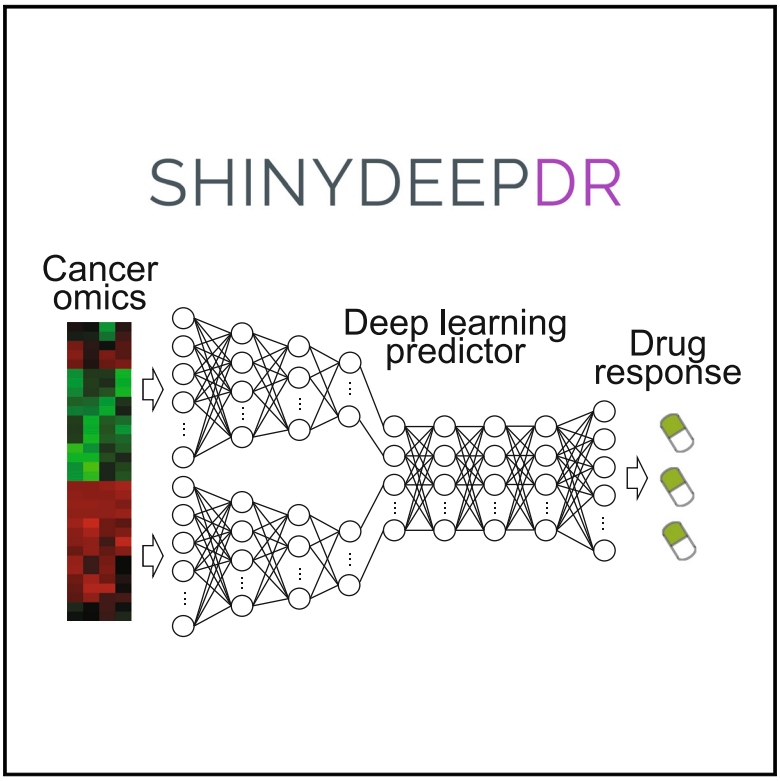
Descriptor

Patterns

shinyDeepDR: A user-friendly R Shiny app for predicting anti-cancer drug response using deep learning

Graphical abstract



Highlights

- shinyDeepDR is a web tool for predicting responses to 265 anti-cancer compounds
- It is applicable for researching both cancer cell lines and tumors
- Its interactive web interface facilitates result interpretation and investigation
- It identifies promising targets for an "undruggable" mutation in liver cancer

Authors

Li-Ju Wang, Michael Ning, Tapsya Nayak, ..., Yufei Huang, Yidong Chen, Yu-Chiao Chiu

Correspondence

cheny8@uthscsa.edu (Y.C.), yuc250@pitt.edu (Y.-C.C.)

In brief

shinyDeepDR is an online platform designed for the computational screening of 265 potential anti-cancer drugs. It utilizes a deep learning model called DeepDR. Users can simply input mutation and/or gene expression data from a cancer sample. shinyDeepDR performs two primary functions without the need for coding expertise or high-performance computers. The authors showcased the capability of the tool by identifying potential drugs for an "undruggable" mutation in hepatocellular carcinoma, the predominant type of liver cancer.

Wang et al., 2024, Patterns *5*, 100894
February 9, 2024
https://doi.org/10.1016/j.patter.2023.100894

CellPress

Patterns

CellPress
OPEN ACCESS

Descriptor

shinyDeepDR: A user-friendly R Shiny app for predicting anti-cancer drug response using deep learning

Li-Ju Wang,[1] Michael Ning,[1] Tapsya Nayak,[2] Michael J. Kasper,[1] Satdarshan P. Monga,[3,4,5] Yufei Huang,[5,6,7,8] Yidong Chen,[2,9,*] and Yu-Chiao Chiu[1,4,5,10,*]

[1]Cancer Therapeutics Program, University of Pittsburgh Medical Center Hillman Cancer Center, Pittsburgh, PA 15232, USA
[2]Greehey Children's Cancer Research Institute, University of Texas Health San Antonio, San Antonio, TX 78229, USA
[3]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA
[4]Pittsburgh Liver Research Center, University of Pittsburgh Medical Center and University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA
[5]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA
[6]Cancer Virology Program, University of Pittsburgh Medical Center Hillman Cancer Center, Pittsburgh, PA 15232, USA
[7]Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
[8]Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15261, USA
[9]Department of Population Health Sciences, University of Texas Health San Antonio, San Antonio, TX 78229, USA
[10]Lead contact
*Correspondence: cheny8@uthscsa.edu (Y.C.), yuc250@pitt.edu (Y.-C.C.)
https://doi.org/10.1016/j.patter.2023.100894

THE BIGGER PICTURE Understanding how different genomic attributes affect drug responses in cancer is crucial for personalized oncology. Deep learning, an advanced computational method, has demonstrated significant potential in identifying and predicting these intricate interactions. One such example is the DeepDR model, which predicts how cancer cells respond to drugs. However, not all researchers have the computational resources and programming expertise to leverage this potential. Here, we introduce shinyDeepDR to bridge this gap by providing an intuitive and user-friendly web platform to access DeepDR. In the broader scope, we envision that tools like shinyDeepDR will advance cancer research by making sophisticated computational models more FAIR (findable, accessible, interoperable, and reusable).

Development/Pre-production: Data science output has been rolled out/validated across multiple domains/problems

SUMMARY

Advancing precision oncology requires accurate prediction of treatment response and accessible prediction models. To this end, we present shinyDeepDR, a user-friendly implementation of our innovative deep learning model, DeepDR, for predicting anti-cancer drug sensitivity. The web tool makes DeepDR more accessible to researchers without extensive programming experience. Using shinyDeepDR, users can upload mutation and/or gene expression data from a cancer sample (cell line or tumor) and perform two main functions: "Find Drug," which predicts the sample's response to 265 approved and investigational anti-cancer compounds, and "Find Sample," which searches for cell lines in the Cancer Cell Line Encyclopedia (CCLE) and tumors in The Cancer Genome Atlas (TCGA) with genomics profiles similar to those of the query sample to study potential effective treatments. shinyDeepDR provides an interactive interface to interpret prediction results and to investigate individual compounds. In conclusion, shinyDeepDR is an intuitive and free-to-use web tool for *in silico* anti-cancer drug screening.

INTRODUCTION

Understanding and predicting relationships between cancer genomics and the response of cancer cells to anti-cancer treatments are key to identify new therapeutic agents and develop effective drug repurposing strategies. With the advent of high-throughput sequencing techniques and drug screening methods, global consortia have profiled the genomics of thousands of pan-cancer cell lines and tumors, such as the Cancer Cell Line Encyclopedia (CCLE)[1] and The Cancer Genome Atlas (TCGA),[2] and systematically screened hundreds of anti-cancer compounds, such as the Genomics of Drug Sensitivity in Cancer (GDSC).[3] These extensive data resources have led to the development of advanced computational models that predict pharmacogenomic associations between baseline (i.e., pretreatment) genomics of cancer samples and their drug sensitivities. Given the complex and non-linear nature of these pharmacogenomic patterns, likely governed by both genetic and transcriptomic features,[1,4] deep learning models are a viable approach for capturing and predicting such relationships,[5–11] as benchmarked by a recent study.[12]

Our previously developed DeepDR model integrates gene mutation and expression profiles of a cancer sample to predict its response to all compounds in the GDSC library.[5] The model's unique transfer learning design effectively integrates drug screens of cell lines and genomics of tumors, enabling accurate predictions of real-world therapy responses in patients with cancer.[5] It has since attracted extensive attention among researchers.[13–16] However, the implementation of deep learning models, typically using Python, can pose significant challenges for biomedical researchers with limited programming expertise or computational resources. To the best of our knowledge, PaccMann[17] and DrVAEN[12] are the only published web servers that implement a deep learning drug response predictor. Given that both tools rely on gene expression profiles for predicting anti-cancer drug sensitivity, their capacity to encompass genetic context related to gene mutations is constrained, which is yet pivotal in informing current precision oncology practices.[18] Furthermore, the tools were developed solely by data of cell lines, thereby hindering straightforward applications to tumor samples. Due to computational complexity, neither tool supports simultaneous prediction of hundreds of compounds.

To address the unmet needs, here, we present shinyDeepDR, a web-based implementation of DeepDR designed for predicting the sensitivity to a broad panel of 265 anti-cancer drugs. Built on an R Shiny framework integrated with Python, HTML, and JavaScript, shinyDeepDR performs deep learning computation and enables interactive exploration of prediction results. Specifically, our tool integrates mutation and gene expression features, enables prediction of drug responses in both cell line and tumor samples, delivers well-annotated prediction results, facilitates user-friendly analyses, and compares predictions to real data generated from drug screening. With two main query modules, the goal of shinyDeepDR is to empower users to easily search for potentially effective anti-cancer drugs for a cancer sample (cell line or tumor) using mutation and/or expression data (Figure 1).

(1) Module 1 – Find Drug (core module) utilizes DeepDR to predict the query sample's response to 265 approved and investigational anti-cancer compounds in the GDSC library.
(2) Module 2 – Find Sample identifies cell lines or tumors from the CCLE and TCGA databases with similar genomics features to those of the query sample and examines their real or predicted drug responses.

The query results are presented through an intuitive interface consisting of interactive figures and data tables, allowing straightforward downstream analyses, candidate prioritization, and interpretation of results. In this descriptor, we showcase the potential application of shinyDeepDR by studying a prevalent "undruggable" gene mutation in *CTNNB1* and its frequently co-occurring activation in MET in hepatocellular carcinoma (HCC). Given the broad research interest in DeepDR and other deep learning models, we expect that the introduction of shinyDeepDR could increase accessibility of deep-learning-based drug response predictors and facilitate the development of anti-cancer treatments. shinyDeepDR is freely accessible at https://shiny.crc.pitt.edu/shinydeepdr/.

IMPLEMENTATION

DeepDR models, cancer genomics datasets, and drug screening datasets

Full DeepDR model and associated datasets. The core of shinyDeepDR is the DeepDR model, our deep learning predictor for drug sensitivity.[5] In brief, the full DeepDR model takes sample-paired mutation and gene expression data from a cancer sample (cell line or tumor) as inputs and predicts the response to 265 potential anti-cancer compounds included in the GDSC drug screening library.[3] As shown in Figure 1, DeepDR considers binary mutation status in 18,281 genes (1: missense and nonsense mutations and frameshift insertions and deletions; 0: otherwise) and continuous expression levels of 15,363 genes (log2[TPM+1] values, where TPM denotes transcripts per million). As we described in Chiu et al.,[5] genomics data for cell lines and tumors were downloaded from the CCLE/CTD[19–21] and TCGA/TumorMap databases,[2,21] respectively. Drug sensitivity data for cell lines were downloaded from GDSC.[3] The multitask output of DeepDR contains continuous IC_{50} values (on a log [μM] scale) of 265 anti-cancer compounds for a cancer sample. DeepDR was trained with a two-step transfer learning approach. Step 1 is "unsupervised pretraining" of an autoencoder on high-dimensional mutation data (and another autoencoder on gene expression data) using TCGA pan-cancer tumors (n = 9,059) in order to capture tumor-relevant data representation. Step 2 is "supervised fine-tuning" to optimize features that predict drug sensitivity by connecting the pretrained encoders (i.e., the dimension-reducing component of an autoencoder) to a prediction network and then training the entire network using drug sensitivity data from GDSC (622 cell lines). An 80-10-10 partition of cell lines was utilized for model training, validation, and testing, respectively. Technical details of DeepDR, such as data preprocessing, model construction, implementation, benchmarking, and optimization of hyperparameters, were described previously.[5]

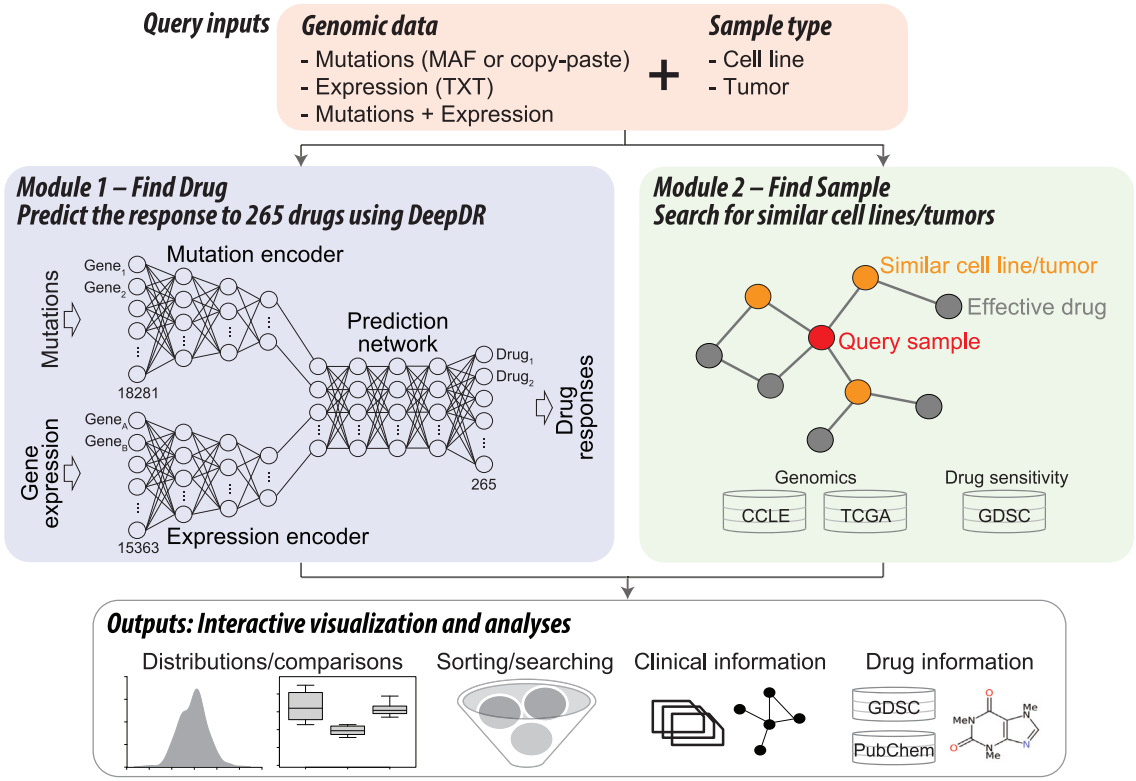


Figure 1. Server overview

shinyDeepDR is an R Shiny implementation of DeepDR, our deep learning model for predicting drug sensitivity.[5] Given mutations and/or gene expression profiles in a cancer sample (human cell line or tumor), shinyDeepDR runs two main modules. "Module 1 – Find Drug" is the core function of shinyDeepDR that uses DeepDR to predict the query sample's response to 265 anti-cancer compounds. DeepDR has a transfer learning scheme that incorporates features of tumors and cell lines (described and evaluated in Chiu et al.[5]). Thus, it is applicable to both tumors and cell lines. "Module 2 – Find Sample" searches for cell lines across the Cancer Cell Line Encyclopedia (CCLE; n = 704) or tumors of The Cancer Genome Atlas (TCGA; n = 9,059) with similar genomics features to the query sample and displays their drug response. We incorporate real drug response data of pan-cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC; 265 compounds × 704 cell lines) and predictions made by DeepDR, as well as predicted data of tumors (265 compounds × 9,059 tumors) from our paper.[5]

Simplified DeepDR models. Since sample-paired mutation and expression data may not always be readily available, we implemented two simplified DeepDR models: one taking mutations alone, and the other taking gene expression alone. We evaluated the performance of the full and two simplified models by per-cell line and per-drug correlation coefficients in testing cell lines. Since the simplified models generally performed less well than the full model,[5] for each simplified design, we trained 10 models and implemented the best-performing one in shinyDeepDR.

Other datasets. For implementing module 2 of shinyDeepDR, we used up-to-date gene mutation (binary status of 18,281 genes, as defined above) and expression data (15,363 log2 [TPM+1] values) in CCLE, downloaded from the Cancer Dependency Map portal (https://depmap.org/portal/; v.22Q2). This yielded a total of 704 cell lines. Drug annotation data were downloaded from the GDSC and PubChem[22] databases.

R and web environment of shinyDeepDR

shinyDeepDR was implemented using R (v.4.2.1) and R Shiny (v.1.7.2). Deep learning computation was performed by using the R TensorFlow library (v.2.9.0) with a backend of TensorFlow (v.2.10.0) in Python 3.8. Other main R libraries for data processing and visualization included Plotly, Tidyverse, visNetwork, and rcdk. shinyDeepDR is hosted on a high-performance computing node at the Center for Research Computing of the University of Pittsburgh. The machine is deployed on the VMWare infrastructure of the Center for Research Computing. The initial virtual machine has 8 Intel cores and 64 GB RAM. To accommodate potential large usage, the computational resources, such as additional CPU cores or more memory, can be quickly increased by re-deploying the Shiny app.

Implementation of "Module 1 – Find Drug"

Module overview. Module 1 is the main query module of shinyDeepDR that predicts a cancer sample's response to 265 anti-cancer compounds of the GDSC library using a prebuilt DeepDR model, taking mutations alone, gene expression alone, or both data types. Since the DeepDR models were originally implemented using Keras v.1.2.2, we updated all models to Keras v.2.9.0 without changing the model parameters and implemented them using the R TensorFlow library. When a user uploads a mutation and/or a gene expression profile, the tool maps the input data to the requirements of DeepDR (mutations of 18,281 genes and expression of 15,363 genes). Each unmapped gene, if any, is imputed by the median value across the 622 cell lines used for constructing the DeepDR models.

An interactive output interface is provided to summarize, sort, search, analyze, and annotate the query sample's predicted responses to 265 compounds.

Input data preparation. shinyDeepDR only supports human genomics data. For mutation data, we recommend using Mutation Annotation Format (MAF) files following the format of cBioPortal[23,24] or TCGA, which contains the columns "Hugo_Symbol," "Variant_Classification," and "Tumor_Sample_Barcode." For gene expression data, to minimize the number of unmapped genes, we recommend users generate their input data following TCGA "mRNA Expression Workflow" with the "Original" setting and GENCODE v.22 annotation. If FPKM (fragments per kilobase of exon per million mapped fragments) values are uploaded, shinyDeepDR provides a conversion tool to TPM using methods described in Pachter et al.[25]

Input sample type. The input interface allows users to specify whether the sample under study is a cancer cell line or tumor in order to provide appropriate reference ranges for predictions. If the query sample is specified as a cell line, our tool provides the ranges of real and predicted log(IC_{50}) values for each drug among 622 cancer cell lines and corresponding percentiles of the predicted value for the query sample. If a tumor is specified, the range and percentile are determined based on predicted responses to the drug across 9,059 TCGA tumors using the full DeepDR model.

Drug-sample association networks. As one of the outputs, our tool generates an interactive drug-sample-drug association network to explore a given drug among the top predictions. When a user selects a drug of interest, a network is created with the drug being the hub, the 10 samples most sensitive to the drug as the first-degree nodes, and the 10 most effective drugs in each sample as the second-degree nodes. The network is constructed using real GDSC data for cell lines and predicted data for tumors by the full DeepDR model.

Implementation of "Module 2 – Find Sample"

Module 2 is designed to supplement module 1 by identifying cancer cell lines or tumors that share similar genomics profiles with the query sample. Given a query sample and a specified sample type, our tool searches through all 704 cell lines of the CCLE or 9,059 tumors of TCGA. The similarity between the query sample and samples in our database is measured and ranked by the Jaccard index for mutated genes among 18,281 genes and by Pearson correlation coefficients among 15,363 genes for gene expression data. When sample-paired data are uploaded, the tool averages the ranks of mutations and gene expression to determine the overall ranking of CCLE or TCGA samples. Real and DeepDR-predicted drug response data for CCLE samples, or predicted data for TCGA samples, are displayed and can be interactively analyzed as in module 1. The tool generates a sample-drug network to show and analyze the drugs that were most effective among CCLE or TCGA samples most similar to the query (Figure 1).

Implementation of a supplemental module for analyzing predictions of TCGA samples with gene alterations

shinyDeepDR has a supplemental module, namely "Analyze TCGA," that provides easy access to pregenerated predictions of drug responses across all TCGA samples as documented in Chiu et al.[5] The module enables users to explore the relationships between drug sensitivities and gene mutations or aberrant expressions. Users can interact with the tool by selecting the symbol of a specific gene and the type of alteration of interest (mutated vs. wild type, top 25% high expression vs. others, or bottom 25% low expression vs. others). The module identifies drugs achieving significantly enhanced responses in tumors harboring specific gene alterations in a pan-cancer or cancer-type-specific manner. Statistical significance is assessed by the one-tailed t test. Additionally, the tool supports the analysis incorporating two genes by conducting a bivariate analysis that compares drug responses among four sample groups, which are defined by the presence or absence of the two gene alterations. This extended analysis provides additional insights into the potential interplay between the two genes and their impact on drug responses. Statistical significance is assessed by the one-way analysis of variance (ANOVA) test. The results generated by the module are presented through a comprehensive data table and boxplots, enabling users to explore the specific drug responses, associated statistical information, and results interpretation.

RESULTS

shinyDeepDR usage

shinyDeepDR is an interactive web tool that provides a user-friendly platform to identify and study drugs that may be effective in a cancer sample by using our deep learning predictor (DeepDR) and leveraging high-throughput screening data (Figure 2A). To begin the analysis, a user can simply upload a mutation profile (MAF file or copy-and-paste list of mutated genes) and/or a gene expression profile (TXT file) of a human sample, either a cell line or tumor (patient-derived model or research tumor sample) (Figure 2B). The input interface allows users to specify whether the sample is a cell line or tumor in order to use appropriate databases and reference ranges. Module 1 runs DeepDR and generates the response to 265 potential anti-cancer compounds. Empowered by the transfer learning design, DeepDR is capable of predicting drug response in both cell lines and tumors.[5] The output interface contains interactive figures, networks, and tables to inform drug prioritization; reference ranges derived from real and predicted responses within CCLE (n = 622) or TCGA (n = 9,059) samples; cancer type/subtype analyses; drug structure visualization and drug annotations (such as mechanism of action, stage of clinical development, and PubChem information); and drug-sample associations (examples in Figure 2C). Module 2 was specifically designed to complement the information provided by module 1. Its objective is to identify cancer samples in our database (cancer cell lines in CCLE or tumors in TCGA) that exhibit genomics profiles comparable to the query sample and then analyze the samples' sensitivities to drugs. Output panels include prioritization of similar CCLE or TCGA samples, detailed clinical information of the samples, and an interactive network to visualize which drugs may be effective in these samples. Comprehensive documentation, help information regarding each input and output panel, and examples are provided with the web tool.

Performance evaluation

The shinyDeepDR web server is highly efficient. Running each module takes about 10 s. In our publication providing details

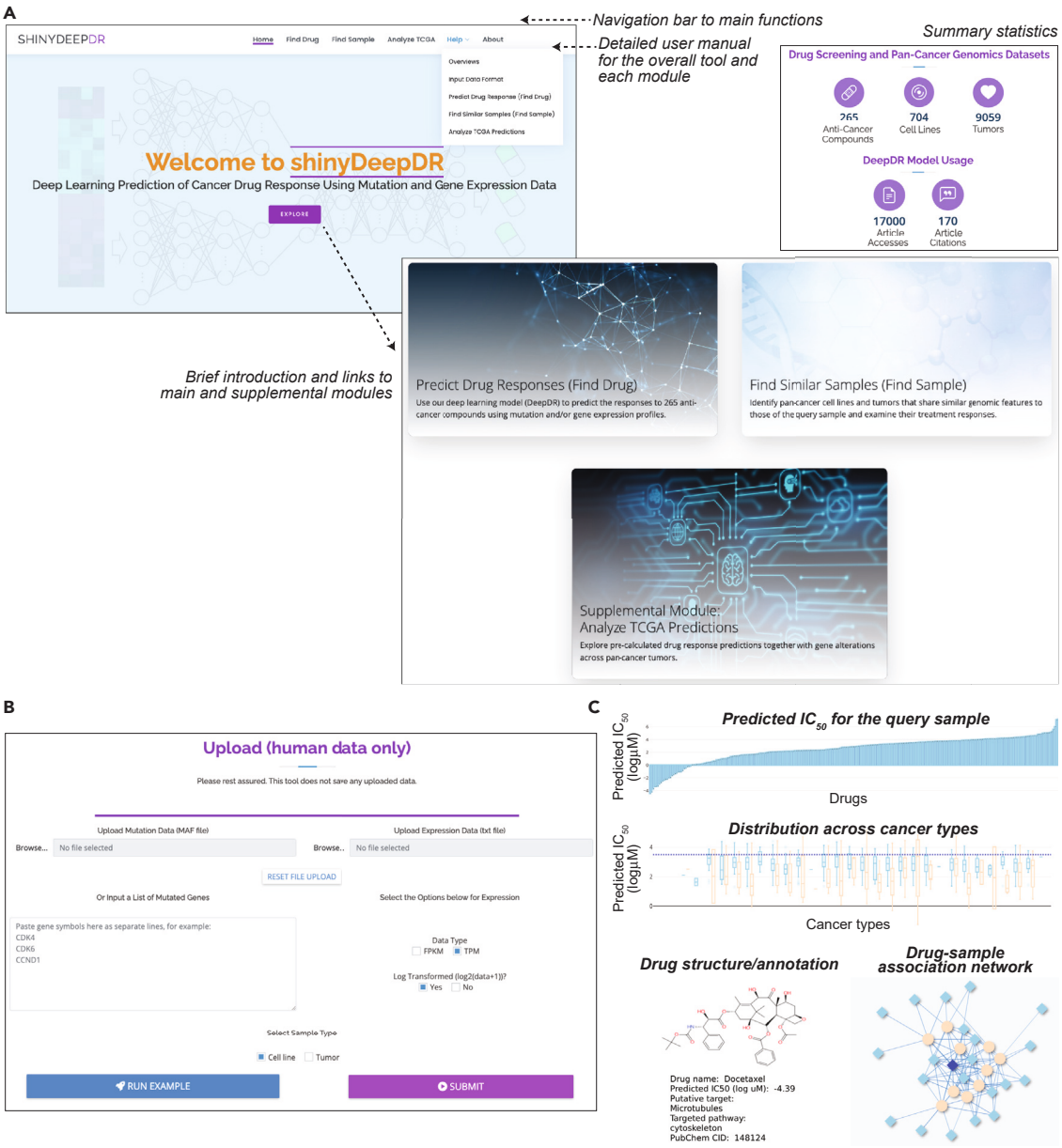


Figure 2. Server usage

(A) Screenshots of the homepage. The homepage provides a concise summary of our tool and the datasets used, as well as links to the main and supplemental modules. Users can test each module by using our built-in example files. A detailed user manual and figure/table legends are provided along with individual output panels and in the help pages.

(B) Screenshots of the input interface. shinyDeepDR takes the inputs of gene mutation (users can input a list of mutated genes into the text box or upload an MAF file) and/or expression profiles (a TXT file) of a cancer sample. Users can specify the sample type (cell line or tumor) to use appropriate reference ranges for prediction results, as well as the data type and normalization of gene expression data.

(C) Screenshots of the output panels. Query results are presented via a user-friendly interface of interactive figures and networks, as well as easy-to-use data tables, enabling meaningful downstream analyses and interpretation of results. We also provide detailed clinical and biochemical annotations to cell lines, tumors, and drugs in the tool.

for DeepDR,[5] we systematically evaluated prediction performance using hold-out cancer cell lines by multiple measures, including mean squared errors in drug response and per-cancer cell line correlation coefficients between real and predicted data (Pearson and Spearman correlation coefficients, 0.74–0.95 and 0.70–0.92, respectively; Table S1). Results with DeepDR were markedly improved over conventional methods, including linear regression, support vector machine, and alternative deep learning models trained either with cell lines alone without transferring features learned from tumors or using principal

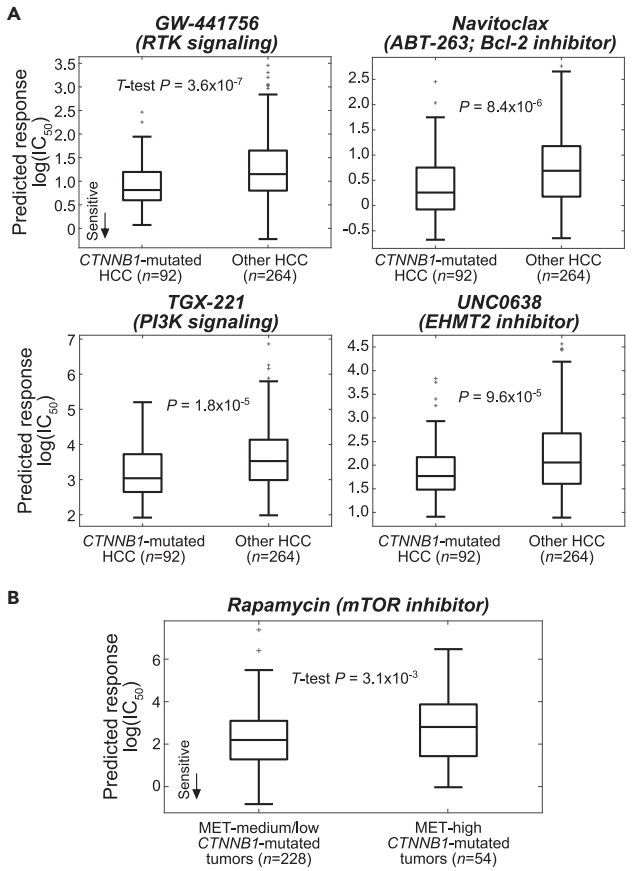


Figure 3. Application of shinyDeepDR to study *CTNNB1* mutations in TCGA

(A) Top-predicted drugs effective against *CTNNB1* mutations in hepatocellular carcinoma (HCC). shinyDeepDR was applied to 356 HCC tumors in TCGA (TCGA-LIHC). Results for four drugs with significantly increased sensitivities (smaller IC_{50} values) in *CTNNB1*-mutated HCC tumors compared to other HCC samples are shown. Statistical significance was assessed by a one-tailed t test. Detailed results of all nine drugs with statistically significant results ($p < 0.0001$) are summarized in Table S3. This plot can be generated with the "Analyze TCGA" module of shinyDeepDR by selecting the "LIHC: liver hepatocellular carcinoma" project and the "CTNNB1_mut" gene alteration. We edited the output plot using the Adobe Illustrator software to ensure the visual quality for the publication purpose.

(B) Validation of MET-modulated response to rapamycin in *CTNNB1*-mutated tumors. Pan-cancer tumors harboring *CTNNB1* mutations were studied for their sensitivity to rapamycin, with or without activated MET. Activation of MET was determined by the 75th percentile of its expression levels across all TCGA tumors. Statistical significance was determined by a one-tailed t test. To generate the figure, a user can utilize the same module of shinyDeepDR, configuring it for a pan-cancer analysis ("PanCan") with two gene alterations of interest set as "CTNNB1_mut" and "MET_exp_high." In order to assess the statistical significance between the two groups using a one-tailed t test instead of the ANOVA provided by the tool, the results table together with sample groups can be downloaded from our tool, and a subsequent statistical test can be performed using user-friendly software such as Microsoft Excel or Prism.

components to replace encoder outputs. Furthermore, we validated the predictions with real-world clinical data from patients, such as an approved estrogen receptor agonist (tamoxifen) for breast cancer, approved drugs targeting the *EGFR* mutations (afatinib and gefitinib) for non-small cell lung cancer, and an investigational compound, CX-5461, for the potential to treat hematopoietic malignancies. To enable the tool's use for samples with only single-omics data (mutation or gene expression), we have implemented the two best-performing simplified DeepDR models in shinyDeepDR (see implementation). We confirmed that their per-cell line performance was comparable to that of the full model that takes both data types (Table S1). Additionally, we evaluated the performance with respect to individual drugs obtained in the full and simplified models (Table S2) and provided those results on the web server, along with prediction results.

Case study 1: Identifying potential treatments for *CTNNB1*-mutated HCCs

HCC is the most dominant type of liver cancer in adults and accounts for ~90% of all primary liver tumors. Liver cancer is the only cancer type that continues to show increased incidence and mortality rates in both men and women in the United States,[26] indicating an urgent need to identify better therapeutic targets.[27] Unfortunately, one of the most prevalent mutations in HCC—*CTNNB1* that encodes β-catenin—remains undruggable.[28,29] As a case study, we used shinyDeepDR to predict drug sensitivity among HCC samples in TCGA (TCGA-LIHC; n = 356) to identify potential therapeutic targets for β-catenin. For each tumor, shinyDeepDR predicted response to 265 compounds using the full DeepDR model. These prediction results are made available to users through a supplemental query module ("Analyze TCGA") of shinyDeepDR. We identified nine drugs that were significantly more effective in tumors harboring *CTNNB1* mutations (n = 92) compared to others (n = 264) (one-tailed t test, $p < 0.0001$; Figure 3A; Table S3). These drugs target critical cancer pathways, such as cell apoptosis (navitoclax, a Bcl-2 inhibitor, and serdemetan, a p53 activator); *EHMT2* and chromatin histone methylation (UNC0638); RTK (GW-441756 and SB52334) and PI3K/Akt (TGX-221) signaling pathways; and DNA-dependent protein kinases (NU-7441). Since *CTNNB1* mutations are considered undruggable, targeting synthetic lethal interactions may represent a promising therapeutic strategy.[30,31] Indeed, our prediction identified signaling pathways that are functionally relevant to the WNT/β-catenin pathway. For instance, RTK signaling, which may induce PI3K/Akt signaling,[32] alternatively activates β-catenin and could be a potential therapeutic approach for treating subsets of cancers driven by aberrant β-catenin activation.[33] Additionally, UNC0638 is an inhibitor of *EHMT2*, another key gene that activates the Wnt/β-catenin signaling pathway in HCC.[34] Of note, navitoclax was significantly more effective in *CTNNB1*-mutated cancer cells in an earlier drug screening study by the Broad Institute, namely the Cancer Therapeutics Response Portal (CTRP).[35] Taken together, our analyses revealed known drugs and identified promising candidates that warrant further *in vitro* and *in vivo* investigations.

Case study 2: Verifying the modulatory effect of MET on the response of *CTNNB1*-mutated tumors to rapamycin

Co-occurring alterations in *CTNNB1* mutations and Met activation represent approximately 11% of HCC cases, according to our re-analyses of TCGA and French cohorts.[36,37] To devise therapeutic strategies for this co-occurrence, we previously

established murine HCC models that co-expressed clinically relevant mutant-*CTNNB1* and human MET using sleeping beauty transposon/transposase and hydrodynamic tail vein injection.[38] In these mice, HCC burden decreased significantly after rapamycin treatment, which inhibits mammalian target of rapamycin (mTOR). Furthermore, combination therapy of rapamycin and GC-1 (a coincidental Met inhibitor) led to an even stronger response.[38] To verify this finding in human tumor data, we predicted the response of *CTNNB1*-mutated tumors to rapamycin and explored the association with MET. Because of the limited sample size for such a bivariate analysis, we expanded the analysis to the pan-cancer scale (TCGA; n = 9,059). We used the supplemental query module of shinyDeepDR to perform an integrative analysis of predicted drug responses of tumors with gene expression data. The analysis revealed that *CTNNB1*-mutated tumors with medium-to-low MET expression (below 75th percentile; n = 228) would respond significantly better than those with highly expressed Met (n = 54) (one-tailed t test, p = 0.0031; Figure 3B). Thus, data from our tool provide additional evidence of the interplay of *CTNNB1* and MET with mTOR, as well as the impact of that interplay on treatment response.

DISCUSSION

To our knowledge, shinyDeepDR is the first web tool that integrates cancer mutation and gene expression features to predict and inform drug responses for both cell line and tumor samples. The R Shiny framework offers a user-friendly and efficient interface for interactive visualization and analysis. It provides interpretable and thoroughly annotated results, as well as cross-references to existing drug screening data. Our case studies demonstrate that shinyDeepDR produces meaningful predictions and unveils promising targets for an undruggable gene mutation in HCC, aligning consistently with findings from *in vivo* models. We anticipate that the tool will further enhance with the introduction of newly developed deep learning architectures[10] or expansive compound libraries, such as the Profiling Relative Inhibition Simultaneously in Mixtures (PRISM).[39] Although shinyDeepDR offers a useful platform for predicting anti-cancer drug responses using deep learning, it is crucial to recognize its limitations. The performance of the tool is intrinsically linked to the quality of the input datasets. The model was trained and tested using large-scale genomics data from TCGA and CCLE, which adhere to standardized protocols and rigorous quality standards.[1,2] Suboptimal data quality can potentially undermine the prediction accuracy. The present version of DeepDR and shinyDeepDR focuses on gene expression and mutation data and might not encompass all factors affecting drug responses. Continued research is warranted to create a comprehensive approach that integrates key biological, environmental, clinical, and other omics factors to enhance the predictions. Currently, shinyDeepDR is intended solely for research purposes, as explicitly stated on the website. While our tool prioritizes data privacy by not storing user-uploaded datasets, the accountability for ethical data management falls upon the end users. The prediction results should not be used for clinical interpretation or decision-making.

EXPERIMENTAL PROCEDURES

Resources availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yu-Chiao Chiu (yuc250@pitt.edu).

Materials availability

This study did not generate any new materials.

Data and code availability

shinyDeepDR is freely accessible at https://shiny.crc.pitt.edu/shinydeepdr/ without a login requirement. Results are displayed directly on the website, and users do not need to provide an e-mail address or contact information. All codes and data of shinyDeepDR are deposited in Figshare (https://doi.org/10.6084/m9.figshare.23925255).[40] Python codes and results of the DeepDR model can be downloaded from our publication[5] and GitHub repository (https://github.com/chenlabgccri/DeepDR). All cell line and tumor genomics, drug screening data, and drug annotation data are publicly available at the sources described in the implementation section and in cited references.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2023.100894.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R00CA248944 to Y.-C.C.; 3R00CA248944-04S1 from the Office of the Director to Y.-C.C.; R03OD036494 to Y.-C.C.; UL1RR025767 and P30CA054174 to Y.C.; U01CA279618 to Y.H.; R01CA250227 and R01CA251155 to S.P.M.; P30CA047904 to the University of Pittsburgh Medical Center Hillman Cancer Center and Y.-C.C.; P30DK120531 to the Genomics and Systems Biology Core of the Pittsburgh Liver Research Center, S.P.M., and Y.-C.C.; and S10OD028483 to the Center for Research Computing of the University of Pittsburgh); the Cancer Prevention and Research Institute of Texas (RP160732 and RP220662 to Y.C. and RP190346 to Y.C. and Y.H.); the Fund for Innovation in Cancer Informatics (Major Grant to Y.-C.C. and Y.C.); and the Leukemia Research Foundation (New Investigator Research Grant to Y.-C.C.). The funding sources had no role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript. We also thank Dr. Fangping Mu at the Center for Research Computing of the University of Pittsburgh for establishing the web server that hosts shinyDeepDR.

AUTHOR CONTRIBUTIONS

Conceptualization and tool design, L.-J.W., Y.C., and Y.-C.C.; tool implementation, L.-J.W., M.N., T.N., and M.J.K.; dataset collection, L.-J.W., T.N., and Y.-C.C.; data analysis, L.-J.W. and Y.-C.C.; data interpretation, L.-J.W., S.P.M., Y.H., Y.C., and Y.-C.C.; writing – original draft, L.-J.W., Y.C., and Y.-C.C.; writing – review & editing, all authors; funding acquisition, S.P.M., Y.H., Y.C., and Y.-C.C.; supervision, Y.C. and Y.-C.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: August 3, 2023
Revised: November 10, 2023
Accepted: November 14, 2023
Published: January 12, 2024

REFERENCES

1. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The

Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

2. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.
3. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740–754.
4. Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. *32*, 1202–1212.
5. Chiu, Y.C., Chen, H.I.H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.J., Huang, Y., and Chen, Y. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC Med. Genom. *12*, 18.
6. Manica, M., Oskooei, A., Born, J., Subramanian, V., Sáez-Rodríguez, J., and Rodríguez Martínez, M. (2019). Toward Explainable Anticancer Compound Sensitivity Prediction via Multimodal Attention-Based Convolutional Encoders. Mol. Pharm. *16*, 4797–4806.
7. Chang, Y., Park, H., Yang, H.J., Lee, S., Lee, K.Y., Kim, T.S., Jung, J., and Shin, J.M. (2018). Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. Sci. Rep. *8*, 8857.
8. Bazgir, O., Ghosh, S., and Pal, R. (2021). Investigation of REFINED CNN ensemble learning for anti-cancer drug sensitivity prediction. Bioinformatics *37*, i42–i50.
9. Kuenzi, B.M., Park, J., Fong, S.H., Sanchez, K.S., Lee, J., Kreisberg, J.F., Ma, J., and Ideker, T. (2020). Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. Cancer Cell *38*, 672–684.e6.
10. Chiu, Y.C., Chen, H.I.H., Gorthi, A., Mostavi, M., Zheng, S., Huang, Y., and Chen, Y. (2020). Deep learning of pharmacogenomics resources: moving towards precision oncology. Briefings Bioinf. *21*, 2066–2083.
11. Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., and Cheng, F. (2019). deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics *35*, 5191–5198.
12. Jia, P., Hu, R., and Zhao, Z. (2023). Benchmark of embedding-based methods for accurate and transferable prediction of drug response. Briefings Bioinf. *24*, bbad098.
13. Goecks, J., Jalili, V., Heiser, L.M., and Gray, J.W. (2020). How Machine Learning Will Transform Biomedicine. Cell *181*, 92–101.
14. Bhinder, B., Gilvary, C., Madhukar, N.S., and Elemento, O. (2021). Artificial Intelligence in Cancer Research and Precision Medicine. Cancer Discov. *11*, 900–915.
15. Yao, H., Liang, Q., Qian, X., Wang, J., Sham, P.C., and Li, M.J. (2020). Methods and resources to access mutation-dependent effects on cancer drug treatment. Briefings Bioinf. *21*, 1886–1903.
16. Baptista, D., Ferreira, P.G., and Rocha, M. (2021). Deep learning for drug response prediction in cancer. Briefings Bioinf. *22*, 360–379.
17. Cadow, J., Born, J., Manica, M., Oskooei, A., and Rodríguez Martínez, M. (2020). PaccMann: a web service for interpretable anticancer compound sensitivity prediction. Nucleic Acids Res. *48*, W502–W508.
18. Mateo, J., Steuten, L., Aftimos, P., André, F., Davies, M., Garralda, E., Geissler, J., Husereau, D., Martinez-Lopez, I., Normanno, N., et al. (2022). Delivering precision oncology to patients with cancer. Nat. Med. *28*, 658–665.
19. Cancer Cell Line Encyclopedia Consortium; Genomics of Drug Sensitivity in Cancer Consortium (2015). Pharmacogenomic agreement between two cancer cell line data sets. Nature *528*, 84–87.
20. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.
21. Newton, Y., Novak, A.M., Swatloski, T., McColl, D.C., Chopra, S., Graim, K., Weinstein, A.S., Baertsch, R., Salama, S.R., Ellrott, K., et al. (2017). TumorMap: Exploring the Molecular Similarities of Cancer Samples in an Interactive Portal. Cancer Res. *77*, e111–e114.
22. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Farrell, C.M., Feldgarden, M., Fine, A.M., Funk, K., et al. (2023). Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Res. *51*, D29–D38.
23. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.
24. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.
25. Pachter, L. (2011). Models for transcript quantification from RNA-Seq. Preprint at arXiv. https://doi.org/10.48550/arXiv.1104.3889.
26. Cronin, K.A., Scott, S., Firth, A.U., Sung, H., Henley, S.J., Sherman, R.L., Siegel, R.L., Anderson, R.N., Kohler, B.A., Benard, V.B., et al. (2022). Annual report to the nation on the status of cancer, part 1: National cancer statistics. Cancer *128*, 4251–4284.
27. Chen, B., Garmire, L., Calvisi, D.F., Chua, M.S., Kelley, R.K., and Chen, X. (2020). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. Nat. Rev. Gastroenterol. Hepatol. *17*, 238–251.
28. Russell, J.O., and Monga, S.P. (2018). Wnt/beta-Catenin Signaling in Liver Development, Homeostasis, and Pathobiology. Annu. Rev. Pathol. *13*, 351–378.
29. Llovet, J.M., Pinyol, R., Kelley, R.K., El-Khoueiry, A., Reeves, H.L., Wang, X.W., Gores, G.J., and Villanueva, A. (2022). Molecular pathogenesis and systemic therapies for hepatocellular carcinoma. Nat. Can. (Ott.) *3*, 386–401.
30. Li, S., Topatana, W., Juengpanich, S., Cao, J., Hu, J., Zhang, B., Ma, D., Cai, X., and Chen, M. (2020). Development of synthetic lethality in cancer: molecular and cellular classification. Signal Transduct. Targeted Ther. *5*, 241.
31. Setton, J., Zinda, M., Riaz, N., Durocher, D., Zimmermann, M., Koehler, M., Reis-Filho, J.S., and Powell, S.N. (2021). Synthetic Lethality in Cancer Therapeutics: The Next Generation. Cancer Discov. *11*, 1626–1635.
32. Regad, T. (2015). Targeting RTK Signaling Pathways in Cancer. Cancers *7*, 1758–1784.
33. Krejci, P., Aklian, A., Kaucka, M., Sevcikova, E., Prochazkova, J., Masek, J.K., Mikolka, P., Pospisilova, T., Spoustova, T., Weis, M., et al. (2012). Receptor tyrosine kinases activate canonical WNT/beta-catenin signaling via MAP kinase/LRP6 pathway and direct beta-catenin phosphorylation. PLoS One *7*, e35826.
34. Guo, Y., Zhao, Y.R., Liu, H., Xin, Y., Yu, J.Z., Zang, Y.J., and Xu, Q.G. (2021). EHMT2 promotes the pathogenesis of hepatocellular carcinoma by epigenetically silencing APC expression. Cell Biosci. *11*, 152.
35. Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell *154*, 1151–1161.
36. Cancer Genome Atlas Research Network Electronic address wheeler@bcmedu; Cancer Genome Atlas Research Network (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell *169*, 1327–1341.e23.
37. Nault, J.C., De Reyniès, A., Villanueva, A., Calderaro, J., Rebouissou, S., Couchy, G., Decaens, T., Franco, D., Imbeaud, S., Rousseau, F.,

et al. (2013). A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology *145*, 176–187.

38. Adebayo Michael, A.O., Ko, S., Tao, J., Moghe, A., Yang, H., Xu, M., Russell, J.O., Pradhan-Sundd, T., Liu, S., Singh, S., et al. (2019). Inhibiting Glutamine-Dependent mTORC1 Activation Ameliorates Liver Cancers Driven by beta-Catenin Mutations. Cell Metabol. *29*, 1135–1150.e6.

39. Corsello, S.M., Nagari, R.T., Spangler, R.D., Rossen, J., Kocak, M., Bryan, J.G., Humeidi, R., Peck, D., Wu, X., Tang, A.A., et al. (2020). Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. Nat. Can. (Ott.) *1*, 235–248.

40. Wang, L.-J., Ning, M., Nayak, T., Kasper, M.J., Monga, S.P., Huang, Y., Chen, Y., and Chiu, Y.-C. (2023). shinyDeepDR: A User-Friendly R Shiny App for Predicting Anti-cancer Drug Response Using Deep Learning. Figshare. https://doi.org/10.6084/m9.figshare.23925255.