



# Integrated single-dose kinome profiling data is predictive of cancer cell line sensitivity to kinase inhibitors

Chinmaya U. Joisa<sup>1</sup>, Kevin A. Chen<sup>2</sup>, Matthew E. Berginski<sup>3</sup>, Brian T. Golitz<sup>4</sup>, Madison R. Jenner<sup>3,5</sup>, Gabriela Herrera Loeza<sup>5</sup>, Jen Jen Yeh<sup>2,3,5</sup> and Shawn M. Gomez<sup>1,3</sup>

<sup>1</sup> Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University, Chapel Hill, NC, United States of America

<sup>2</sup> Department of Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

<sup>3</sup> Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

<sup>4</sup> Eshelman Institute for Innovation, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

<sup>5</sup> Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

## ABSTRACT

Protein kinase activity forms the backbone of cellular information transfer, acting both individually and as part of a broader network, the kinome. Their central role in signaling leads to kinome dysfunction being a common driver of disease, and in particular cancer, where numerous kinases have been identified as having a causal or modulating role in tumor development and progression. As a result, the development of therapies targeting kinases has rapidly grown, with over 70 kinase inhibitors approved for use in the clinic and over double this number currently in clinical trials. Understanding the relationship between kinase inhibitor treatment and their effects on downstream cellular phenotype is thus of clear importance for understanding treatment mechanisms and streamlining compound screening in therapy development. In this work, we combine two large-scale kinome profiling data sets and use them to link inhibitor-kinome interactions with cell line treatment responses (AUC/IC<sub>50</sub>). We then built computational models on this data set that achieve a high degree of prediction accuracy (R<sup>2</sup> of 0.7 and RMSE of 0.9) and were able to identify a set of well-characterized and understudied kinases that significantly affect cell responses. We further validated these models experimentally by testing predicted effects in breast cancer cell lines and extended the model scope by performing additional validation in patient-derived pancreatic cancer cell lines. Overall, these results demonstrate that broad quantification of kinome inhibition state is highly predictive of downstream cellular phenotypes.

Submitted 10 February 2023  
Accepted 3 October 2023  
Published 16 November 2023

Corresponding author  
Shawn M. Gomez, smgomez@unc.edu

Academic editor  
Abhishek Tyagi

Additional Information and  
Declarations can be found on  
page 15

DOI 10.7717/peerj.16342

© Copyright  
2023 Joisa et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Bioinformatics, Cell Biology, Computational Biology, Oncology

**Keywords** Cancer, Drug response, Predictive modeling, Machine learning, Kinase inhibitor treatment, Cell phenotypes, Drug sensitivity, Data integration, Cell signaling

## INTRODUCTION

Computational drug screening has recently emerged as a powerful approach to integrate vast amounts of cancer cell line multi-omics data into predictive models, with the goal of predicting downstream phenotypic responses such as growth and viability. Advancing rapidly with recent developments in machine learning, these methods have the potential to predict outcomes for large drug libraries with minimal experimental cost, reducing the number of drug candidates fed into downstream validation efforts. Most current approaches to the prediction of drug response use baseline cell-line multi-omics data (e.g., mutation status, gene expression, copy number variation, *etc.*) and quantitative structure-activity relationships (QSAR) to map drug structure characteristics onto their biological phenotypes. However, information describing drug-target interactions, especially at the protein level, remains underutilized due to the unique nature of associated data acquisition methods. For example, the recent DREAM challenge ([Costello et al., 2014](#)) hosted by the National Cancer Institute (NCI) for drug-response predictions saw the winning team utilize high throughput drug screening data along with baseline gene expression features ([Gönen & Margolin, 2014](#)) and achieved at most 80% accuracy in predicting cell line responses in a binary, yes/no fashion.

As a central component of cellular information transfer, protein kinases are enzymes that have also shown ([Gambacorti-Passerini et al., 2011](#)) promise as therapeutic targets, with initial success being found through the development of Imatinib (Gleevec). Drugs that inhibit kinases (“kinase inhibitors”) are now one of the fastest growing clinical drug classes (74 FDA approved as of 2022), but around 1/3rd of all known kinases still have relatively unknown function and few chemical tools exist to interrogate and expand this knowledge. To explore the potential of the kinome as a therapeutic target, recent work has focused on profiling the full breadth of targets for kinase inhibitors, especially since many inhibitors have significant off target effects as a result of targeting the conserved ATP-binding pocket. Continued improvements in high-throughput assays such as Kinobead/MS ([Reinecke et al., 2019](#)), KINOMEscan (©DiscoverX), and KiNativ ([Patricelli et al., 2011](#)) now enable measurement of a given inhibitor’s interactions across 250–500 kinases, providing a snapshot of its effect on the physiological kinome. We refer to this kinome-wide profiling data as the “kinome inhibition state” of a given inhibitor. This ability to generate drug-target interaction data on a large scale for a compound class is relatively unique, providing a novel means to leverage knowledge of off-target effects for drug response prediction.

The DepMap portal database ([Corsello et al., 2020](#)) contains thorough multi-omic characterization of ~1,000 cancer cell lines of all types, and corresponding cell viability measurements for about ~1,500 repurposed compounds, ~250 of which are kinase inhibitors. Using this data, we can connect kinase inhibitor phenotypes of cell viability to their “kinome inhibition states” and build models to predict the cellular responses to treatment with different kinase inhibitors. We have previously shown that these kinome states obtained solely through the kinobeads assay for clinical inhibitors are highly predictive of cancer cell viability, and also validated these predictions experimentally ([Berginski et al., 2022](#)). However, the kinobeads assay is unique and requires dedicated lab personnel

to run, restricting its use to relatively few labs. In contrast, the KINOMEscan assay is a popular and easily accessible commercial alternative that assays a panel of ~500 native and mutant kinases recombinantly. Large amounts of KINOMEscan data have been deposited online by various groups (Koleti *et al.*, 2018; Wells *et al.*, 2021), including data for inhibitors developed against understudied kinases. These altogether account for four times as many inhibitors profiled (~800) when compared to the data available from the kinobeads assay, representing a massive expansion of publicly available inhibitor state data. However, due to the uncharacterized nature of the inhibitors in the large KINOMEscan data set, only a small number of them (~40) have been tested in the DepMap screening database, compared to ~200 inhibitors from the kinobeads set.

In this work, we describe a framework to create an integrated kinome inhibition state data set by combining kinobeads and KINOMEscan data, and then leverage the entirety of this data into predictive models. This combined set contains single-dose inhibitor profiling data for a total of ~800 kinases and kinase interacting proteins, spanning almost 1,000 kinase inhibitors that target a diverse section of the overall kinome space. When leveraged within a machine learning framework, and supplemented with baseline gene expression data, we are able to predict the sensitivity of ~450 cancer cell lines in the DepMap screening dataset, with a reasonable  $R^2$  of ~0.7. Using this model, we were able to generate sensitivity predictions for 1.2 million inhibitor-cell line combinations, many of them targeted towards understudied kinases. We then experimentally validated these predictions in well characterized breast cancer cell lines seen by the model, as well as primary derived pancreatic cancer cell lines. We find reasonable agreement between predicted and observed outcomes in most compounds, seeing an expected drop in performance for understudied compounds and unique patient-derived cell lines. Together, these results show that there is a strong and predictive relationship between the state of the kinome (its “kinotype”) and downstream cellular phenotypes, while further suggesting potential opportunities for leveraging computational models in inhibitor therapy design.

## METHODS

Portions of this text were previously published as part of a preprint (<https://www.biorxiv.org/content/10.1101/2022.12.06.519165v1>).

### Data sources

The primary data sources we used can be split into two categories: the integrated kinome profiling data set and the cancer cell line set:

The following were downloaded from the respective Supplementary Materials to create the integrated set of kinome profiling data:

1. Kinome profiling data from the kinobeads assay
  - a. *Klaeger et al.* (2017)
2. Kinome profiling data from KINOMEscan assay
  - a. LINCS: kinome profiling datasets for individual compounds downloaded programmatically from <http://lincs.hms.harvard.edu/db/datasets/> (Koleti *et al.*, 2018)

- b. Kinome profiling data for the PKIS drug set was downloaded from the Supplementary Data of <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181585> (Drewry et al., 2017)
- c. Kinome profiling data for the KCGS drug set was downloaded from the Supplementary Data of <https://www.mdpi.com/1422-0067/22/2/566> (Wells et al., 2021) and from internal data provided by SGC-UNC.

The following were downloaded from the DepMap portal (<https://depmap.org/portal/download/all/>) to create the set of cancer cell line sensitivities and their gene expression characteristics:

1. DepMap secondary repurposing screen (“secondary-screen-dose–response-curve-parameters.csv”)
2. CCLE gene expression set (“CCLE\_expression.csv”)

The high-throughput screening data gathered in PDAC patient-derived cell lines was gathered from Lipner et al. (2020) with methods as described in Berginski et al. (2021a).

### Data preprocessing

The scripts implementing these descriptions are all available through github.

*Klaeger et al. Kinobead Kinase Inhibition Profiles:* As previously described (Berginski et al., 2022), we read the values from the [Supplemental Data](#) table into R and produced a filtered list of kinase and kinase interactor relative intensity values. We imputed missing values with the default “no interaction” value of 1, and truncated likely outlier values to the 99.99 percentile (3.43).

*KINOMEScan Inhibition Profiles:* We read in the three datasets mentioned above into R and concatenated them into a single combined data set. All the individual data sets contain identical protein lists because of the same assay type. Values are reported as “Percent Control”, a ratio of protein pulled down in experimental condition (with inhibitor) vs control condition (without inhibitor). These were divided by 100 to convert the scale to 0–1 to match the Kinobeads relative intensity data.

*Creating the Combined Kinome Inhibition Profiling Set:* We took the kinobeads dataset and the KINOMEScan dataset and concatenated them into one large set containing inhibitor-kinase interaction states for ~800 total kinases and kinase interactors. We left out assays that included recombinantly mutated kinases but left those with naturally occurring post-translational modifications. The vast majority (99.95%) of the inhibitor-kinase pairs represented was unique for either assay type, but for the 0.05% inhibitor-kinase pairs, we took the mean value of the measurements across the two assay types. Additionally, any missing values were imputed with the default “no interaction” value of 1. In the end we were left with kinome inhibition states for ~1,000 kinase inhibitors.

*Dataset of Cancer Cell Line Sensitivity to Drugs from DepMap:* The DepMap repurposing dataset contains cell viability measurements across multiple doses, but since our dataset of kinome states is restricted to single-dose measurements, we extracted single summary statistics of cell line sensitivity to kinase inhibitors: Dose–response area under the curve (AUC) and half-maximal inhibitory concentration (IC<sub>50</sub>). We extracted these by reading in the “secondary-screen-dose–response-curve-parameters” dataset into R, which contains

curve parameters for a log–logistic curve fit to the cell viability dose response curve and filtered it down to cell line name, IC<sub>50</sub>, AUC and other associated metadata.

*Matching of Kinase Inhibitors between Profiling Dataset and Cell-Line Sensitivity Dataset:* The compound names from each dataset were read into R, and the package Webchem ([Szöcs et al., 2020](#)) was used to retrieve PubChem compound IDs. The two sets of compound names were then matched based on these reference IDs. There were 252 matches between the two sets, forming a final set of ~70,000 inhibitor-cell line combinations.

*Baseline Gene Expression:* As described before ([Berginski et al., 2022](#)) the RNAseq data provided in the “CCLE\_expression.csv” file needed no modifications while preprocessing. Our only modification was to add identifiers to each gene label (“exp\_”), to ensure that kinome inhibition data and expression data related to the same gene weren’t accidentally combined.

*String:* The STRING database ([Szklarczyk et al., 2021](#)) was processed as described previously ([Berginski et al., 2022](#)) to annotate kinases and kinase interacting genes.

## Modeling techniques

To assess our models we used a random 10-fold cross validation strategy. The number of features was varied as specified by the feature selection scheme described in the results section. We compared the performance of three model types using this strategy: LASSO (Least Absolute Shrinkage and Selection Operator) regression using the glmnet engine ([Friedman, Hastie & Tibshirani, 2010](#)), random forest using the ranger engine ([Wright & Ziegler, 2017](#)) and gradient boosting using the XGBoost (eXtreme Gradient Boosting) engine ([Chen & Guestrin, 2016](#)). Model performance was assessed by the R<sup>2</sup> value between predicted and actual outcome within the cross-validation scheme. For each model type, we tuned sets of 30 hyperparameters to find the best possible performer as follows:

1. LASSO
  - a. Penalty (1E–10–0.9)
2. Random Forest
  - a. Trees (100–2000)
3. XGBoost
  - a. Trees (100–1000)
  - b. Tree Depth (4–30)

After final model selection, we fit the model on the entire dataset and then made predictions on inhibitor-cell line pairs not found in the original DepMap screening data.

## Compound testing

BT-474, HCC1806, SUM-159 and SKBR-3 cells were assayed as described previously ([Berginski et al., 2022](#)). Briefly, cells were grown in ATCC recommended media and seeded in triplicate at 4000, 2000, 4000 and 500 cells per well respectively. 24 h after seeding, cells were treated with inhibitors at 30 μM, 3 μM, 1 μM, 300 nM, 100 nM, 30 nM, 10 nM, and 3 nM, along with the appropriate DMSO controls. seeded at, in white flat-bottom 96-well plates (Corning). Seventy-two hours post-treatment, cells were lysed with CellTiter-Glo (Promega) and luminescence was read using the PHERAstar FS microplate reader (BMG

Labtech) and gain adjustments were conducted for each cell line. Notably, we used a 3-day treatment period instead of the 5-day treatment period used in the highly multiplexed PRISM assay, since we saw over-confluence in our 96-well plate assay in times longer than 3 days. However, it is likely that summary measures of trends in response like IC<sub>50</sub> and AUC are minimally affected by the small difference in treatment times. Data were averaged over replicates normalized row-wise to the DMSO-only (0.1% on cells) control samples on each plate to calculate relative viability. Quality checks were performed to look at the data distribution and the presence of spatial bias on a plate. A quality control metric of <120% of DMSO was applied to all rows analyzed. The full file of cell viability curves is available in [Table S1](#).

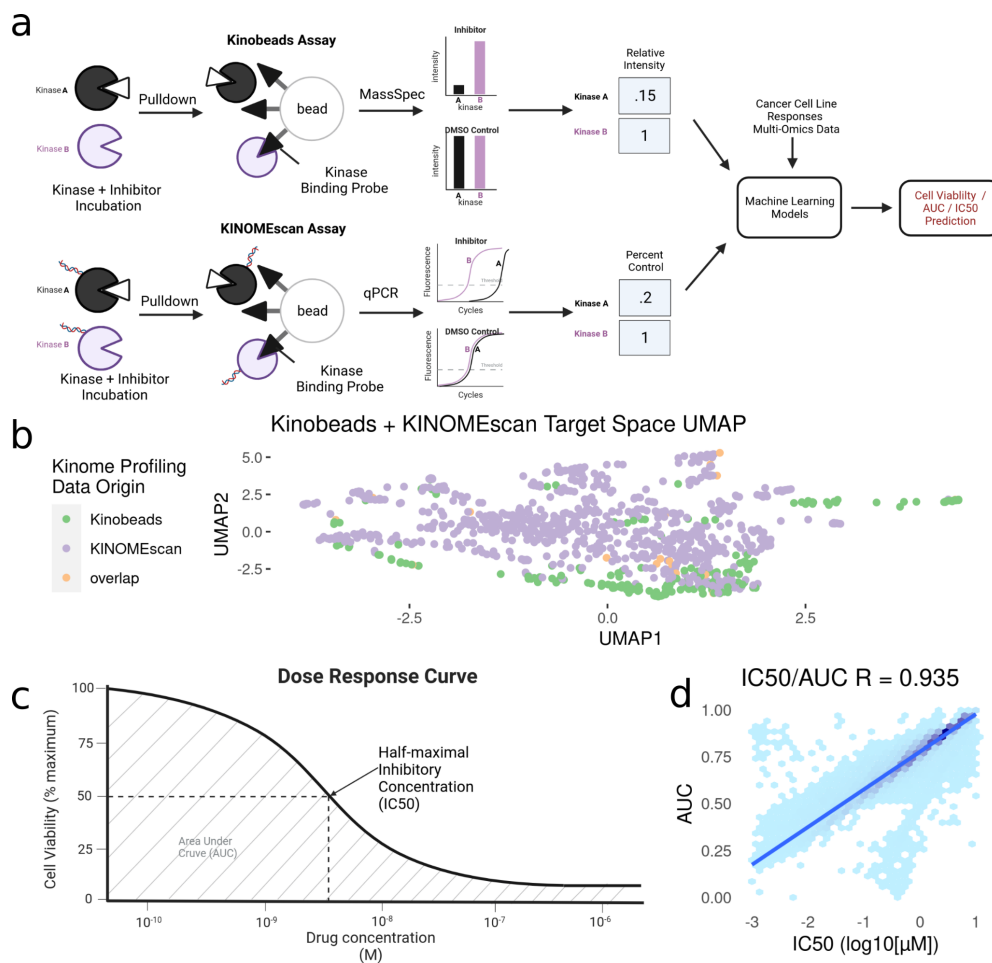
The functions “ComputeAUC” and “ComputeIC50” from the R package *dr4pl* ([Gadagkar & Call, 2015](#)) was used to fit a four-parameter log-logistic curve to the cell viability data, and extract AUC and IC<sub>50</sub> values from the 9-point cell viability curves.

## RESULTS

### Creating an integrated set of kinome profiling data across a wide chemical space

Kinase inhibitors have been profiled using a number of assays, but for this study we have focused on a specific subset of kinase inhibitors that have been assayed using the kinobead/MS-based method ([Klaeger et al., 2017](#)) or the KINOMEscan<sup>®</sup> (DiscoverX) method. These methods assess their specific kinase targets as well as the magnitude of inhibition of each kinase in response to different inhibitor concentrations ([Klaeger et al., 2017](#)). We combined kinome profiling datasets from [Klaeger et al. \(2017\)](#) (Kinobeads), LINCS ([Koleti et al., 2018](#)) (KINOMEscan), and UNC ([Drewry et al., 2017](#)) (KINOMEscan), filtering down to profiles measured only at 1 $\mu$ M. For the small amount of overlap between datasets, the mean inhibition value was taken across drug-kinase combinations. Given that both assays measure the engagement of inhibitors to kinases, most of the proteins that appear in assay results are either known kinases or closely associated proteins. Specifically, kinase inhibitor profiles include measurements on all wild-type and phosphorylated kinases (~500), along with a set of associated proteins (~300). As such, we will refer to this data as “kinase inhibition states”, and the profile of each individual drug as its “kinome inhibition state” ([Fig. 1A](#)).

After integration, we were left with a final set of ~1,000 compounds with corresponding information on their inhibitor-induced kinome states, describing changes in ~800 kinases and kinase interactors. To summarize the relationship between inhibition states induced by all inhibitors, we performed a UMAP dimensionality reduction ([McInnes et al., 2018](#)) on the dataset, only using kinases profiled in both assays for visualization ([Fig. 1B](#)). The UMAP coordinates represent the aggregate effect of each inhibitor on the kinome, *i.e.*, it is a representation of the uniqueness of its kinome inhibition state. Inhibitors that have similar effects on the kinome will have similar coordinates, while disparate inhibitors will have coordinates that are far apart. Using this, we can examine the diversity of kinome space targeting in our dataset, based on the origin of kinome profiling data. Our analysis



**Figure 1** Modeling pipeline and target variable overview. (A) Schematic of Kinobead/MS (upper) and KINOMEScan assay data integration into machine learning models predicting IC<sub>50</sub> and AUC. (B) Visualization of UMAP dimensionality reduction on the combined kinome profiling data set, each point represents a single compound's position in the target space, and colors representing the origin of kinome profiling data. Target variables for modeling: (C) Extraction of IC<sub>50</sub> and AUC from a drug's dose response curve in a given cell line. (D) Correlation and Scales of IC<sub>50</sub> vs AUC values. Blue line indicates a linear model fit through the data.

Full-size DOI: 10.7717/peerj.16342/fig-1

shows that integrating KINOMEScan data for 800 inhibitors vastly increases the kinome space targeted (Fig. 1B), compared to kinobeads alone.

However, it is important to note that despite the similarity in the assay readouts, the KINOMEScan and kinobeads assays are very different molecularly, and care needs to be taken while integrating these assay types. To assess the degree of difference between the two assay types, we retained all kinases captured by both assays (~250) and measured Pearson's correlation of inhibition states for the same compound between both assays. We found a reasonable degree of correlation (Pearson's  $R \sim 0.5$ ) across all data. In addition, we additionally found a much higher degree of agreement between data types (90%) when both assays were cut off at 80% inhibition (a "strong" hit in either assay) (Fig. S1). As a

result, keeping in mind the advantages of integrating both assay types, we moved forward with connecting these integrated kinome inhibition states to cell line responses.

### **Connecting inhibited kinome states to cancer cell line sensitivities from the DepMap repurposing screen**

To connect these kinase inhibitors and their induced inhibition states with their corresponding phenotypes in cancer cell lines, we make use of the DepMap repurposing screen, which uses the PRISM assay (Yu *et al.*, 2016) to run highly multiplexed cell viability assays. This dataset contains cell viability measurements for over 1,500 drugs profiled in 450 cell lines. From within this data, we found ~200 drugs for which we also have corresponding profiling data as described above.

The DepMap repurposing dataset provides cell viability measurements across multiple drug doses. However, since our dataset of kinome states is restricted to single-dose measurements, we extracted two single summary statistics for describing cell line sensitivity to kinase inhibitors (Fig. 1C): Dose–response area under the curve (AUC) and half-maximal inhibitory concentration (IC<sub>50</sub>). These properties are highly correlated with each other, having a Pearson’s correlation coefficient ~0.9 (Fig. 1D). We extracted these properties from DepMap and matched them to our kinome states. The final integrated dataset has ~250 drugs tested across ~450 cell lines, representing ~70,000 inhibitor–cell line combinations representing nearly all cancer types.

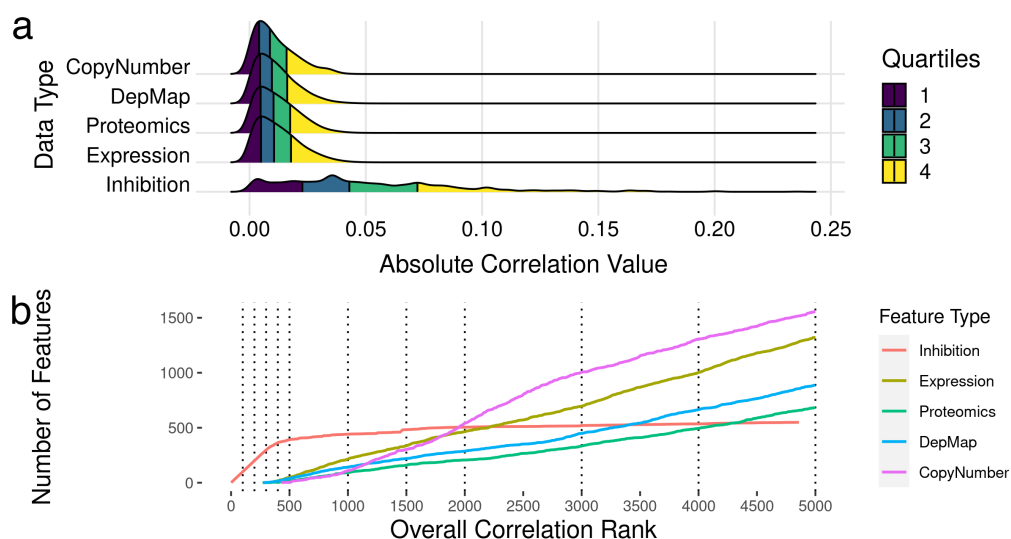
### **Examining bivariate association of features to cell line sensitivities provides a means for feature selection**

The 450 cell lines tested in the DepMap dataset also have accompanying baseline RNAseq gene expression data, so we integrated the ~20,000 TPM values for each cell line into the kinome-state and cell line sensitivity dataset. This adds baseline cell line-specific gene expression information to our cell line agnostic inhibitor-induced kinome states.

To identify which features were most correlated with drug sensitivity, we examined bivariate associations of each of the ~21,000 features (kinome inhibition states, baseline gene expression, gene essentiality, protein expression and copy number variation) were compared individually to the outcome of cell line sensitivities (dose response AUC and IC<sub>50</sub>), and their Pearson’s R correlation was calculated. We additionally calculated Spearman’s correlation coefficients, and ensured all correlated features were significant at the 0.5 level. The full table of feature correlations with significance values is attached in Table S2. We found that the most correlated feature is the drug-induced kinase inhibition state of TP53RK with a correlation coefficient  $R \sim 0.3$ , while the most correlated baseline gene expression value was OGFRL1 with a correlation coefficient  $R \sim 0.05$ . Overall, inhibitor-induced kinome states showed stronger correlation with cell line sensitivity metrics (Fig. 2A) despite there being 40× more baseline gene expression features than kinome states. However, it is important to note that this large difference in overall correlation may be partly due to the sample imbalance in abundance of kinome inhibition states (~200 per cell line) compared to baseline gene expression (one per cell line).

After exploring the relationship between features and cell line sensitivity, we sought to use machine learning models to integrate these features into predictive models of cell line





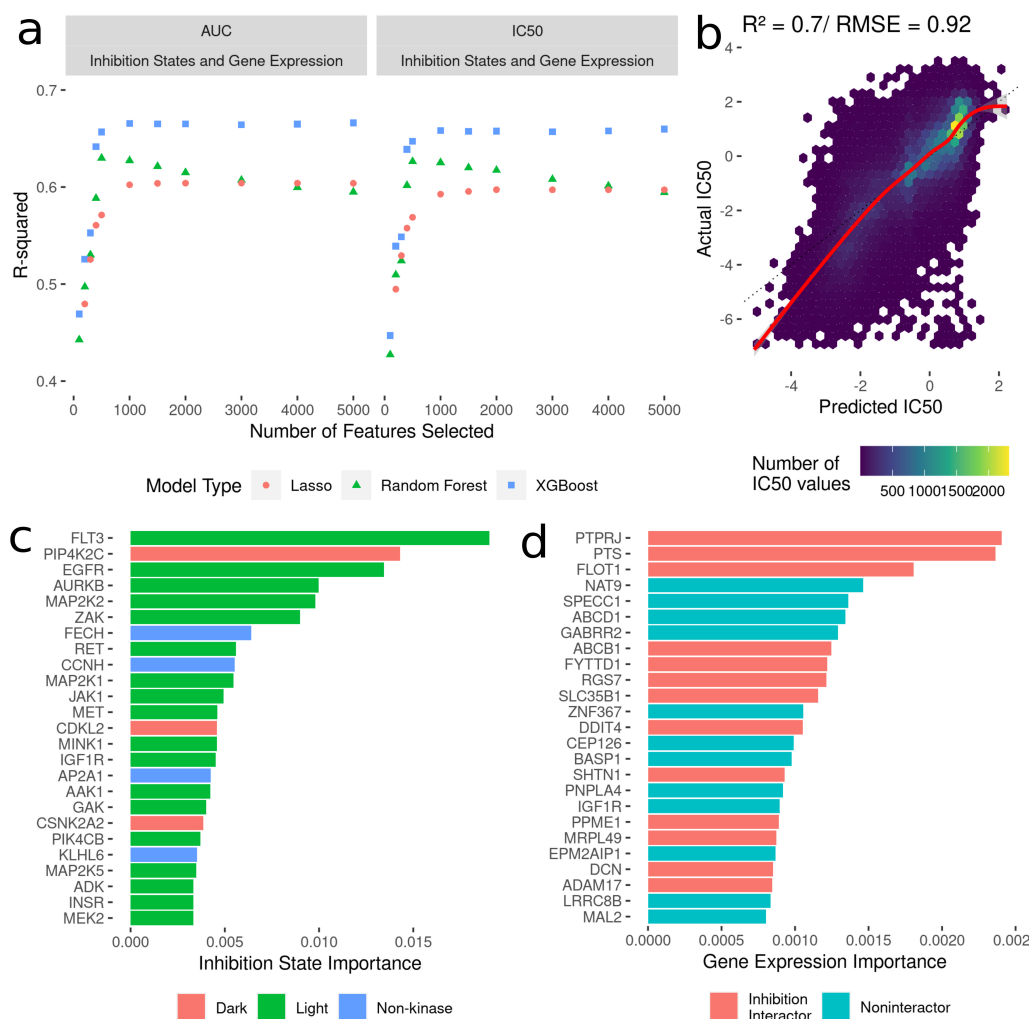
**Figure 2** Feature selection by bivariate association with cancer cell line sensitivity. (A) Ridgeline plot showing distributions of correlations with drug IC<sub>50</sub>s and AUC values across the data types included in analysis. (B) Plots showing what order classes of features are selected from the ranked set of inhibition states and baseline gene expression values. The dotted lines indicate the discrete increments of feature rank cutoffs at which model performance was tested. Kinase inhibition states were the most informative feature within the first ~300, after which gene expression features started to show predictive value.

Full-size [DOI: 10.7717/peerj.16342/fig-2](https://doi.org/10.7717/peerj.16342/fig-2)

sensitivity to kinase inhibitor treatment. To create a feature selection metric, all features were ranked by the absolute value of their Pearson's correlation to the cell sensitivity outcomes. Features were incorporated into models based on this ranking, *i.e.*, including only a top-ranked subset of best-correlated features. Seeking models with the best balance of predictive power and feature number, we tested model performance with different numbers of features, again selected starting from the most informative feature. We tested models with 100, 200, 300, 400, 500, 1,000, 2,000, 3,000, 4,000, and 5,000 features to see how adding more and more lower ranked features affected model performance. [Figure 2B](#) shows the ranking for all the feature sets, and the order they are added into the model.

### Machine learning models predict cancer cell line sensitivity from a combination of kinome inhibition states and baseline transcriptomics

To build machine learning models to predict cancer cell line AUC and IC<sub>50</sub> in response to treatment with kinase inhibitors, the highest ranked 100–5000 features were selected from the dataset linking drug-induced kinome states to cancer cell line responses ([Fig. 2B](#)). We compared three model types: LASSO regression, random forest and XGBoost. All models were trained with 10-fold random cross validation to minimize overfitting on the training data. Here, 10% of the dataset was randomly sampled and held out for testing, with the remaining 90% used for training. This was repeated ten times, and model performance was averaged across all the ten folds. This ensures comparable accuracy of the model predictions on new kinase inhibitors and cell lines. For each “batch” of feature numbers modelled from 100–5,000 we tuned sets of 30 hyperparameters for all model types ([Fig. 3A](#)). The R<sup>2</sup>



**Figure 3** Development of models to predict cancer cell line sensitivities to kinase inhibitors by integrating single-dose kinome profiling data. (A) Model performance metrics ( $R^2$ ) for LASSO (orange dot), Random Forest (green triangle) and XGBoost (blue square). (B) Scatterplot of predicted  $IC_{50}$  values from the best-performing model vs actual  $IC_{50}$  values. The red line indicated a smooth fit through the data points. (C) Horizontal bar plot showing model importance of individual kinase inhibition states by shapley values. (D) Horizontal bar plot showing model importance of individual baseline gene expression by shapley values.

Full-size DOI: 10.7717/peerj.16342/fig-3

value between predicted and actual value was utilized as the metric for model comparison. Overall, the 5000 feature XGBoost model performed the best with a cross-validation  $R^2$  of  $\sim 0.7$  (Fig. 3B).

Since tree-based machine learning models like XGBoost offer in-built explainability, it is possible to interrogate and explain which features were most important in predicting the outcome of cell line sensitivities. These importances generated *via* Shapley values (Rozemberczki et al., 2022) show kinase inhibition states to be overwhelmingly more important for predicting cell line responses when compared to baseline gene expression.

Kinases involved in cell cycle and proliferation are overrepresented in the top 25 features (Fig. 3C) (MAP2K, MEK2, CDKL5 *etc.*), and we further annotated each kinase with whether they are considered by current NIH guidelines as understudied (“Dark”) or well-characterized (“Light”) (Berginski *et al.*, 2021b; Oprea *et al.*, 2018). Interestingly four kinase interactor proteins are included as well, suggesting that interactions between inhibitors and non-kinases (off-target effects) have important consequences for cell viability. Baseline gene expression features show much lower model importances (Fig. 3D), but 40% of the top 25 genes have known interactions with kinases whose inhibition states are used in the model.

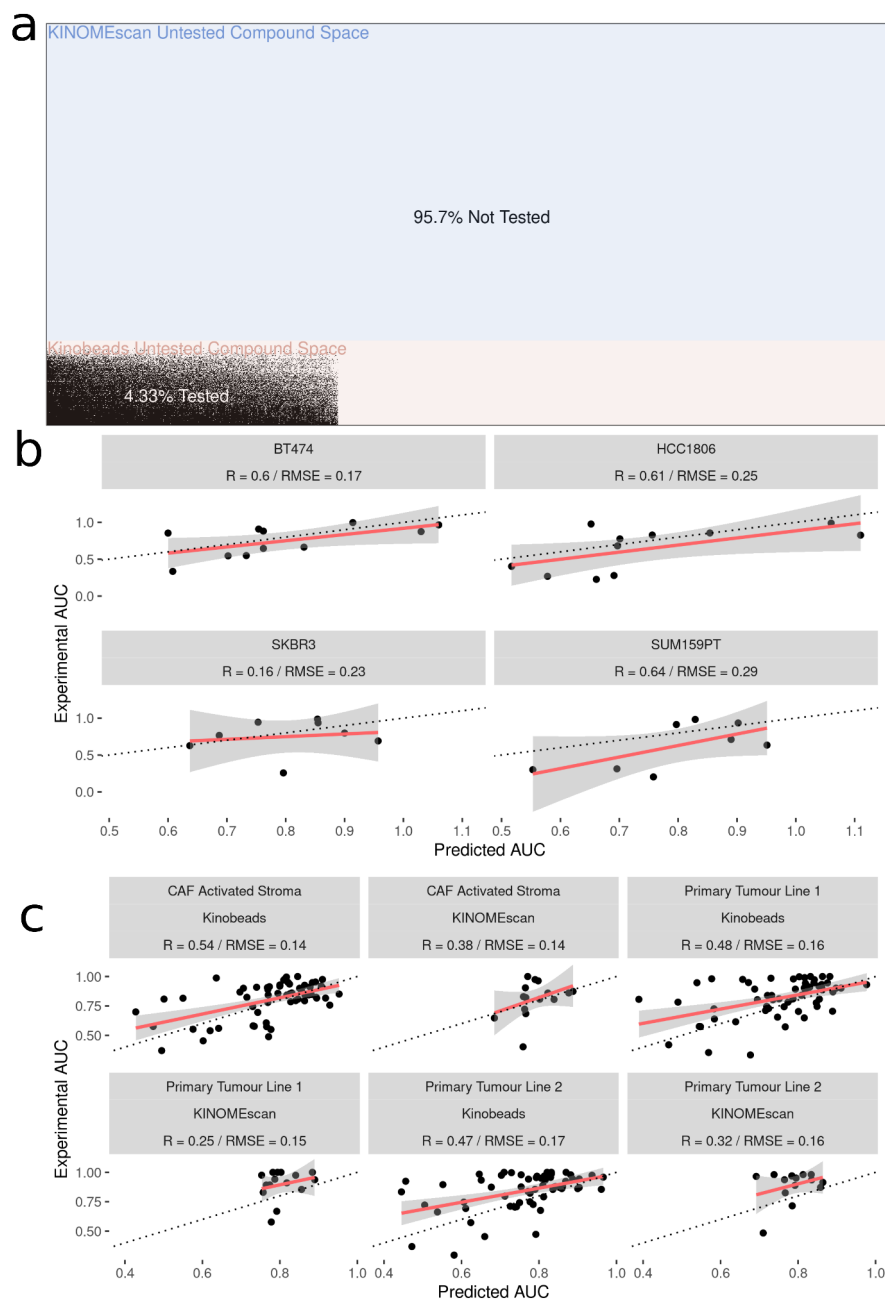
### **Inclusion of various multi-omics data with kinome inhibition states and gene expression did not improve model predictive performance**

In addition to the baseline gene expression data, all the cell lines in the DepMap database have three other profiling data types available: copy number variation, gene essentiality from CRISPR/KO, and baseline proteomics. To see if inclusion of these data into models would improve predictions, we integrated these with the modeling dataset of kinome inhibition states and gene expression and used identical modeling strategies described above to select correlated features, build, and evaluate LASSO, random forest, and XGBoost models (Fig. S2). We found that adding in the various multi-omic data types did not significantly outperform the models limited to kinase inhibition states combined with baseline gene expression ( $R^2$  of  $\sim 0.69$  for predicting  $IC_{50}$ ).

### **Experimental validation of model predictions were successful in characterized and novel cell lines**

After fitting the model on 70,000 cell line-drug combinations, predictions were made on 1.2 million unseen (not previously seen by the model) drug-cell line combinations (Fig. 4A). Approximately 90% of the untested inhibitors were associated with KINOMEscan datasets. As an initial validation, we tested a subset of the predictions in well-characterized breast cancer cell lines, two unseen by the models and two previously seen (HER2 positive: SKBR-3, BT-474 and two triple negative: SUM159, HCC1806). We analyzed the performance of the model on experimental data for unseen drug-cell line combinations, arriving at an R value  $\sim 0.6$  for all but one (SKBR3) breast cancer cell line (Fig. 4B). Notably, all the drugs tested had kinome profiling data from the Kinobeads assay.

We then further validated the model by predicting inhibitor effects from collected RNAseq data in tumor (two samples) and stroma (one sample) derived cell lines from PDAC patients (Lipner *et al.*, 2020; Berginski *et al.*, 2021a). Importantly, these patient-derived cell lines were profiled for baseline gene expression in-house and represent a challenging and highly heterogeneous transcriptional landscape which the model has not seen before. Dose–response AUC predictions were made by the model for 58 drugs with kinome profiling data from the Kinobeads assay and 18 drugs with kinome profiling data from the KINOMEscan assay. The model predicted AUCs were compared to experimentally generated AUCs, revealing an average  $R \sim 0.5$  for drugs with kinome profiling data from the Kinobeads assay tested in patient stroma-derived cell lines (Fig. 4C), and  $R \sim 0.4$  for drugs with kinome profiling data from the KINOMEscan assay. On the other hand, in



**Figure 4** Experimental validation of model in breast cancer cell and patient-derived PDAC cell lines. (A) Visualization of the space of compound (Y-axis) and cell line (X-axis) combinations that have been tested (white) and not tested (black) with colors denoting the origin of the drug kinome profiling data as Kinobeards (Pink) and KINOMEscan (blue). (B) Scatter plot showing relationship between AUC's predicted by model and experimentally generated AUC's for drugs not yet tested by PRISM in breast cancer cell lines (C) Scatter plot showing relationship between AUC's predicted by model and experimentally generated AUC's for drugs from Kinobeards and KINOMEscan tested in primary-tumor and stroma PDAC cell lines.

Full-size DOI: 10.7717/peerj.16342/fig-4

patient tumor-derived cell lines, drugs from kinobeads had model accuracy  $R \sim 0.49$ , and drugs from KINOMEScan had  $R \sim 0.3$  (Fig. 4C).

## DISCUSSION

Kinase inhibitors are one of the fastest growing classes of targeted cancer therapies, but only a small fraction of the druggable kinome has been explored to date (Laufer & Bajorath, 2022). We have previously used data describing kinobeads-derived kinome inhibition states to predict cell viability in cancer cell lines in response to clinical kinase inhibitors and shown high prediction accuracy. However, public databases of kinome inhibition states derived through the more accessible KINOMEScan assay cover many uncharacterized compounds and vastly widen the kinome space that can be targeted. In this work, we created a large integrated set of inhibitor-altered kinome states characterized by the kinobeads and KINOMEScan assays, representing a broad space of kinome targets and including a host of tool compounds targeting understudied kinases. We linked these kinome inhibition states to cancer cell line responses to kinase inhibitors (dose response AUC and  $IC_{50}$ ) and then built machine learning models that integrate these kinome states with cell line baseline gene expression values to predict cell line responses to kinase inhibitors. Finally, we predicted cell line sensitivity to previously untested kinase inhibitors in breast cancer and patient-derived PDAC cell lines and validated them experimentally.

Prediction of therapy response for cancer cell lines has been demonstrated through various methods, mostly utilizing chemical structure information, baseline gene expression and gene mutation status. Drug-target interaction data is relatively under-utilized for phenotype prediction, but offers opportunities for biological hypothesis building, especially for compounds with uncharacterized mechanisms of action. Kinome profiling data provides an exciting opportunity to use a wide array of functionally relevant drug-target interactions, and is almost unique (except for GPCR inhibitors and HDAC inhibitors) in terms of ability to assay potential off target interactions. While we have previously shown that kinome profiling data generated from the kinobeads assay is informative for cell viability prediction, KINOMEScan data is more readily accessible, publicly available for many uncharacterized compounds, and easier to generate. In addition, integration of this data vastly broadens the kinome space capable of being targeted and increases the number of inhibitors that can be virtually screened as potential therapeutics.

It is important to note that this model linking kinome inhibition states to cell line response is generalizable to any human cancer sample, provided the sample has baseline transcriptomic data available. We have shown in this work and previously that models can reasonably extrapolate to kinase inhibitors that have not been tested before in well-characterized cell lines. Significantly, in this work we have extended the scope of the model by using novel RNAseq data from patient-derived PDAC cell lines, and testing kinase inhibitors against both tumor and stroma cell lines and achieving reasonable prediction accuracy.

Using tree-based models like gradient boosting lends us the ability to explain to some degree which features most influenced cell viability. Shapley importance values

generated from the best-performing model show that the inhibition states of kinases had overwhelmingly more predictive power compared to baseline gene expression values, with FLT3 as the most important feature. FLT3 mutations are observed in 30% of acute myeloid leukemia (AML) patients, and various FLT3 inhibitors are commonly prescribed for treatment (*Antar et al., 2020*). However, the study dataset contains a majority of cell lines from non-small cell lung cancer (NSCLC), and FLT3 inhibitors have recently shown promise in preclinical studies by abrogating DNA damage (*Ryu et al., 2019*). Although the baseline gene expression features had considerably lower predictive power in the model, it is important to note that they provided crucial cell line specific context to the model, essentially “tuning” the model for specific cellular contexts. Interestingly, 40% of all the top 50 gene expression features were annotated as known kinase interactors in the STRING database. An additional strength of modeling cancer response from kinase-drug interactions is the generation of new hypotheses for understudied kinases. For example, the understudied (*Essegian et al., 2020*) “Dark” kinases PIP4K2C and CSNK2A2 appear in the top 25 features of the best-performing model, suggesting possible functional roles in cancer cell viability. Interestingly, PIP4K2C expression has been associated with outcomes in Acute Myeloid Leukemia (AML) (*Lima et al., 2019*), while CSNK2A2 has been associated significantly with prognoses of 14 different cancer types (*Strum, Gyenis & Litchfield, 2022*).

There are still many limitations to the results reported in this study. The creation of a combined kinome profiling dataset involves gluing together results from different assay types. Although the assays produce the same output (ratio of kinase in treatment sample to kinase in control), there may be numerous methodological artifacts that add noise to the data. We have attempted to address this by analyzing model performance on each assay type individually, and we can see that responses to inhibitors with data originating from KINOMEScan are noisier and more difficult to predict than inhibitors with data from kinobeads. This discrepancy is potentially due to the imbalance in training data availability for the KINOMEScan inhibitors, with only 15 inhibitors having annotated cell line sensitivity data available. In the future, as more cell line sensitivity testing is performed for compounds in the KINOMEScan dataset, model performance for this assay may improve. Additionally, model performance also decreases when shifting from using gene expression data obtained from well-characterized cancer cell lines to using in-house gene expression of novel patient-derived cell lines. This is potentially due to the innate and significant heterogeneity that exists in such samples and because the models have been trained on baseline transcriptomics set of the given ~450 cell lines.

Additionally, there are important limitations to our initial analysis of correlations to cell line responses. While we observe higher correlations overall for kinome inhibition states, these are complicated by the sample imbalance in our abundance of drug-specific inhibition states (~200 per cell line) to gene expression values (one per cell line). This imbalance, characterized by the ratio of drug-specific kinome inhibition states to cell-line specific baseline gene expressions, is not only inherent to our study design but also reflective of the clinical scenario we aim to model. In a real-world setting, a patient is likely to have a single point of genomic profiling, most plausibly from a pre-treatment biopsy. Subsequent acquisition of multiple “cell-induced” genomic signatures in response

to a variety of treatments is not only challenging but also often impractical in a clinical setting. This underscores the importance and value of drug-specific descriptors, like the kinome inhibition states in our study, which can be obtained from high-throughput assays such as kinobeads and kinomescan. Consequently, while the imbalance does complicate interpretation, it also reinforces the relevance and applicability of our approach. We caution that the observed higher correlation for kinome inhibition states may be partly due to this sample ratio imbalance. Nonetheless, our primary measure of the relative importance of the data types is through the described multivariate machine learning analyses. We believe that these methods, which leverage the full complexity of the data, provide a more robust and comprehensive assessment of the information contained in each data type.

It should be possible to extend these models to incorporate multiple kinase inhibitors in combination. This is significant, given the frequency of resistance to targeted cancer monotherapies (Lovly & Shaw, 2014) and the potential to escape the resistance acquisition from dynamic kinome reprogramming through multi-inhibitor combinations (Yesilkanal et al., 2021). Thus, another area of future work is to combine the kinome inhibition states of multiple inhibitors to gain an understanding of their dual effect on the kinome, connecting them to biological phenotypes that arise in response to inhibitor combinations to eventually build models and predict effective kinome-targeting combination therapies.

While targeted therapies such as kinase inhibitors have had significant clinical impact, much work remains to better understand how modulation of the kinome leads to both desirable and undesirable phenotypic effects. The results presented here provide an example of one approach where knowledge of the inhibition state of the kinome can be linked to downstream phenotypes through predictive models, greatly enhancing our ability to predict the effects of existing inhibitor treatments as well as facilitating the design of novel targeted therapies in the future.

## ACKNOWLEDGEMENTS

We would like to thank Madison Jenner for providing data access for our PDAC validation. We would like to thank UNC Research Computing for access to the computational resources necessary for this work.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by grants through the National Institutes of Health (Grant #s CA274298, CA233811, CA238475, DK116204). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

National Institutes of Health: CA274298, CA233811, CA238475, DK116204.

## Competing Interests

Shawn M. Gomez is an Academic Editor for PeerJ.

## Author Contributions

- Chinmaya U. Joisa conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Kevin A. Chen conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Matthew E. Berginski conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Brian T. Golitz performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Madison R. Jenner performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Gabriela Herrera Loeza performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Jen Jen Yeh analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Shawn M. Gomez conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

## Data Deposition

The following information was supplied regarding data availability:

The code is available at GitHub and Zenodo:

-[https://github.com/gomezlab/kinomescan\\_viability\\_prediction](https://github.com/gomezlab/kinomescan_viability_prediction).

-Chinmaya Joisa. (2023). Integrated single-dose kinome profiling data is predictive of cancer cell line sensitivity to kinase inhibitors. <https://doi.org/10.5281/zenodo.8415100>.

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.16342#supplemental-information>.

## REFERENCES

- Antar AI, Otrrock ZK, Jabbour E, Mohty M, Bazarbachi A. 2020. FLT3 inhibitors in acute myeloid leukemia: ten frequently asked questions. *Leukemia* 34(3):682–696 DOI 10.1038/s41375-019-0694-3.
- Berginski ME, Jenner MR, Joisa CU, Herrera Loeza SG, Golitz BT, Lipner MB, Leary JR, Rashid NU, Johnson GL, Yeh JJ, Gomez SM. 2021a. Kinome state is predictive of cell viability in pancreatic cancer tumor and stroma cell lines. *BioRxiv* DOI 10.1101/2021.07.21.451515.
- Berginski ME, Joisa CU, Golitz BT, Gomez SM. 2022. Kinome inhibition states and multiomics data enable prediction of cell viability in diverse cancer types. *BioRxiv* DOI 10.1101/2022.04.08.487646.



- Berginski ME, Moret N, Liu C, Goldfarb D, Sorger PK, Gomez SM. 2021b. The dark kinase knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Research* **49**(D1):D529–D535 DOI [10.1093/nar/gkaa853](https://doi.org/10.1093/nar/gkaa853).
- Chen T, Guestrin C. 2016. XGBoost: a scalable tree boosting system. ArXiv preprint. [arXiv:1603.02754](https://arxiv.org/abs/1603.02754).
- Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, Humeidi R, Peck D, Wu X, Tang AA, Wang VM, Bender SA, Lemire E, Narayan R, Montgomery P, Ben-David U, Garvie CW, Chen Y, Rees MG, Lyons NJ, McFarland JM, Wong BT, Wang L, Dumont N, O’Hearn PJ, Stefan E, Doench JG, Harrington CN, Greulich H, Meyerson M, Vazquez F, Subramanian A, Roth JA, Bittker JA, Boehm JS, Mader CC, Tsherniak A, Golub TR. 2020. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer* **1**(2):235–248 DOI [10.1038/s43018-019-0018-6](https://doi.org/10.1038/s43018-019-0018-6).
- Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, Bansal M, Ammad-ud din M, Hintsanen P, Khan SA, Mpindi J-P, Kallioniemi O, Honkela A, Aittokallio T, Wennerberg K, Collins JJ, Gallahan D, Singer D, Saez-Rodriguez J, Kaski S, Gray JW, Stolovitzky G. 2014. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology* **32**(12):1202–1212 DOI [10.1038/nbt.2877](https://doi.org/10.1038/nbt.2877).
- Drewry DH, Wells CI, Andrews DM, Angell R, Al-Ali H, Axtman AD, Capuzzi SJ, Elkins JM, Ettmayer P, Frederiksen M, Gileadi O, Gray N, Hooper A, Knapp S, Laufer S, Luecking U, Michaelides M, Muller S, Muratov E, Denny RA, Saikatendu KS, Treiber DK, Zuercher WJ, Willson TM. 2017. Progress towards a public chemogenomic set for protein kinases and a call for contributions. *PLOS ONE* **12**(8):e0181585 DOI [10.1371/journal.pone.0181585](https://doi.org/10.1371/journal.pone.0181585).
- Essegian D, Khurana R, Stathias V, Schürer SC. 2020. The clinical kinase index: a method to prioritize understudied kinases as drug targets for the treatment of cancer. *Cell Reports. Medicine* **1**(7):100128 DOI [10.1016/j.xcrm.2020.100128](https://doi.org/10.1016/j.xcrm.2020.100128).
- Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**(1):1–22.
- Gadagkar SR, Call GB. 2015. Computational tools for fitting the hill equation to dose-response curves. *Journal of Pharmacological and Toxicological Methods* **71**:68–76 DOI [10.1016/j.vascn.2014.08.006](https://doi.org/10.1016/j.vascn.2014.08.006).
- Gambacorti-Passerini C, Antolini L, Mahon F-X, Guilhot F, Deininger M, Fava C, Nagler A, Della Casa CM, Morra E, Abruzzese E, D’Emilio A, Stagno F, Le Coutre P, Hurtado-Monroy R, Santini V, Martino B, Pane F, Piccin A, Giraldo P, Assouline S, Durosini MA, Leeksa O, Pogliani EM, Puttini M, Jang E, Reiffers J, Piazza R, Valsecchi MG, Kim D-W. 2011. Multicenter independent assessment of outcomes in chronic Myeloid Leukemia patients treated with Imatinib. *Journal of the National Cancer Institute* **103**(7):553–561 DOI [10.1093/jnci/djr060](https://doi.org/10.1093/jnci/djr060).

- Gönen M, Margolin AA. 2014. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* 30(17):i556–i563 DOI 10.1093/bioinformatics/btu464.
- Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, Reinecke M, Ruprecht B, Petzoldt S, Meng C, Zecha J, Reiter K, Qiao H, Helm D, Koch H, Schoof M, Canevari G, Casale E, Depaolini SR, Feuchtinger A, Wu Z, Schmidt T, Rueckert L, Becker W, Huenges J, Garz A-K, Gohlke B-O, Zolg DP, Kayser G, Vooder T, Preissner R, Hahne H, Tonisson N, Kramer K, Gotze K, Bassermann F, Schlegl J, Ehrlich H-C, Aiche S, Walch A, Greif PA, Schneider S, Felder ER, Ruland J, Medard G, Jeremias I, Spiekermann K, Kuster B. 2017. The target landscape of clinical kinase drugs. *Science* 358(6367):eaan4368 DOI 10.1126/science.aan4368.
- Koleti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, Vidovic D, Forlin M, Kelley TT, D’Urso A, Allen BK, Torre D, Jagodnik KM, Wang L, Jenkins SL, Mader C, Niu W, Fazel M, Mahi N, Pilarczyk M, Clark N, Shamsaei B, Meller J, Vasiliauskas J, Reichard J, Medvedovic M, Ma’ayan A, Pillai A, Schurer SC. 2018. Data portal for the Library of Integrated Network-Based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Research* 46(D1):D558–D566 DOI 10.1093/nar/gkx1063.
- Laufer S, Bajorath J. 2022. New horizons in drug discovery—understanding and advancing different types of kinase inhibitors: seven years in kinase inhibitor research with impressive achievements and new future prospects. *Journal of Medicinal Chemistry* 65(2):891–892 DOI 10.1021/acs.jmedchem.1c02126.
- Lima K, Coelho-Silva JL, Sarti Kinker G, Pereira-Martins DA, Traina F, Fernandes PACM, Markus RP, Lucena-Araujo AR, Machado-Neto JA. 2019. PIP4K2A and PIP4K2C transcript levels are associated with cytogenetic risk and survival outcomes in acute myeloid leukemia. *Cancer Genetics* 233–234:56–66.
- Lipner MB, Peng XL, Jin C, Xu Y, Gao Y, East MP, Rashid NU, Moffitt RA, Loeza SGH, Morrison AB, Golitz BT, Vaziri C, Graves LM, Johnson GL, Yeh JJ. 2020. Irreversible JNK1-JUN inhibition by JNK-IN-8 sensitizes pancreatic cancer to 5-FU/FOLFOX chemotherapy. *JCI Insight* 5(8):e129905 DOI 10.1172/jci.insight.129905.
- Lovly CM, Shaw AT. 2014. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. *Clinical Cancer Research* 20(9):2249–2256 DOI 10.1158/1078-0432.CCR-13-1610.
- McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold approximation and projection. *Journal of Open Source Software* 3(29):861 DOI 10.21105/joss.00861.
- Oprea TI, Bologa CG, Brunak S, Campbell A, Gan GN, Gaulton A, Gomez SM, Guha R, Hersey A, Holmes J, Jadhav A, Jensen LJ, Johnson GL, Karlson A, Leach AR, Ma’ayan A, Malovannaya A, Mani S, Mathias SL, McManus MT, Meehan TF, Von Mering C, Muthas D, Nguyen D-T, Overington JP, Papadatos G, Qin J, Reich C, Roth BL, Schurer SC, Simeonov A, Sklar LA, Southall N, Tomita S, Tudose I, Ursu O, Vidovic D, Waller A, Westergaard D, Yang JJ, Zahoranszky-Kohalmi G. 2018.

- Unexplored therapeutic opportunities in the human genome. *Nature Reviews. Drug Discovery* **17**(5):317–332 DOI [10.1038/nrd.2018.14](https://doi.org/10.1038/nrd.2018.14).
- Patricelli MP, Nomanbhoy TK, Wu J, Brown H, Zhou D, Zhang J, Jagannathan S, Aban A, Okerberg E, Herring C, Nordin B, Weissig H, Yang Q, Lee J-D, Gray NS, Kozarich JW. 2011.** *In situ* kinase profiling reveals functionally relevant properties of native kinases. *Chemistry & Biology* **18**(6):699–710 DOI [10.1016/j.chembiol.2011.04.011](https://doi.org/10.1016/j.chembiol.2011.04.011).
- Reinecke M, Heinzlmeir S, Wilhelm M, Médard G, Klaeger S, Kuster B. 2019.** Kinobeads: a chemical proteomic approach for kinase inhibitor selectivity profiling and target discovery. In: *Target discovery and validation: methods and strategies for drug discovery*. Hoboken: Wiley-VCH Verlag GmbH & Co. KGaA, 97–130 DOI [10.1002/9783527818242.ch4](https://doi.org/10.1002/9783527818242.ch4).
- Rozemberczki B, Watson L, Bayer P, Yang H-T, Kiss O, Nilsson S, Sarkar R. 2022.** The shapley value in machine learning. ArXiv preprint. [arXiv:2202.05594](https://arxiv.org/abs/2202.05594).
- Ryu H, Choi H-K, Kim HJ, Kim A-Y, Song J-Y, Hwang S-G, Kim J-S, Kim D-U, Kim E-H, Kim J, Ahn J. 2019.** Antitumor activity of a novel tyrosine kinase inhibitor AIU2001 due to abrogation of the DNA damage repair in non-small cell lung cancer cells. *International Journal of Molecular Sciences* **20**(19):1–17 DOI [10.3390/ijms20194728](https://doi.org/10.3390/ijms20194728).
- Strum SW, Gyenis L, Litchfield DW. 2022.** CSNK2 in cancer: pathophysiology and translational applications. *British Journal of Cancer* **126**(7):994–1003 DOI [10.1038/s41416-021-01616-2](https://doi.org/10.1038/s41416-021-01616-2).
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, Von Mering C. 2021.** The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**(D1):D605–D612 DOI [10.1093/nar/gkaa1074](https://doi.org/10.1093/nar/gkaa1074).
- Szöcs E, Stirling T, Scott ER, Scharmüller A, Schäfer RB. 2020.** Webchem: an R package to retrieve chemical information from the web. *Journal of Statistical Software* **93**:1–17.
- Wells CI, Al-Ali H, Andrews DM, Christopher Asquith RM, Axtman AD, Dikic I, Ebner D, Ettmayer P, Fischer C, Frederiksen M, Futrell RE, Gray NS, Hatch SB, Knapp S, Lucking U, Michaelides M, Mills CE, Muller S, Owen D, Picado A, Saikatendu KS, Schroder M, Stolz A, Tellechea M, Turunen BJ, Vilar S, Wang J, Zuercher WJ, Willson TM, Drewry DH. 2021.** The Kinase Chemogenomic Set (KCGS): an open science resource for kinase vulnerability identification. *International Journal of Molecular Sciences* **22**(2):566 DOI [10.3390/ijms22020566](https://doi.org/10.3390/ijms22020566).
- Wright MN, Ziegler A. 2017.** Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**:1–17.
- Yesilkanal AE, Johnson GL, Ramos AF, Rosner MR. 2021.** New strategies for targeting kinase networks in cancer. *The Journal of Biological Chemistry* **297**(4):101128 DOI [10.1016/j.jbc.2021.101128](https://doi.org/10.1016/j.jbc.2021.101128).

Yu C, Mannan AM, Yvone GM, Ross KN, Zhang Y-L, Marton MA, Taylor BR, Crenshaw A, Gould JZ, Tamayo P, Weir BA, Tsherniak A, Wong B, Garraway LA, Shamji AF, Palmer MA, Foley MA, Winckler W, Schreiber SL, Kung AL, Golub TR. 2016. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nature Biotechnology* 34(4):419–423  
[DOI 10.1038/nbt.3460](https://doi.org/10.1038/nbt.3460).