDA-library of ~4500 compounds (Table 3).

Notably, this methodology exhibits a remarkable level of agnosticism towards the specific type of cancer tissue from which the cells are derived. This finding is further underscored when investigating the composition of the ten most informative cancer cells (according to their tissue types) when predicting example cell lines from different tissues of origin (Fig. S3 and Fig. 3). Only haematopoietic and lymphoid tissue cancer appears to prefer similar cancer types for efficient prediction. In addition, we found that the specific selection of drugs that is included in

the panel, as well as the size of the panel, have a profound impact on how patients cluster (Fig. S4), consistent with previous studies<sup>24</sup>. The clustering becomes more robust when increasing the size of the panels, suggesting that overall, specific cell lines tend to be similar in their general response (hints at the concept of digital twins in precision medicine).

Lastly, this methodology demonstrated strong performance on the RX dataset, which includes tissue cultures exposed to drugs shortly after biopsy. This serves as a crucial validation, highlighting the approach's potency with clinical samples, even when the drug library is limited and the number of patients is small. In nearly every case where hits were present, at least one of these was identified. Only in 3 out of 21 cell lines, where there were only 1 or 2 hits, did the approach fail to identify one. However, the methodology was still successful; for every cell line, the mean viability of the top 5 hits was significantly lower than that of the entire library, suggesting that the methodology correctly identified effective compounds from this library (Table S12).

Clinicians have already seen promising outcomes using PDC data to guide treatment decisions, particularly regarding drug resistance. Our methodology offers a streamlined and cost-effective solution, requiring only a fraction of a drug library for initial screening and capable of delivering a tailored selection of effective drugs within a week, revolutionising treatment timelines. Recognising the diverse drug susceptibilities of cancer cell lines, our method supports the formulation of strategic drug cocktails to optimise treatment, minimise redundancy, and prevent resistance.

It is important to note that the methodology presented in this work is based on data from cell-based assays. This means that challenges may arise when studying certain classes of drugs such as immunomodulators. Acknowledging the growing significance of modulators influencing the microenvironment and immune system, future iterations of this work must address these nuances. Strategic drug cocktails, including modulators and various therapeutic agents, are likely to be at the core of future precision medicine. Our methodology stands poised to contribute to these efforts by facilitating potent and tailored therapies.

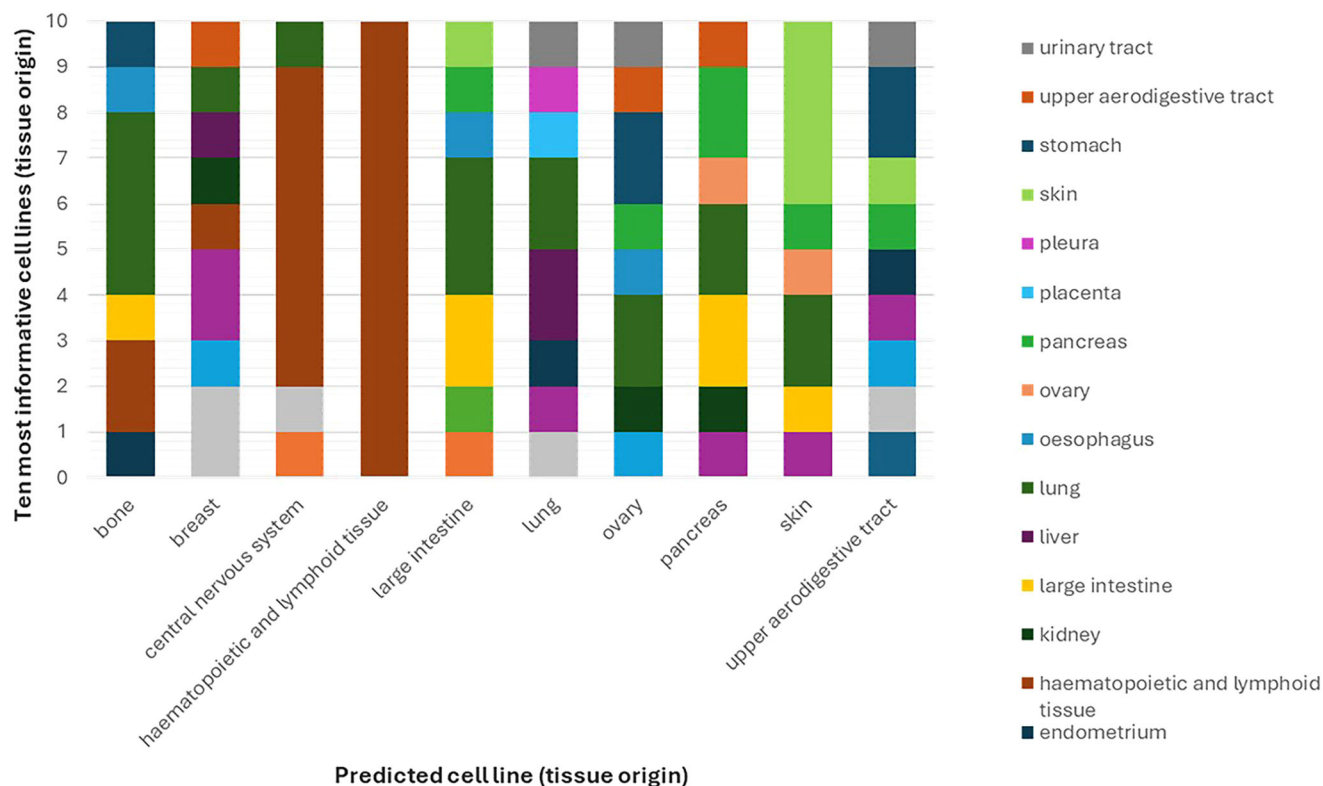
## Methods

The RX dataset constituted tissue cultures derived from 24 patients with different cancer types screened against a drug library of 35 FDA-approved drugs using a cell viability assay (Table S1).

### Tissue collection and drug screening

This study was conducted in accordance with relevant UK guidelines and regulations, and the ethical principles of the Declaration of Helsinki. Ethical approval was provided by the North West - Greater Manchester South Research Ethics Committee (Reference22/NW/0237). Informed consent was obtained from all participants (or their parent/legal guardian in case of minors).

Fresh samples of tumour specimen were acquired from hospitals and clinics in the UK. Upon arrival, the samples were washed and processed according to the dissociation kit's instructions (Miltenyi Biotec, cat. 130-095-929) and gentleMACS dissociator (Miltenyi Biotec). After washing in PBS and centrifugation, the cell pellet was resuspended in prewarmed medium and incubated for 48 h at 37 °C, 5% CO<sub>2</sub>. Medium was changed to growth medium, and cells were grown to ~1 × 10<sup>6</sup> cells. Growth medium was changed every second day. Drugs were transferred to 384-well plates (Grenier cat. 781091) to result in a final



**Fig. 3 | Feature selection results.** For each of 10 cell lines being predicted (defined by their tissue of origin), the most influential cell lines were identified, and their respective tissue of origin reported. Fifty percent of the drug library was used in the probing panel.

**Table 5 | Number of patients and drugs present in each dataset after data pruning according to specified thresholds**

Dataset	n-patients	n-drugs	n-drugs (selective)
GDSC1	809	322	266
GDSC2	707	165	137
PRISM	522	4515	4199
RX	24	35	-

All drugs were used for the RX dataset as the drug library only constituted 35 compounds.

concentration of 1  $\mu\text{M}$  using a D300e digital dispenser (Tecan). Cell suspension was transferred to the wells with a density of 2000 cells per well in 50  $\mu\text{L}$  growth medium. Cells were incubated with the drugs for 72 h at 37  $^{\circ}\text{C}$ , 5%  $\text{CO}_2$ . Hoechst was used to stain the cells and acquire cell counts. The resultant values were used to calculate viability in relation to negative DMSO and positive Benzethonium chloride controls. The generated dataset was named RX.

### Datasets

Four datasets were used in this study; the publicly available sets GDSC1, GDSC2 and PRISM<sup>25,26</sup> datasets, as well as the additional dataset (RX) described above.

### Processing datasets

In the case of GDSC1 and GDSC2, we excluded cell lines with more than 20% missing drug measurements from the analysis. Similarly, drugs untested in over 20% of the remaining cell lines were also removed from the study. For the larger PRISM dataset, we applied a more stringent threshold of 10% for cell line exclusion and 20% for drug exclusion. The resulting number of cell lines and drugs after this data pruning is summarised in Table 5. For the RX dataset, given the limited size, we excluded

cell lines with over five missing values, and drugs missing in over five cell lines.

### Target values

For GDSC1 and GDSC2, we transformed target values to  $\log(\text{IC}_{50})^{-1}$ , with activity values above 6 ( $\text{IC}_{50} \leq 1 \mu\text{M}$ ) considered active. In the case of PRISM, we utilised  $\log_2(\text{fold-change})^{-1}$  values as the target values, with activity values above 1.7 considered active. For RX, the dataset contained viability values normalised against DMSO controls, a value of 100 corresponded to 100% viability compared to DMSO controls. An active compound was indicated by the reduction of viability to  $\leq 30\%$ .

### Data completion

Since all the drug response datasets were incomplete, we considered two approaches for addressing this issue. The first involved filling the missing values in the matrices with zeros. Alternatively, we employed imputation using TML, an approach that generates a predictive model based on available values in the matrix to predict missing values. In this latter approach we used the sci-kit learn implementation GradientBoostingRegressor ( $n_{\text{estimators}} = 50$ ,  $\text{max\_depth} = 1$ ). We excluded all test and validation samples (cell lines) ensuring data integrity. The remaining cell lines were designated as samples, and drugs were treated as features. Missing drug response values for each drug were predicted. For each drug under prediction, we generated a model based on all other drugs with fewer than 20 missing entries, any such entries were filled with zeros, and the missing values for the relevant drug were predicted and incorporated into a newly imputed 'TML-matrix'.

### Model training and validation

To train the predictive models, we split the datasets into training, validation, and test sets. GDSC1, GDSC2, and PRISM followed predefined splits to prevent data leakage. The data was randomly split into 80% training, 10% validation, and 10% test sets. We performed cell and drug panel selection on the training set using a correlation-based approach. Samples with the lowest

50% variance were first excluded. A Pearson correlation matrix was computed across all remaining samples, which were then ranked by median correlation. Starting with the sample exhibiting the lowest median correlation, any samples exceeding a user-defined correlation threshold relative to it were excluded. The next available sample in the original ranking was then evaluated in the same manner, and the process was repeated until no samples remained. The low-redundancy and diverse samples selected through this process constituted the final panel. Due to its small size, RX was analysed using leave-one-out cross-validation.

### Regression

We conducted a comparison of regression models using the Sci-kit learn implementations RandomForestRegressor and GradientBoostingRegressor, both of which have shown strong performance in similar tasks<sup>27–30</sup>. We explored a predefined set of configurations, including varying the number of trees for both learners and adjusting the max\_depth parameter for GradientBoostingRegressor. This exploration was sufficient to build effective prototype systems. More comprehensive parameter optimisation across a wider range of learners can be performed in future studies. Feature importance scores were obtained using the random forest model to identify influential input features.

### Evaluation metrics

To assess the performance of these regression models, we employed a range of evaluation metrics. These metrics included Spearman R, Pearson R, root mean squared error (RMSE), standard deviation, hit rates and the accuracy of top *n* selections. These metrics served to provide a comprehensive evaluation of the model's predictive accuracy, robustness and suitability for the given task. Due to the smaller size of the RX dataset, leave-one-out cross-validation was used, and the metrics we chose were Spearman R, mean viability of the top 5 predicted hits, number of drugs predicted as hits as well as how many of these that were actual hits.

### Data availability

The RX dataset can be found at <https://github.com/abbiAR/DrugResponsePrediction>. The GDSC1/2 datasets are available at <https://www.cancerrxgene.org/>, we used the datasets with calculated IC50 values. The PRISM datasets are available at <https://depmap.org/repurposing/>, we used the primary screen dataset with collapsed replicates and calculated logfold change values.

### Code availability

The code can be found at <https://github.com/abbiAR/DrugResponsePrediction>.

Received: 19 December 2024; Accepted: 5 May 2025;

Published online: 13 June 2025

### References

- Dolgin, E. The future of precision cancer therapy might be to try everything. *Nature* **626**, 470–473 (2024).
- Izumchenko, E. et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **28**, 2595–2605 (2017).
- Dobrolecki, L. E. et al. Patient-derived xenograft (PDX) models in basic and translational breast cancer research. *Cancer Metastasis Rev.* **35**, 547–573 (2016).
- de Witte, C. J. et al. Patient-derived ovarian cancer organoids mimic clinical response and exhibit heterogeneous inter- and inpatient drug responses. *Cell Rep.* **31**, 107762 (2020).
- Chen, X. et al. Breast cancer patient-derived whole-tumor cell culture model for efficient drug profiling and treatment response prediction. *Proc. Natl. Acad. Sci. USA* **120**, e2209856120 (2023).
- Wensink, G. E. et al. Patient-derived organoids as a predictive biomarker for treatment response in cancer patients. *npj Precis. Oncol.* **5**, 30 (2021).
- Liebers, N. et al. Ex vivo drug response profiling for response and outcome prediction in hematologic malignancies: the prospective non-interventional SMARTrial. *Nat. Cancer* **4**, 1648–1659 (2023).
- Snijder, B. et al. Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study. *Lancet Haematol.* **4**, e595–e606 (2017).
- Kornauth, C. et al. Functional precision medicine provides clinical benefit in advanced aggressive hematologic cancers and identifies exceptional responders. *Cancer Discov.* **12**, 372–387 (2022).
- van de Wetering, M. et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).
- Weeber, F. et al. Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *Proc. Natl. Acad. Sci. USA* **112**, 13308–13311 (2015).
- Sachs, N. et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell* **172**, 373–386.e10 (2018).
- Su, C. et al. The efficacy of using patient-derived organoids to predict treatment response in colorectal cancer. *Cancers* **15**, 805 (2023).
- Vamathevan, J. et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **55**, 263–274 (2015).
- Burd, A. et al. Precision medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary efficacy of the Beat AML Master Trial. *Nat. Med.* **26**, 1852–1858 (2020).
- Horak, P. et al. Comprehensive genomic and transcriptomic analysis for guiding therapeutic decisions in patients with rare cancers. *Cancer Discov.* **11**, 2780–2795 (2021).
- Letai, A. Functional precision cancer medicine-moving beyond pure genomics. *Nat. Med.* **23**, 1028–1035 (2017).
- Prasad, V. Perspective: the precision-oncology illusion. *Nature* **537**, S63 (2016).
- Olier, I. et al. Transformational machine learning: learning how to learn from many related scientific problems. *Proc. Natl. Acad. Sci. USA* **118**, e2108013118 (2021).
- Kauvar, L. M. et al. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **2**, 107–118 (1995).
- Helal, K. Y. et al. Public domain HTS fingerprints: design and evaluation of compound bioactivity profiles from PubChem's bioassay repository. *J. Chem. Inf. Model.* **56**, 390–398 (2016).
- Malani, D. et al. Implementing a functional precision medicine tumor board for acute myeloid leukemia. *Cancer Discov.* **12**, 388–401 (2022).
- Pouryahya, M. et al. Pan-cancer prediction of cell-line drug sensitivity using network-based methods. *Int. J. Mol. Sci.* **23**, 1074 (2022).
- 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 (2013).
- Corsello, S. M. et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **1**, 235–248 (2020).
- Laufkötter, O. et al. Combining structural and bioactivity-based fingerprints improves prediction performance and scaffold hopping capability. *J. Cheminformatics* **11**, 54 (2019).
- Yang, M. et al. Machine learning models based on molecular fingerprints and an extreme gradient boosting method lead to the discovery of JAK2 inhibitors. *J. Chem. Inf. Model.* **59**, 5002–5012 (2019).
- Sheridan, R. P. et al. Extreme gradient boosting as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **56**, 2353–2360 (2016).

30. SCIKIT, Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

### Acknowledgements

None of the above work would have been possible without the availability of cancer-derived cell lines. We would like to express gratitude to all patients who have made this, as well as many other studies contributing to the understanding and treatment of cancer, a possibility. This work has been supported by grants from the UK Engineering and Physical Sciences Research Council (EPSRC) [EP/R022925/2, EP/W004801/1 and EP/X032418/1].

### Author contributions

A.A. designed the study, analysed the data, and wrote the main manuscript text. G.G. was involved in the work producing the Vali-RX dataset. R.K. and L.S. supervised the study. All authors reviewed the work and the manuscript.

### Competing interests

G.G. works with Vali-RX, a CRO in the area of drug development. The remaining authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-025-00937-2>.

**Correspondence** and requests for materials should be addressed to Abbi Abdel-Rehim.

**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) 2025