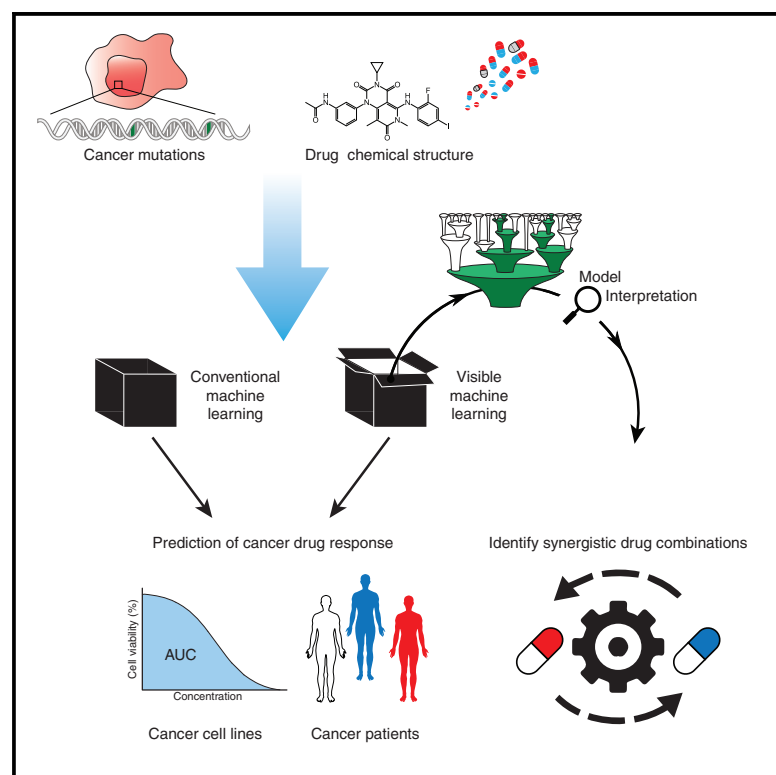


Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells

Graphical Abstract



Authors

Brent M. Kuenzi, Jisoo Park,
Samson H. Fong, ...,
Jason F. Kreisberg, Jianzhu Ma,
Trey Ideker

Correspondence

tideker@ucsd.edu

In Brief

Kuenzi et al. develop DrugCell, an interpretable deep learning model that simulates the response of human cancer cells to therapy. DrugCell predictions might generalize to patient tumors and can be used to design synergistic drug combinations that significantly improve treatment outcomes.

Highlights

- Development of an interpretable deep learning model of human cancer cells
- Model interpretations represent synergistic drug combination opportunities
- Predicted combinations improve progression-free survival in PDX models
- Response predictions stratify ER-positive breast cancer patient clinical outcomes



Article

Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells

Brent M. Kuenzi,^{1,5} Jisoo Park,^{1,5} Samson H. Fong,^{1,2} Kyle S. Sanchez,¹ John Lee,¹ Jason F. Kreisberg,¹ Jianzhu Ma,⁴ and Trey Ideker^{1,2,3,6,*}

¹Division of Genetics, Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA

²Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

³Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA

⁴Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

⁵These authors contributed equally

⁶Lead Contact

*Correspondence: tideker@ucsd.edu

<https://doi.org/10.1016/j.ccell.2020.09.014>

SUMMARY

Most drugs entering clinical trials fail, often related to an incomplete understanding of the mechanisms governing drug response. Machine learning techniques hold immense promise for better drug response predictions, but most have not reached clinical practice due to their lack of interpretability and their focus on monotherapies. We address these challenges by developing DrugCell, an interpretable deep learning model of human cancer cells trained on the responses of 1,235 tumor cell lines to 684 drugs. Tumor genotypes induce states in cellular subsystems that are integrated with drug structure to predict response to therapy and, simultaneously, learn biological mechanisms underlying the drug response. DrugCell predictions are accurate in cell lines and also stratify clinical outcomes. Analysis of DrugCell mechanisms leads directly to the design of synergistic drug combinations, which we validate systematically by combinatorial CRISPR, drug-drug screening *in vitro*, and patient-derived xenografts. DrugCell provides a blueprint for constructing interpretable models for predictive medicine.

INTRODUCTION

Each year dozens of new therapies enter clinical trials for the potential treatment of various types of cancer, but fewer than 4% will ultimately gain approval by the US Food and Drug Administration (Wong et al., 2019). Although many factors contribute to this challenge, a major failure is in understanding how or why a particular cancer responds to therapy. The problem becomes particularly acute for cancers that are not associated with strong targetable genetic drivers (e.g., *BCR-ABL* fusion, *EGFR* mutation, or *EML4-ALK* translocation), since cancers without these known drivers lack clear biomarkers with which to stratify drug response. A better basic understanding of the molecular pathways governing drug sensitivity would help greatly in determining which patients should be treated and with which drugs.

There has recently been a great deal of interest in applying advances in artificial intelligence, including machine learning and deep learning, to classic problems in biomedicine (Topol, 2019). Whereas popular applications include disease diagnosis from biomedical images and interpretation of electronic medical records (Esteva et al., 2019; Rajkomar et al., 2019; Wainberg et al., 2018), machine learning models are also of high interest in predicting drug responses (Barretina et al., 2012; Costello et al., 2014; Garnett et al., 2012; Iorio et al., 2016; Zeng et al.,

2019). In a typical application (reviewed in Table S1), the model uses the 'omics profile of a cell line or tissue sample as input to predict the 50% inhibitory concentration (IC_{50}) of a drug. For example, Iorio et al. (2016) built elastic net models to predict the drug IC_{50} of cancer cell lines given their profiles of gene mutations and expression levels; a range of predictive accuracy is observed, depending on the compound. Using the same dataset, Cortés-Ciriano et al. (2016) showed that predictive performance could in some cases be improved using a random forest model linked to a measure of statistical confidence in each prediction. Deep neural networks (Baptista et al., 2020; Chiu et al., 2019; Menden et al., 2013; Sakellaropoulos et al., 2019) and variational autoencoders (Rampásek et al., 2019) have also been applied to drug response prediction, with significant performance gains noted depending on the drug and disease context.

Owing to the significant molecular heterogeneity observed across tumors, there are often many different molecular features and feature combinations that can lead a model to predict a particular drug response. What these features are, and whether they are distinct or functionally interrelated, can be very difficult to interpret, however. The reason is that most machine learning models are “black boxes,” optimized for prediction accuracy without knowledge of or attention to the biological mechanisms underlying predicted outcomes (Ching et al., 2018). To address these difficulties,



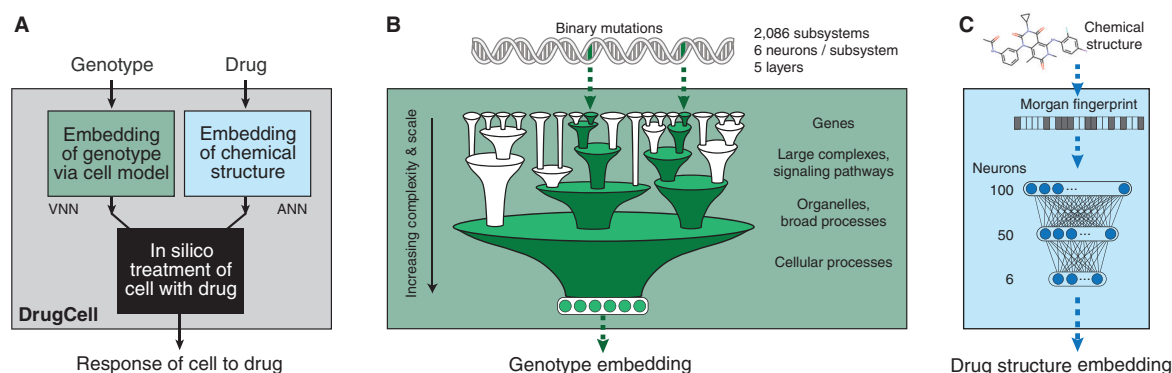


Figure 1. DrugCell Design

(A) DrugCell uses a modular neural network design that combines conventional artificial neural networks (ANN) with a visible neural network (VNN) to make drug response predictions.

(B) Binary encodings of individual genotypes are processed through a VNN with architecture guided by a hierarchy of cell subsystems, with multiple neurons assigned per subsystem.

(C) Compound chemical structures are processed through an ANN using the Morgan fingerprint as input features.

model interpretation is now a rapidly growing subfield within machine learning, with a growing arsenal of approaches for achieving models with not only high predictive accuracy, but also high descriptive accuracy (Murdoch et al., 2019). One major strategy has been to use prior knowledge or data to add structure to the model, which can then be interpreted. Applied to genomics, such a strategy has been used to recast the thousands of measured molecular features of a tumor as states on a much smaller number of functional modules (Cortés-Ciriano et al., 2016; Yang et al., 2019). For example, a recent study mapped raw molecular measurements to a set of pre-defined metabolic pathways drawn from prior knowledge bases; the states of these pathways predict antibiotic resistance in *Escherichia coli*, with particular pathway features emerging as candidate mechanisms of resistance (Yang et al., 2019). Organization of molecular features into predictive modules can also be accomplished using prior data as opposed to literature-curated knowledge. Such an approach was recently exemplified by Deep-Profile, which analyzed a large collection of leukemia expression profiles to extract a low-dimensional representation of these data as a set of functional gene modules; these modules are then used as interpretable features for drug response prediction (Dincer et al., 2018). Apart from model-based approaches, a second major strategy to increase model interpretability has been to perform *post hoc* analysis of model features or feature weights to interpret the underlying drug response mechanisms (Chiu et al., 2019; Iorio et al., 2016; Murdoch et al., 2019). For example, the weights assigned to each input gene by a black-box neural network model are subjected to gene set enrichment analysis (Subramanian et al., 2005) to identify pathways regulating the predicted drug response (Sakellaropoulos et al., 2019). These pathways, however, were not used during modeling or validated experimentally.

To more explicitly link the structure of a machine learning model to cellular functions, we recently developed a visible neural network (VNN) simulating a simple eukaryotic cell, *Saccharomyces cerevisiae* (Ma et al., 2018; Yu et al., 2018). This model, called DCell, was made mechanistically interpretable, or “visible,” by directly mapping the neurons of a deep neural network into a large hierarchy of known and putative molecular components and pathways. DCell is able to accurately predict

the impact of genetic mutations on cellular growth response and, simultaneously, identify the most relevant molecular pathways driving those predictions. Building from this paradigm, we now describe DrugCell, a VNN that simulates the response of human cancer cells to therapeutic chemical compounds. DrugCell couples the inner workings of the model to the hierarchical structure of human cell biology, allowing for response predictions for any drug in any cancer and intelligent design of effective combination therapies.

RESULTS

Design and Training of an Interpretable Neural Network of Drug Response

The cellular drug response is a complex phenomenon that depends on both biological and chemical factors (Turner et al., 2015). Current black-box models of drug response that use both these factors have begun to reach the limits of predictive performance (Table S1). We therefore aimed to design a model that maintains this high level of predictive capability while gaining mechanistic interpretability of the model predictions. To capture both determinants of drug response in an interpretable model, we devised DrugCell as a neural network with two branches (Figure 1A, STAR Methods). The first branch was a VNN modeling the hierarchical organization of molecular subsystems in a human cell, drawn from 2,086 biological processes documented in the Gene Ontology (GO) database (Ashburner et al., 2000) (Figure S1A). Each of these subsystems, from those involving small protein complexes (e.g., β -catenin destruction complex) to larger signaling pathways (e.g., MAPK signaling pathway) to overarching cellular functions (e.g., glycolysis), was assigned a bank of artificial neurons to represent the state of that subsystem (Figure 1B). Connectivity of neurons was set to mirror the biological hierarchy, so that neurons accept inputs only from child subsystems and send outputs only to parent systems, with connection weights determined during training. The use of multiple neurons per subsystem (here six, see STAR Methods) allowed cellular subsystems to be multifunctional, with distinct states able to adopt a range of values along multiple dimensions

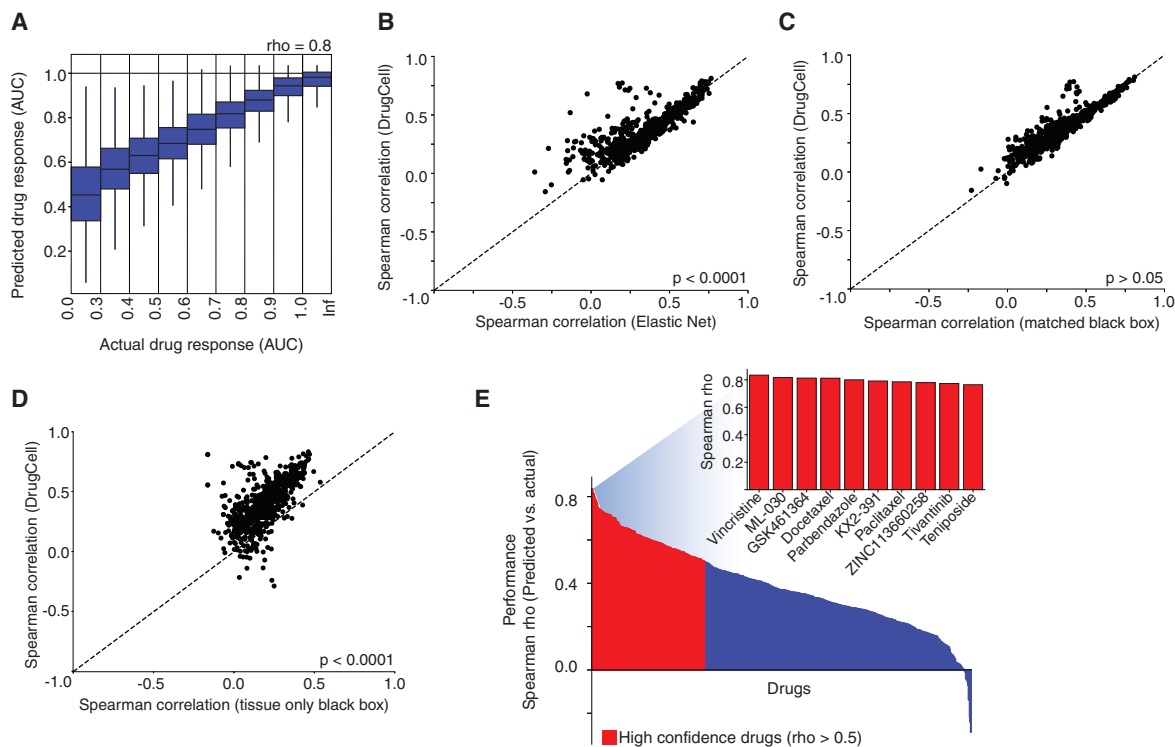


Figure 2. Predictive Performance

(A) Predicted versus actual drug responses across all (cell line, drug) pairs studied. Box plots show the 25th, 50th, and 75th percentiles of values in each bin; whiskers show maximum and minimum values.

(B–D) Scatterplots of the predictive performance (Spearman rho between actual and predicted drug response across 684 drugs) of DrugCell versus three alternative models: (B) elastic net, (C) matched black-box neural network, and (D) tissue-only black-box neural network. Points represent individual drugs; points above the diagonal represent drugs better predicted by DrugCell.

(E) Waterfall plot of predictive performance for each drug in the dataset (y axis), ranked from highest to lowest (x axis). “High confidence” drugs are highlighted in red ($\rho > 0.5$). The inset shows the performance for the top 10 best predicted drugs.

(Copley, 2012). The input layer of the hierarchy mapped to the mutation status of genes. The six neurons at the VNN output, corresponding to the root of the hierarchy, represented the embedded state of the whole cell based on its genotype (Figure 1B). In total, the VNN used 12,516 neurons distributed hierarchically across six distinct layers (STAR Methods, Figure S1B). The second branch of DrugCell was a conventional artificial neural network (ANN) embedding the Morgan fingerprint of a drug, a canonical vector representation of chemical structure (Figure 1C, STAR Methods) (Rogers and Hahn, 2010). Outputs from the two branches of the model, the VNN embedding cell genotype and the ANN embedding drug structure, were combined in a single layer of neurons, which were then integrated to generate the response of a given genotype to a particular treatment (Figure 1A).

To train the model, we harmonized data from two large cancer drug screening resources: the Cancer Therapeutics Response Portal (CTRP) v2 and the Genomics of Drug Sensitivity in Cancer (GDSC) database (Seashore-Ludlow et al., 2015; Yang et al., 2013). The combined dataset consisted of 509,294 cell line-drug pairs, covering 684 drugs and 1,235 cell lines (Figure S1C, STAR Methods). All major tissue types were represented, with hematopoietic and lung lineages the most prevalent (Figure S1D). Each cell-line genotype was represented by a binary vector

recording the mutational status (1 = mutated, 0 = non-mutated) of the top 15% most frequently mutated genes in cancer ($n = 3,008$; median mutated genes per cell line = 73; Figure S1E). Each drug’s chemical structure was represented by an average of 81 activated bits in the Morgan fingerprint vector, with each bit typically representing fewer than 10 molecular fragments (Figures S1F and S1G). DrugCell was trained to associate each genotype-drug pair with its corresponding drug response, measured by the area under the dose-response curve (AUC, STAR Methods). The DrugCell model and its codebase are available for public download on GitHub (<https://github.com/idekerlab/DrugCell>).

Interpretable Modeling of Drug Response Has No Performance Loss

We first sought to assess the prediction accuracy of DrugCell using the Spearman correlation (ρ) between predicted and observed AUC values in 5-fold cross validation (STAR Methods). The total accuracy over all cell line-drug pairs was $\rho = 0.80$ (Figure 2A). Further insight was achieved by computing the prediction accuracy for each drug individually, revealing a subpopulation of drugs with very high prediction accuracy (30% of drugs with $\rho > 0.5$) amid a much wider general distribution (range -0.29 to $+0.83$, median 0.37). These accuracies were

significantly higher than those achieved for elastic net (median $\rho = 0.35$), a state-of-the-art regression technique used in many previous approaches to drug response prediction (Eskio-cak et al., 2017; Iorio et al., 2016; Kuenzi et al., 2019; Potts et al., 2015) (Figure 2B). DrugCell's drug-by-drug predictive performance was not significantly different from that of a conventional black-box ANN with matching numbers of neurons, layers, and connections (Figure 2C). It was also comparable to previous efforts to incorporate chemical features of drugs into the response prediction (e.g., structure and physiochemical properties such as solubility, lipophilicity, and molecular weight), and it outperformed models that predict response using biological features alone (e.g., expression of biomarkers, point mutation, copy number variation, and microsatellites; Table S1). Finally, since knowledge of tissue type can be predictive of drug response even in the absence of other information (Iorio et al., 2016), we considered that some of the performance of these models might be due to their ability to recognize the tissue type of a cell line from its input data (i.e., its mutational profile). Accordingly, we compared DrugCell with an equivalent neural network model trained on drug structure and a tissue label only (STAR Methods). DrugCell vastly outperformed this tissue-only model (median $\rho = 0.18$; Figure 2D), indicating that the model had learned information from somatic mutations beyond the tissue of origin.

Compounds for which DrugCell predictions were most accurate came from diverse target classes, including chemotherapeutics (e.g., vincristine, teniposide) and targeted therapies (e.g., GSK461364 targeting PLK1, KX2-391 targeting Src; Figure 2E). DrugCell maintained the specificity of the training data in that its predictions were specific to individual classes of drugs (e.g., MEK inhibitor predictions were highly specific) and did not simply reflect general drug toxicity (Figure S2A). Predictive performance for a drug did not strongly correlate with the number of cell line-drug pairs used for training, nor with the structural complexity of a compound (number of activated bits; Figures S2B and S2C). We did find that compounds eliciting a larger range of cell-line responses tended to be more predictable (Figure S2D). Similarly, individual cell lines (Figure S2E) and tissue types (Figure S2F), which elicit a large range of responses, were in general highly predictable.

DrugCell Learns Mechanisms that Mediate Specific Drug Responses

Having evaluated predictive ability, we next turned to mechanistic interpretation. This task was aided by the two model branches, which dissect the effects of genotype on the configuration of cell systems (genotype embedding) from the effects of chemical structure on drug activity within the cell (drug embedding, Figure 1A). We visually inspected these embeddings by plotting the top two principal components (Figures 3A–3E). The genotype embedding from the VNN revealed a separation of genotypes according to mutations known to confer specific drug sensitivities, such as activating mutations in BRAF (Figure 3A) that promote sensitivity to the MEK inhibitor selumetinib (Figure 3B). The genotype embedding also distinguished mutations leading to drug resistance, such as mutations in EGFR (Yin et al., 2019), LKB1 (Shimamura et al., 2013), or BRAF (Ma et al., 2017) (Figure 3C) that confer resistance to the BET-family inhibitor JQ-1 (Figure 3D). We similarly inspected the DrugCell embeddings of

individual subsystems within the VNN and found that many were in agreement with subsystem activities measured experimentally by an independent analysis of protein abundances and phosphorylation states using reverse-phase protein arrays (RPPAs; Figure S3A; STAR Methods). For example, DrugCell accurately captured MAPK pathway activity within the subsystem embedding of *Regulation of MAPK cascade* (Figure S3B), which significantly correlated with ERK1/2 phosphorylation (Figure S3C). Overall, the majority of DrugCell subsystems were well correlated with the RPPA measurements of those subsystems (note bimodal distribution of correlation in Figure S3A). Other accurately captured subsystems included *Proteolysis* (Figure S3D), *Regulation of PI3K signaling* (Figure S3E), and *Cell-cycle arrest* (Figure S3F).

Inspection of the drug embedding from the ANN revealed a stratification of drugs based on their mechanisms of action within major drug target classes (Figure 3E). The distance between each pair of drugs in the chemical structure embedding did not correlate with their overall chemical similarity (Figure S4A), consistent with previous studies of drug activity and chemical structure (Breinig et al., 2015). Since the training data consisted solely of drugs and drug-like molecules, the chemical structural embedding did not stratify drugs on chemical features such as membrane permeability (Figure S4B), solubility (Figure S4C), or pharmacodynamic properties (Lipinski; Figure S4D). Together these results suggest that DrugCell is able to learn key features of the genotype that govern drug sensitivity and resistance, as well as features of chemical structure that govern drug biological activity.

Since DrugCell's VNN is structured according to the hierarchy of biological subsystems comprising a human cell, its output (genotype embedding) is the result of state changes in particular subsystems within that hierarchy. To identify the most important of these subsystems, we scored subsystems by the degree to which their states were significantly more predictive of a drug response than the states of their child subsystems using the relative local improvement in predictive power metric (RLIPP, STAR Methods) (Ma et al., 2018). As an initial proof of concept, we used RLIPP scoring to identify subsystems important for the cellular response to taxol (paclitaxel), an agent that stabilizes microtubules (Figures 2E and 3F, Table S2). Among the top scores for paclitaxel, many subsystems were metabolic processes (hypergeometric $p < 0.05$; Figures 3G and 3H), including *Response to cAMP* (top score) along with *Insulin secretion in response to glucose* and *Response to glucose*. We confirmed by inspection that the states of these subsystems had the ability to stratify paclitaxel sensitive versus resistant cell lines (e.g., *Response to cAMP* subsystem, Figure 3I). Given these underlying metabolic pathways, we hypothesized that paclitaxel efficacy might be modulated by metabolic perturbation. We therefore exposed A427 cells to three different treatments – paclitaxel, the glycolysis inhibitor 2-deoxyglucose (2-DG), or a combination of the two – and found that the combination was substantially more effective than either individual compound (Figure 3J).

A similar analysis was performed for the next (second-most) important subsystem, *Regulation of ubiquitin-protein transferase activity* (Figure 3G, Table S2). We combined paclitaxel with perturbation of ubiquitin-dependent protein degradation via the proteasome inhibitor bortezomib (Figure S5A). We found

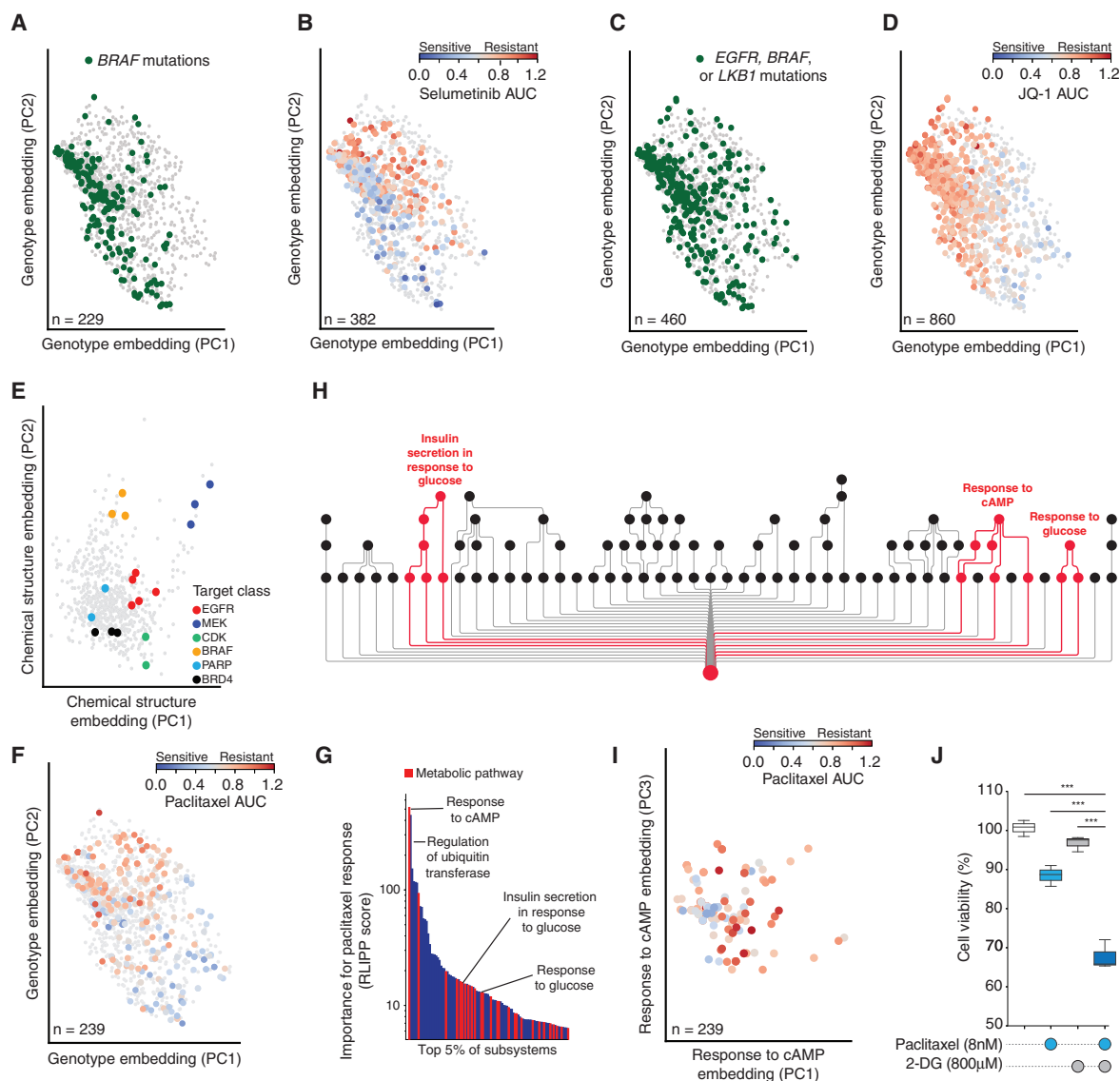


Figure 3. Characterization of Cancer Cell States Learned by DrugCell

(A–D) Genotype embeddings of each cell line, showing the first two principal components (PC). Points are cell lines, with colors indicating specific drug responses or genetic markers according to the panel. (A and C) Green denotes cell lines harboring mutations in *BRAF* or in *EGFR*, *BRAF*, or *LKB1*, respectively. Gray denotes cell lines without mutations in these genes. (B and D) Blue-to-red gradient represents the response to selumetinib or JQ-1, respectively. Gray denotes cell lines not tested against that drug.

(E) Drug structure embedding. Points are drugs, with colors indicating drug target classes.

(F) Genotype embeddings of each cell line as in (A–D), but with blue-to-red gradient representing response to paclitaxel.

(G) Waterfall plot of top 5% of subsystems (x axis) important for paclitaxel response by RLIPP score (y axis). Subsystems capturing metabolic pathways are highlighted in red.

(H) Visualization of select subsystems highlighted in (G), comprising a sub-hierarchy of the full DrugCell model. Red is used to trace the branches of the hierarchy related specifically to regulation of glycolysis.

(I) Response to cAMP subsystem embedding. Points are cell lines, blue-to-red gradient represents response to paclitaxel.

(J) Boxplot of the relative cell viability of treatment with DMSO, paclitaxel, 2-deoxyglucose (2-DG), or the combination at the indicated concentrations in A427 cells. Data are representative of drug treatments performed in biological and technical triplicates. The boxes represent the interquartile range (IQR) bisected by the median, whiskers represent the maximum and minimum range of the data that do not exceed 1.5 times the IQR. *** $p < 0.0001$ from a t test.

that these treatments were antagonistic, consistent with recent findings showing that glycolysis is subject to negative physical regulation by ubiquitin ligases at the cytoskeleton (Park et al., 2020). Ubiquitin and subsystems were also identified for docetaxel, a sister compound (Table S3). Notably, these DrugCell

pathways were not identified by earlier analyses of genetic mutations (Table S4) and were distinct from those identified by differential mRNA expression analysis of paclitaxel sensitive versus resistant lines (Figures S5B and S5C). Unlike the glycolytic perturbations emerging from DrugCell analysis,

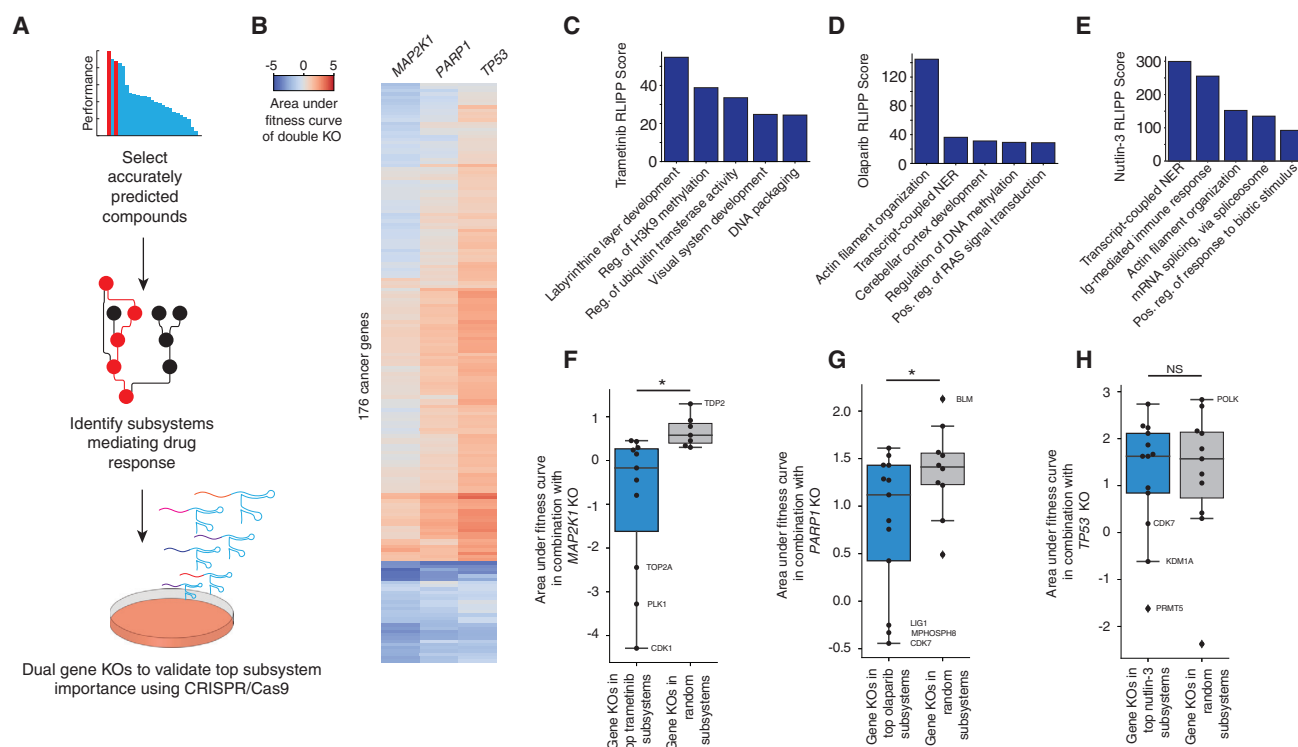


Figure 4. Systematic Validation of Identified Mechanisms of Sensitivity Using CRISPR/Cas9

(A) Workflow of systematic analysis using CRISPR/Cas9.

(B) Heatmap of the area under the fitness curves for 176 cancer genes in combination with *MAP2K1*, *PARP1*, and *TP53*.

(C–E) Bar plots of the RLIPP scores of the top five subsystems for (C) trametinib, (D) olaparib, and (E) nutlin-3.

(F–H) Boxplots of the area under the fitness curve following CRISPR/Cas9-mediated knockout of (F) *MAP2K1*, (G) *PARP1*, and (H) *TP53* in combination with highly weighted genes within the top five subsystems identified by DrugCell for each parent drug compared with random. Select genes are labeled. The boxes represent the IQR bisected by the median, and whiskers represent the maximum and minimum range of the data that do not exceed 1.5 times the IQR. * $p < 0.05$ from a t test, NS denotes not significant.

combination treatments suggested by differentially expressed pathways were not successful at enhancing paclitaxel efficacy (Figure S5D).

Moving beyond paclitaxel to examine the important subsystems identified for other drugs, we found that some of these subsystems corresponded to previously identified mechanisms of drug sensitivity, while many others were novel pathways warranting further investigation. In particular, we examined 60 drugs for which pan-cancer diagnostic gene mutations had been reported by an earlier analysis of the GDSC dataset using type II error ANOVA modeling (Iorio et al., 2016). For a number of drugs, DrugCell recovered the previously reported diagnostic gene(s) within the top subsystem (4 drugs) or top 10 subsystems (14 drugs, upper 0.4th percentile of subsystems). For the vast majority, however (56 drugs), DrugCell achieved better predictive performance by consulting additional, or different, markers than had been previously reported (Table S4).

Given the extent of novel drug response pathways, we sought to systematically investigate the indicated mechanisms (Figure 4A; STAR Methods), focusing on trametinib, a MEK1 inhibitor; olaparib, a PARP1 inhibitor; and nutlin-3, an MDM2 antagonist that stabilizes and activates p53. CRISPR knockouts of each of the three drug targets (*MEK1*, *PARP1*, *TP53*) were combined with knockouts of each gene in a custom CRISPR/Cas9 library,

which had broad representation of cancer signaling pathways (MCF7 cells; Figure 4B). The top five important subsystems in the response to each drug were identified (RLIPP analysis; Figures 4C–4E), along with the genes in those subsystems covered by the CRISPR library. Combinatorial disruption of *MAPK1* with genes in trametinib subsystems (Figure 4C) resulted in significantly more cell killing than observed for genes from random unimportant subsystems (Figure 4F). A similar cell killing effect (Figure 4G) was observed for combinatorial disruption of *PARP1* with genes in olaparib subsystems (Figure 4D). In contrast, combinatorial disruption of *TP53* with genes in nutlin-3 subsystems (Figure 4E) had effects on cell growth that were not significantly different from random (Figure 4H). This result was expected, as *TP53* knockout has the opposite effect compared with nutlin-3, which leads to p53 activation. These results, together with the preliminary results from paclitaxel, provide systematic support for the importance of top response pathways identified by DrugCell.

Identified Subsystems Represent Synergistic Drug Combination Opportunities

The parallel pathway inhibition theory of drug synergy (Yeh et al., 2009) holds that two drugs will be synergistic if they inhibit separate pathways that regulate a common essential function

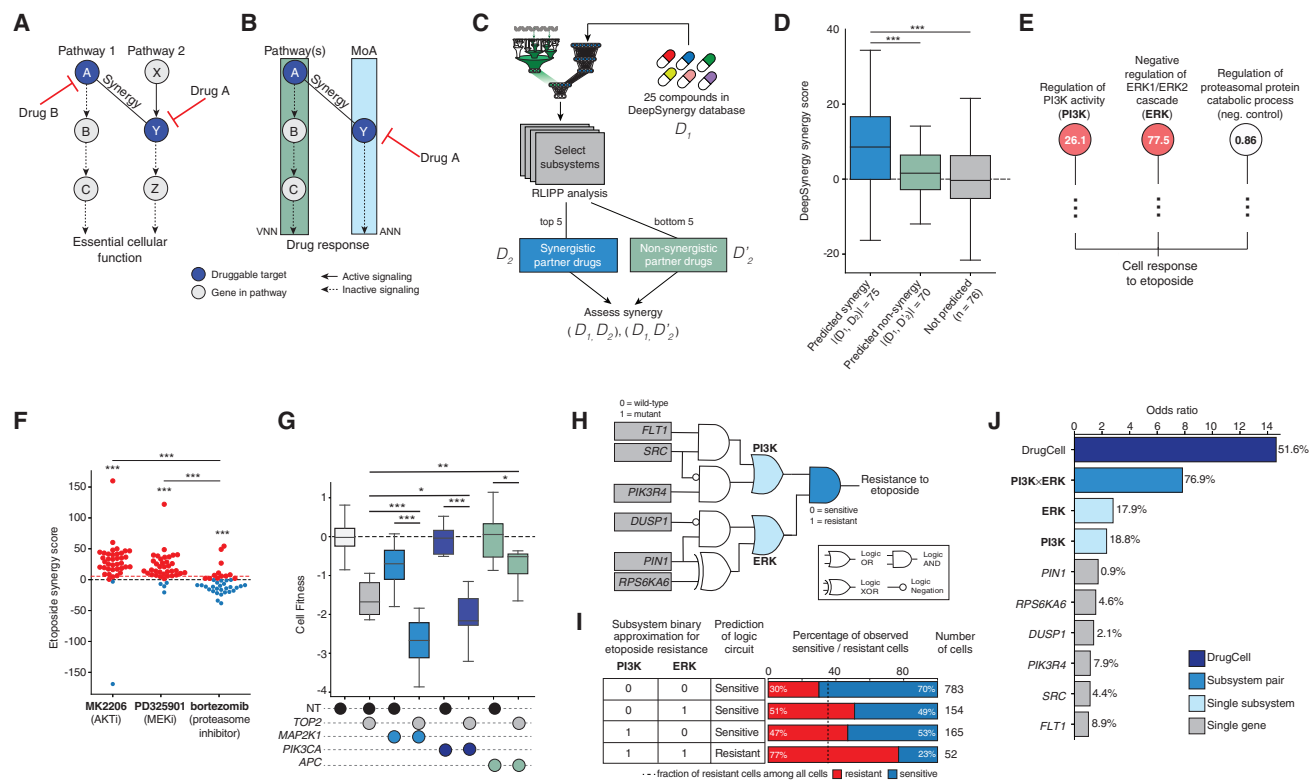


Figure 5. Discovery and Validation of Synergistic Mechanisms

(A) Parallel pathway theory of drug synergy, in which a pathway 2 is targeted by the mechanism of action (MoA) of drug A, and synergy is achieved by simultaneously targeting parallel pathway 1 with drug B.

(B) Logic learned by DrugCell for drug A, in which pathway 1 arises as a predicted mechanism of the VNN.

(C) Workflow demonstrating systematic design and assessment of pairwise combinations of drugs.

(D) Boxplots of DeepSynergy synergy scores for predicted drug combinations, predicted non-synergistic combinations, and random combinations. The boxes represent the IQR bisected by the median, and whiskers represent the maximum and minimum range of the data that do not exceed 1.5 times the IQR. ***p < 0.0001.

(E) Representative subsystems used by DrugCell to simulate etoposide sensitivity (red nodes), along with a negative control branch (white node). RLIPP scores are displayed inside each node. Subsystem names are abbreviated.

(F) Bee swarm plot of the Loewe synergy scores observed upon combination of etoposide with MK2206, PD325901, or bortezomib. Drug combinations were chosen based on subsystems identified in (E). Red dotted line indicates the mean of all Loewe synergy scores in the dataset (Figure S6). ***p < 0.0001. *** without bars represent t test against the synergy score distribution of the full dataset (Figure S6), or bortezomib negative control, as indicated. Red points are cell lines for which synergy is observed. Blue points are cell lines for which antagonism is observed.

(G) Boxplots of the relative cell growth of A549 cells following CRISPR/Cas9-mediated knockout of MAP2K1, PIK3CA, or APC (negative control) in combination with TOP2 or a non-targeting control (NT). Data are reflective of two independent transductions. ***p < 0.0001, *p < 0.01, **p < 0.001.

(H) Boolean logic circuit approximating how the mutational status of genes in the PI3K and ERK subsystems is translated to an etoposide response by DrugCell.

(I) Truth table showing translation of PI3K and ERK states to a binary drug response output. The percentage of observed sensitive versus resistant cells for each state is shown. Dotted line indicates baseline percentage of etoposide-resistant samples among all cell lines.

(J) Odds ratios of etoposide response prediction for DrugCell, the ERK and PI3K logic functions from (H), and individual genes from (H). Percentages of cell lines with an alteration to that biomarker are also shown. Odds ratios are against a background of cell lines that are wild type with respect to this circuit.

(Figure 5A). The branched architecture of the DrugCell model (Figure 1A) mirrors this parallel pathway structure, in that the biological activity of a drug is learned by the drug embedding branch, and the parallel pathways are learned by the genotype embedding branch (Figure 5B). Subsystems important for predicting a drug response may therefore represent synergistic drug combination opportunities. Exactly such parallelism was used to nominate the combination treatments in the above analysis (i.e., 2-DG as synergistic with paclitaxel).

To further explore this concept, we used RLIPP scores to rank subsystems regulating sensitivity to 25 drugs in the DeepSynergy database (Preuer et al., 2018), in which all pairs of 25 drugs had

been tested across a panel of 39 cell lines (Figure 5C). We then analyzed the top 5 and bottom 5 DrugCell subsystems for each of these compounds to nominate synergistic and non-synergistic drug combinations. We found that drug combinations nominated by DrugCell were strongly and significantly enriched for synergistic cell killing outcomes, in contrast to combinations predicted to be non-synergistic or random combinations (Figure 5D).

One such example was etoposide, a topoisomerase inhibitor that leads to DNA damage (Table S5). Among the top etoposide subsystems were the major kinase signaling pathways PI3K-AKT (Regulation of PI3K activity, PI3K; Figure 5E) and RAF-MEK-ERK (Negative regulation of ERK1/ERK2 cascade, ERK;

Figure 5E). Indeed, etoposide synergized strongly with AKT and MEK inhibition across the majority of cell lines tested in DeepSynergy (Figure 5F). We further validated the observed synergy by deleting the target of etoposide, *TOP2*, using CRISPR/Cas9 gene editing in A549 cells, either alone or in combination with core genes in PI3K-AKT signaling (*PIK3CA*) or RAF-MEK-ERK signaling (*MAP2K1*). We observed that deletion of *TOP2* with either *PIK3CA* or *MAP2K1* demonstrated significant loss of cell viability compared with single-gene knockout (Figure 5G). APC, whose subsystem (β -catenin destruction complex) was not identified by RLIPP (Table S5), did not show this same pattern (Figure 5G). Similarly, etoposide did not synergize with the proteasome inhibitor bortezomib (Figure 5F), consistent with the proteasome subsystem not being identified by DrugCell (Figure 5E).

Further inspection suggested that the relationship between PI3K signaling, ERK signaling, and etoposide sensitivity captured by DrugCell could be roughly approximated by a logic function integrating the mutational status of six genes (Figures 5H and 5I; STAR Methods). Among these, *FLT1* (Das et al., 2005) and *PIN1* (Mathur et al., 2011) had previously been shown to regulate etoposide response, whereas *DUSP1*, *PIK3R4*, *SRC*, and *RPS6KA6* had not. Considered individually, any one of these genes was mutated rarely in cancer cell lines, with limited power to predict etoposide sensitivity versus resistance (mutation frequencies 0.9%–8.9%; odds ratios <2; Figure 5J). Considered as an integrated circuit, however, these gene mutations converge on **PI3K** or **ERK** subsystems to create a powerful network-based biomarker of drug response (odds ratio 7.8; Figure 5J). We also noted that these two pathways represent only a portion of the full DrugCell model, which predicts etoposide sensitivity with an odds ratio of 14.3.

DrugCell Improves Progression-Free Survival of Patient-Derived Xenograft Models

We next wished to move beyond cell lines to predict and interpret drug responses in the *in vivo* setting of patient-derived xenograft models (PDX; Figure 6A, STAR Methods). To do so, we accessed the PDX Encyclopedia (Gao et al., 2015), in which 399 PDX tumors of varying tissue types had been screened against a total of 40 different monotherapies and 27 combination therapies. The genotypes of each PDX had also been established (Gao et al., 2015), which were provided to DrugCell to make response predictions to each monotherapy. We considered a PDX tumor to be sensitive to a therapy (DrugCell (+)) if its predicted AUC was beneath the median predicted for all PDX-drug pairs; otherwise this tumor was labeled as insensitive (DrugCell (–)). DrugCell (+) tumors demonstrated significantly longer progression-free survival (PFS) than DrugCell (–) tumors (2.19 versus 1.58 months, $p = 9.4 \times 10^{-10}$, log rank test). However, given the overall insensitivity of these PDX tumors to monotherapy, corresponding to the short observed PFS observed for both DrugCell classes, we wished to evaluate how well DrugCell is able to suggest effective drug combinations. We used RLIPP scoring to rank subsystems by importance in mediating drug responses to six primary drugs, filtering this list to those that contained secondary drug targets. The observed PFS of each of these (primary, secondary) combinations was used to estimate the prediction sensitivity and specificity along a receiver oper-

ating characteristic curve (ROC; Figure 6A, STAR Methods). We found that DrugCell was able to accurately identify subsystems that correspond to effective drug combinations in PDX tumors (auROC = 0.75; Figure 6B) with relatively few false positives and negatives (Figure 6C).

For example, DrugCell analysis of BKM-120, a PI3K inhibitor, identified *Negative regulation of ERK1 + ERK2 cascade* as an important subsystem for BKM-120 response, suggesting a combination of PI3K + MAPK pathway inhibitors (BKM-120 + encorafenib). This combination significantly increased PFS across the PDX panel compared with monotherapy (Figure 6D). Similarly, DrugCell identified *DNA damage response, signal transduction by p53 class mediator resulting in cell-cycle arrest* as an important subsystem for abraxane response, suggesting combination chemotherapy with an agent inducing DNA damage and cell-cycle arrest (abraxane + gemcitabine). This combination similarly significantly improved PFS (Figure 6D). For the combinations that were not prioritized by DrugCell (not in top 20% of subsystems by RLIPP), these combinations indeed failed to significantly improve PFS (Figures 6C and 6E). These results suggested that DrugCell has utility in guiding design of combination therapies in patient tumors.

DrugCell Predicts the Response of Estrogen Receptor-Positive Metastatic Breast Cancer Patients to mTOR and CDK4/6 Inhibitors

Last, we sought to evaluate whether DrugCell could be used clinically to stratify cancer patients into responsive and non-responsive patient populations. We obtained and analyzed aggregated clinical trial data (Smyth et al., 2020) from 221 estrogen receptor (ER)-positive metastatic breast cancer patients who had undergone multiple rounds of therapy, including an ER antagonist (fulvestrant) in addition to treatment with an mTOR inhibitor (everolimus) or CDK4/6 inhibitor (ribociclib). For this analysis (STAR Methods), we predicted patient response to either mTOR or CDK4/6 inhibition using our pre-trained DrugCell model. We considered a patient to be DrugCell (+) if they were predicted to be sensitive to either therapy and DrugCell (–) if they were predicted to be insensitive to both therapies. DrugCell (+) patients had significantly longer overall survival than DrugCell (–) patients (48.2 versus 33.6 months, $p = 0.018$; Figure 7A).

We next interrogated the mechanisms underlying the differential sensitivity between DrugCell (+) and DrugCell (–) patients by performing an RLIPP analysis for both the mTOR and the CDK4/6 inhibitors. Notably, we found that both drug responses were modulated by ER-related subsystems (Figures 7B and 7C), consistent with their use in ER-positive breast cancer (Hare and Harvey, 2017; Pernas et al., 2018). We also found that the major mechanisms of action of both drugs were among the top pathways, with PI3K signaling being especially important for response to mTOR inhibitors and CDK activity being important for CDK4/6 inhibitor activity (Figures 7B and 7C). Interestingly, TOR signaling was also identified for CDK4/6 inhibitors (Figure 7B), and CDK activity was identified for mTOR inhibitors (Figure 7C), suggesting that these drugs could be an effective combination therapy, a finding supported by recent preclinical studies (Michaloglou et al., 2018; Occhipinti et al., 2020).

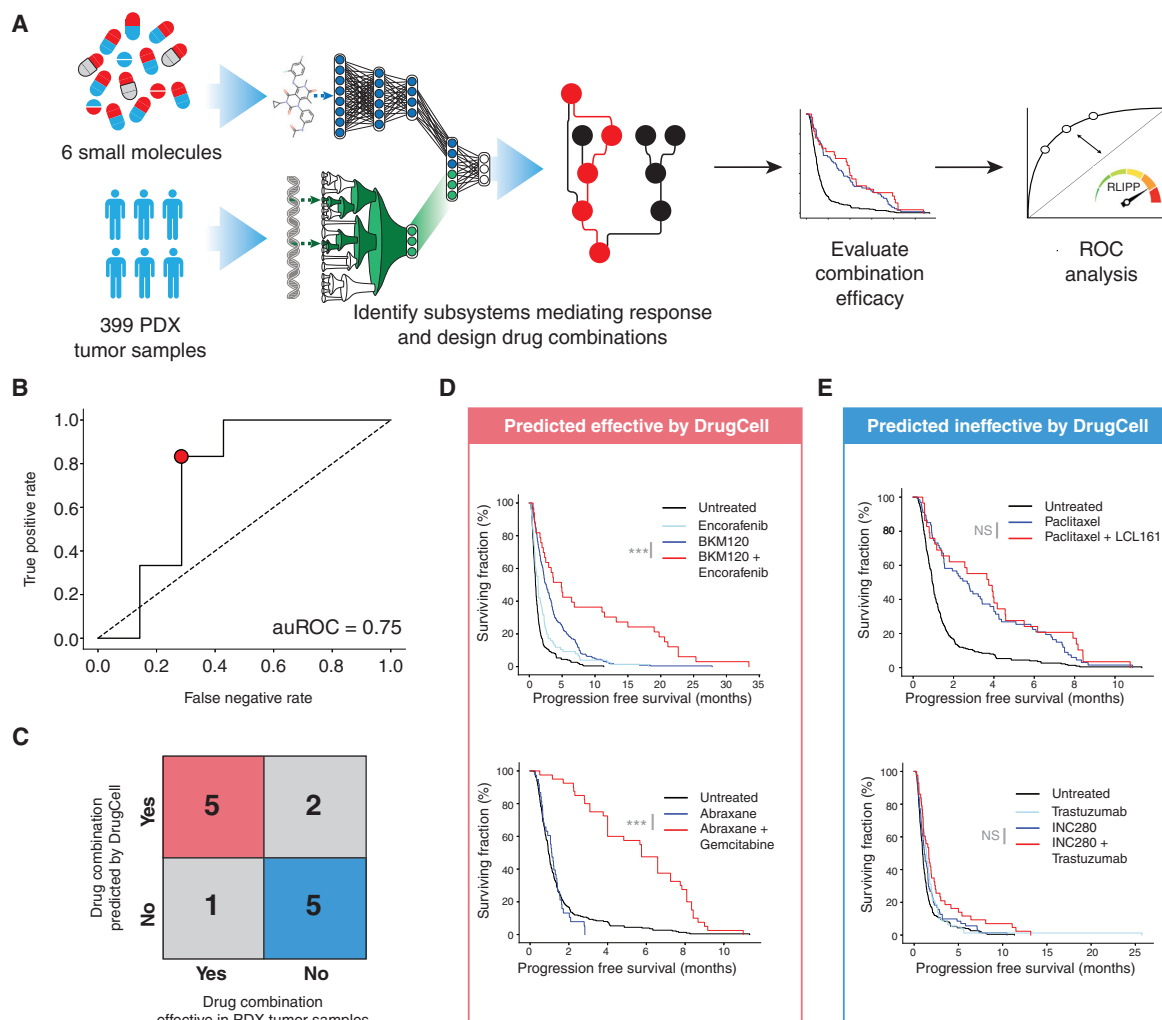


Figure 6. Guiding Combination Therapy in Patient-Derived Xenograft Tumors

(A) Flowchart of analysis procedure.

(B) ROC curve of DrugCell performance in distinguishing effective from ineffective drug combinations.

(C) Error matrix for point indicated in (B) demonstrating best performance of DrugCell against the PDX dataset.

(D) Survival curves for drug combinations predicted to be effective by DrugCell (true positives) showing a significant improvement in progression-free survival.

(E) Survival curves for drug combinations predicted to be ineffective by DrugCell (true negatives) showing a lack of improvement in progression-free survival. p values indicate significance by log rank test. ***p < 0.0001, NS indicates not significant.

With respect to specific genetic alterations, we found that DrugCell (+) patients were much more likely to harbor *AKT1* mutations than DrugCell (–) patients (Figure 7D). In contrast, DrugCell (–) patients had mutations in genes previously associated with drug resistance, including *ESR1* (Reinert et al., 2017), *RB1* (Condorelli et al., 2018), and *PTEN* (Costa et al., 2020) (Figure 7D), suggesting that we had stratified patients based on a complex pattern of mutations leading to therapy resistance. Strikingly, *AKT1* mutation status alone was not predictive of therapeutic response, with *AKT1*-mutant patients actually trending toward shorter overall survival (35.8 versus 43.1 months), although this difference was not statistically significant (Figure 7E). This analysis illustrates how DrugCell can be used to effectively guide clinical treatment decisions with significantly greater precision and insight than single-gene marker studies.

DISCUSSION

Here we have explored an interpretable deep learning model of the structure and function of a human cancer cell in response to treatment. This work advances predictive modeling toward a systematic representation of the biological mechanisms underlying a drug response, a critical direction for precision medicine. Following a model prediction, access to a mechanistic interpretation engages the experimentalist or clinician in reasoning about biological function. For example, analysis of DrugCell's model of etoposide identified a small set of subsystems important for the cellular response and for which targeted drugs were available (Figure 5). This analysis motivated us to perform subsequent experiments to target both genetically and pharmacologically topoisomerase II with either MAPK or PI3K pathways; both of these combinations showed significant synergistic effects.

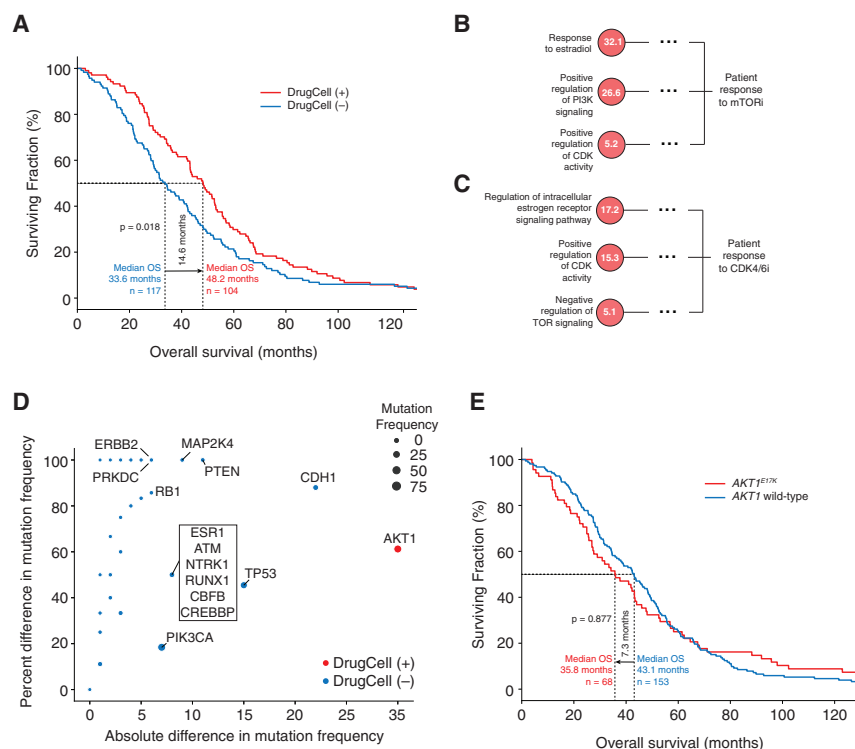


Figure 7. Guiding CDK4/6 and mTOR Inhibitor Therapy in ER-Positive Breast Cancer Patients

(A–C) (A) Survival curves for DrugCell (+) and DrugCell (-) patients treated with CDK4/6 or mTOR inhibitors in any line of therapy. The p value indicates significance by log rank test. (B, C) Important subsystems used by DrugCell to simulate (B) mTOR or (C) CDK4/6 inhibitor sensitivity. Dotted line abbreviates parent subsystems at subsequent layers of the hierarchy. RLIPP scores are displayed inside each node.

(D) Scatterplot of the absolute (x axis) and percentage (y axis) difference in mutation frequencies of genes between DrugCell (+) and DrugCell (-) patients. Red points represent genes mutated more frequently in DrugCell (+) patients. Blue points represent genes mutated more frequently in DrugCell (-) patients. Point size is proportional to overall mutation frequency in the patient population.

(E) Survival curves for AKT1-mutant and wild-type patients treated with CDK4/6 or mTOR inhibitors in any line of therapy. The p value indicates significance by log rank test.

Such engagement of human reasoning and follow-up experimentation helps greatly to increase accountability and trust in the predictions of a machine learning model. In contrast, conventional black-box predictive modeling yields only a model output—the drug response—without further information by which to build trust in the process.

DrugCell is a flexible model that is amenable to both automated and semi-automated combinatorial drug design. First, the importance of each cellular subsystem is scored by DrugCell during a response to monotherapy. These important subsystems are then annotated with second points of intervention, such as the PI3K or ERK pathway in the response to etoposide (Figures 5E–5G). To follow up on this analysis, drug combinations can be selected automatically based on the druggable targets present in top DrugCell subsystems. Alternatively, if DrugCell is being used in a clinical context, its recommendations can be provided to physician-scientists (e.g., a molecular tumor board) who consider the recommended combinations in light of other biological knowledge not explicitly used in modeling, such as potential toxicities and specific information about the case. After careful consideration of all relevant information, the ultimate treatment decision remains in the hands of the physician and the patient. Such need for human accountability is not unique to drug response prediction but is a central tenet of high-stakes applications of machine learning (Rudin, 2019).

Notably, previous models trained on monotherapy responses (Ammad-ud-din et al., 2017; Cortés-Ciriano et al., 2016; Iorio et al., 2016; Zhang et al., 2015) have not attempted to suggest combination therapies. Rather, drug combinations have been predicted using models of synergy trained directly on data from pairwise drug treatments (Preuer et al., 2018). This brute force approach faces the challenge of scalability, given the

combinatorial number of pairwise and higher order drug combinations necessary for training.

If the favorable performance observed in PDX samples (Figure 6) and ER-positive breast cancer patients (Figure 7) continues in further clinical studies, DrugCell and its successors have the potential to substantially expand the set of clinically meaningful mutations. DrugCell translates the mutational status of approximately 3,000 genes into treatment recommendations. Although we have not fully studied which of these genes are absolutely required for DrugCell prediction accuracy, RLIPP analysis suggests that many of them are—1,467 of the 2,086 subsystems are assigned relatively high importance (RLIPP >10) for at least one drug, collectively covering 2,855 genes. This breadth of information contrasts with the fewer genes included in current cancer mutation panels such as MSK-IMPACT or FoundationOne CDx (468 and 324 genes, respectively), which were designed to be queried manually by a physician (Cheng et al., 2015; Harris, 2017). Moreover, since we currently do not understand the clinical implications of the majority of cancer mutations, there is little consensus on what genes should be included in these pan-cancer mutation panels (Nguyen and Gocke, 2017) or on how physicians should act on the results. An increase in the number of clinically meaningful cancer mutations, facilitated by interpretable machine learning models such as DrugCell, could further motivate the case for complete genomic sequencing of cancer patients (Katsanis and Katsanis, 2013; Kuenzi and Ideker, 2020).

Future work may also elect to integrate mutations with additional levels of molecular information such as epigenetic states, gene expression, or microenvironmental influences. This integration could be accomplished by pre-processing multiple layers of information to derive a profile of gene scores for each

cell line or tumor, which would then be input to DrugCell. Extra levels of information could also be integrated by adding new visible or conventional neural network branches alongside existing ones. Alternatively, the effects of specific mutations on gene functions could be incorporated by a metric such as the Combined Annotation-Dependent Depletion score (Rentzsch et al., 2019) or by including gene structural domains as an additional layer of the hierarchy.

Another opportunity is to structure the DrugCell system hierarchy from 'omics data rather than literature curation (GO), as has previously been done in budding yeast (Kramer et al., 2014; Ma et al., 2018). A data-driven, rather than literature-curated, hierarchy has the potential to incorporate new gene-subsystem associations as well as entirely new subsystems into the model. It also has the potential to revise and tailor subsystem definitions in GO, which are generic, to their particular contexts relevant to cancer. For instance, we found that in its current form DrugCell contains a number of subsystems that have misleading labels based on GO naming conventions. For example, *Labyrinthine development* was among the top subsystems for trametinib, which was initially puzzling, but upon further inspection corresponds to MAPK cascade genes with well-known involvement in cancer proliferation (e.g., *MAP2K1*, *MAPK1*, *GRB2*, *FGFR2*). Incorporating data-driven hierarchies into DrugCell provides a route to relabel such subsystems and revise their specific gene contents. Finally, given that DrugCell inputs a full drug structure, it can potentially be used to design compounds *de novo*. Leveraging advancements in reinforcement learning for drug design (Zavoronkov et al., 2019), it may then be possible to design compounds for maximal efficacy against any given genomic background.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Data and Code Availability
 - Materials Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Culture and Reagents
- METHOD DETAILS
 - Defining a Hierarchy of Genes and Cellular Subsystems
 - Pharmacogenomics Data Processing and Morgan Fingerprint Encoding
 - Neural Network Configuration, Training and Evaluation
 - Implementation of Alternative Models for Comparison of Predictive Performance
 - Ranking Important Subsystems in DrugCell
 - Comparing Important DrugCell Subsystems to Predictive Biomarkers Reported by Alternative Models
 - Viability Assays
 - Combinatorial CRISPR-Cas9 Gene Knockouts and Systematic Evaluation
- QUANTIFICATION AND STATISTICAL ANALYSIS

- Assessing the Correspondence of Learned Subsystem Embeddings to Measured Subsystem Activities
- Differential Expression Analysis
- Synergy Determinations
- Translation of Continuous Cell Response (AUC) to Binary Cell Response
- Identification of Boolean Logic Combinations
- PDX Tumor Analysis
- Breast Cancer Patient Analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.ccell.2020.09.014>.

ACKNOWLEDGMENTS

We thank Dr. Gaudenz Danuser for his valuable insights into mechanisms of taxol sensitivity. We gratefully acknowledge the support for this work provided by grants from the National Institutes of Health to T.I. (GM103504, CA209891, ES014811), J.F.K. (CA243885), and B.M.K. (CA212456).

AUTHOR CONTRIBUTIONS

B.M.K., J.P., S.H.F., J.M., J.F.K., and T.I. designed the study and developed the conceptual ideas. B.M.K. collected all the input sources and additional data. J.P., B.M.K., and J.M. implemented the main algorithm. B.M.K. and J.P. implemented all other computational methods and analyses. B.M.K., S.H.F., J.L., and K.S. performed all experiments and B.M.K. analyzed the results. B.M.K., J.P., S.H.F., J.F.K., and T.I. wrote the manuscript with suggestions from other authors.

DECLARATION OF INTERESTS

T.I. is the co-founder and a member of the advisory board for Data4Cure, Inc., and has an equity interest. T.I. has an equity interest in Ideaya BioSciences, Inc., and is a member of the Scientific Advisory Board. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies.

Received: April 8, 2020

Revised: August 7, 2020

Accepted: September 22, 2020

Published: October 22, 2020

REFERENCES

- Ammad-ud-din, M., Khan, S.A., Wennerberg, K., and Aittokallio, T. (2017). Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. *Bioinformatics* 33, i359–i368.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29.
- Bani, M.R., Nicoletti, M.I., Alkharouf, N.W., Ghilardi, C., Petersen, D., Erba, E., Sausville, E.A., Liu, E.T., and Giavazzi, R. (2004). Gene expression correlating with response to paclitaxel in ovarian carcinoma xenografts. *Mol. Cancer Ther.* 3, 111–121.
- Baptista, D., Ferreira, P.G., and Rocha, M. (2020). Deep learning for drug response prediction in cancer. *Brief. Bioinform* bbz171, <https://doi.org/10.1093/bib/bbz171>.
- 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.

- Breinig, M., Klein, F.A., Huber, W., and Boutros, M. (2015). A chemical–genetic interaction map of small molecules using high-throughput imaging in cancer cells. *Mol. Syst. Biol.* 11, 846.
- Chen, Y., Lun, A.T.L., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 5, 1438.
- Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z.Y., Won, H.H., Scott, S.N., et al. (2015). Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT). *J. Mol. Diagn.* 17, 251–264.
- Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interf.* 15, 20170387.
- 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. Genomics* 12, 18.
- Condorelli, R., Spring, L., O'Shaughnessy, J., Lacroix, L., Bailleux, C., Scott, V., Dubois, J., Nagy, R.J., Lanman, R.B., lafrate, A.J., et al. (2018). Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 29, 640–645.
- Copley, S.D. (2012). Moonlighting is mainstream: paradigm adjustment required. *BioEssays* 34, 578–588.
- Cortés-Ciriano, I., van Westen, G.J.P., Bouvier, G., Nilges, M., Overington, J.P., Bender, A., and Malliavin, T.E. (2016). Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* 32, 85–95.
- Costa, C., Wang, Y., Ly, A., Hosono, Y., Murchie, E., Walmsley, C.S., Huynh, T., Healy, C., Peterson, R., Yanase, S., et al. (2020). PTEN loss mediates clinical cross-resistance to CDK4/6 and PI3K α inhibitors in breast cancer. *Cancer Discov.* 10, 72–85.
- 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.
- Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* 13, 21–27.
- Das, B., Yeger, H., Tsuchida, R., Torkin, R., Gee, M.F.W., Thorner, P.S., Shibuya, M., Malkin, D., and Baruchel, S. (2005). A hypoxia-driven vascular endothelial growth factor/flt1 autocrine loop interacts with hypoxia-inducible factor-1 α through mitogen-activated protein kinase/extracellular signal-regulated kinase 1/2 pathway in neuroblastoma. *Cancer Res.* 65, 7267–7275.
- Dincer, A.B., Celik, S., Hiranuma, N., and Lee, S.-I. (2018). DeepProfile: deep learning of cancer molecular profiles for precision medicine. *BioRxiv*. <https://doi.org/10.1101/278739>.
- Eskiocak, B., McMillan, E.A., Mendiratta, S., Kollipara, R.K., Zhang, H., Humphries, C.G., Wang, C., Garcia-Rodriguez, J., Ding, M., Zaman, A., et al. (2017). Biomarker accessible and chemically addressable mechanistic subtypes of BRAF melanoma. *Cancer Discov.* 7, 832–851.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., and Dean, J. (2019). A guide to deep learning in healthcare. *Nat. Med.* 25, 24.
- Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* 21, 1318–1325.
- Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483, 570–575.
- Hare, S.H., and Harvey, A.J. (2017). mTOR function and therapeutic targeting in breast cancer. *Am. J. Cancer Res.* 7, 383–404.
- Harris, J. (2017). FDA Approves FoundationOne CDx, CMS Agrees to Cover (OncLive Novemb).
- Hatzis, C., Bedard, P.L., Birkbak, N.J., Beck, A.H., Aerts, H.J.W.L., Stern, D.F., Shi, L., Clarke, R., Quackenbush, J., and Haibe-Kains, B. (2014). Enhancing reproducibility in cancer drug screening: how do We move forward? *Cancer Res.* 74, 4016–4023.
- Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., et al. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35, W169–W175.
- Hwang, S. (2012). Comparison and evaluation of pathway-level aggregation methods of gene expression data. *BMC Genomics* 13, S26.
- 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.
- Kang, H.C., Kim, I.-J., Park, J.-H., Shin, Y., Ku, J.-L., Jung, M.S., Yoo, B.C., Kim, H.K., and Park, J.-G. (2004). Identification of genes with differential expression in acquired drug-resistant Gastric cancer cells using high-Density Oligonucleotide microarrays. *Clin. Cancer Res.* 10, 272–284.
- Karnaugh, M. (1953). The map method for synthesis of combinational logic circuits. *Trans. Am. Inst. Electr. Eng. Part Commun. Electron.* 72, 593–599.
- Katsanis, S.H., and Katsanis, N. (2013). Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* 14, 415–426.
- Kramer, M., Dutkowski, J., Yu, M., Bafna, V., and Ideker, T. (2014). Inferring gene ontologies from pairwise similarity data. *Bioinformatics* 30, i34–i42.
- Kuenzi, B.M., and Ideker, T. (2020). A census of pathway maps in cancer systems biology. *Nat. Rev. Cancer* 20, 1–14.
- Kuenzi, B.M., Rix, L.L.R., Kinose, F., Kroeger, J.L., Lancet, J.E., Padron, E., and Rix, U. (2019). Off-target based drug repurposing opportunities for tivantinib in acute myeloid leukemia. *Sci. Rep.* 9, 606.
- Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., Vellano, C.P., Roebuck, P., Yu, Q., Eterovic, A.K., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. *Cancer Cell* 31, 225–239.
- Liu, T., Sun, H., Zhu, D., Dong, X., Liu, F., Liang, X., Chen, C., Shao, B., Wang, M., Wang, Y., et al. (2017). TRA2A promoted paclitaxel resistance and tumor progression in triple-negative breast cancers via regulating alternative splicing. *Mol. Cancer Ther.* 16, 1377–1388.
- Loewe, S. (1953). The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 3, 285–290.
- Ma, J., Yu, M.K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., and Ideker, T. (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods.* 15, 290–298.
- Ma, Y., Wang, L., Neitzel, L.R., Loganathan, S.N., Tang, N., Qin, L., Crispi, E.E., Guo, Y., Knapp, S., Beauchamp, R.D., et al. (2017). The MAPK pathway regulates intrinsic resistance to BET inhibitors in colorectal cancer. *Clin. Cancer Res.* 23, 2027–2037.
- Mathur, R., Chandna, S., N Kapoor, P., and S Dwarakanath, B. (2011). Peptidyl prolyl isomerase, Pin1 is a potential target for enhancing the therapeutic efficacy of etoposide. *Curr. Cancer Drug Targets* 11, 380–392.
- Menden, M.P., Iorio, F., Garnett, M., McDermott, U., Benes, C.H., Ballester, P.J., and Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* 8, <https://doi.org/10.1371/journal.pone.0061318>.
- Michaloglou, C., Crafter, C., Siersbaek, R., Delpuech, O., Curwen, J.O., Carnevalli, L.S., Staniszevska, A.D., Polanska, U.M., Cheraghchi-Bashi, A., Lawson, M., et al. (2018). Combined inhibition of mTOR and CDK4/6 is required for optimal blockade of E2F function and long-term growth inhibition in estrogen receptor-positive breast cancer. *Mol. Cancer Ther.* 17, 908–920.
- Moos, P.J., and Fitzpatrick, F.A. (1998). Taxane-mediated gene induction is independent of microtubule stabilization: induction of transcription regulators and enzymes that modulate inflammation and apoptosis. *Proc. Natl. Acad. Sci. U S A* 95, 3896–3901.

- Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U S A* **116**, 22071–22080.
- Nguyen, D., and Gocke, C.D. (2017). Managing the genomic revolution in cancer diagnostics. *Virchows Arch.* **471**, 175–194.
- Nutt, C.L., Noble, M., Chambers, A.F., and Cairncross, J.G. (2000). Differential expression of drug resistance genes and chemosensitivity in glial cell lineages correlate with differential response of oligodendrogliomas and astrocytomas to chemotherapy. *Cancer Res.* **60**, 4812–4818.
- Occhipinti, G., Romagnoli, E., Santoni, M., Cimadamore, A., Sorgentoni, G., Cecati, M., Giulietti, M., Battelli, N., Maccioni, A., Storti, N., et al. (2020). Sequential or concomitant inhibition of cyclin-dependent kinase 4/6 before mTOR pathway in hormone-positive HER2 negative breast cancer: biological insights and clinical implications. *Front. Genet.* **11**, 349.
- Park, J.S., Burckhardt, C.J., Lazcano, R., Solis, L.M., Isogai, T., Li, L., Chen, C.S., Gao, B., Minna, J.D., Bachoo, R., et al. (2020). Mechanical regulation of glycolysis via cytoskeleton architecture. *Nature* **578**, 621–626.
- Pernas, S., Tolaney, S.M., Winer, E.P., and Goel, S. (2018). CDK4/6 inhibition in breast cancer: current practice and future directions. *Ther. Adv. Med. Oncol.* **10**.
- Potts, M.B., McMillan, E.A., Rosales, T.I., Kim, H.S., Ou, Y.-H., Toombs, J.E., Brekken, R.A., Minden, M.D., MacMillan, J.B., and White, M.A. (2015). Mode of action and pharmacogenomic biomarkers for exceptional responders to idemnin B. *Nat. Chem. Biol.* **11**, 401–408.
- Pozdeyev, N., Yoo, M., Mackie, R., Schweppe, R.E., Tan, A.C., and Haugen, B.R. (2016). Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget* **7**, 51619–51625.
- Pratt, D., Chen, J., Welker, D., Rivas, R., Pillich, R., Rynkov, V., Ono, K., Miello, C., Hicks, L., Szalma, S., et al. (2015). NDEx, the network data Exchange. *Cell Syst.* **1**, 302–305.
- Preuer, K., Lewis, R.P.I., Hochreiter, S., Bender, A., Bulusu, K.C., Klambauer, G., and Wren, J. (2018). DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538–1546.
- Rajkumar, A., Dean, J., and Kohane, I. (2019). Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358.
- Rampásek, L., Hidru, D., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. (2019). DrVAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics* **35**, 3743–3751.
- Reinert, T., Saad, E.D., Barrios, C.H., and Bines, J. (2017). Clinical implications of ESR1 mutations in hormone receptor-positive advanced breast cancer. *Front. Oncol.* **7**, 26.
- Rentsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215.
- Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T.J., Piha-Paul, S., Zhou, H., Kardala, E., et al. (2019). A deep learning framework for predicting response to therapy in cancer. *Cell Rep.* **29**, 3367–3373.e4.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784.
- Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* **5**, 1210–1223.
- Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14**, 573–576.
- Shimamura, T., Chen, Z., Soucheray, M., Carretero, J., Kikuchi, E., Tchaicha, J.H., Gao, Y., Cheng, K.A., Cohoon, T.J., Qi, J., et al. (2013). Efficacy of BET bromodomain inhibition in Kras-mutant non-small cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 4325.
- Smyth, L.M., Zhou, Q., Nguyen, B., Yu, C., Lepisto, E.M., Arnedos, M., Hassett, M.J., Lenoue-Newton, M.L., Blauvelt, N., Dogan, S., et al. (2020). Characteristics and outcome of AKT1E17K-mutant breast cancer defined through AACR Project GENIE, a clinicogenomic registry. *Cancer Discov.* **10**, 526–535.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* **102**, 15545–15550.
- Suzuki, S., Horinouchi, T., and Furusawa, C. (2014). Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* **5**, 1–12.
- Topol, E.J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56.
- Turner, R.M., Park, B.K., and Pirmohamed, M. (2015). Parsing interindividual drug variability: an emerging role for systems pharmacology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **7**, 221–241.
- Wainberg, M., Merico, D., Delong, A., and Frey, B.J. (2018). Deep learning in biomedicine. *Nat. Biotechnol.* **36**, 829–838.
- Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y., et al. (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.* **48**, D1031–D1041.
- Wong, C.H., Siah, K.W., and Lo, A.W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273–286.
- Yang, J.H., Wright, S.N., Hamblin, M., McCloskey, D., Alcantar, M.A., Schrübers, L., Lopatkin, A.J., Satish, S., Nili, A., Palsson, B.O., et al. (2019). A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell* **177**, 1649–1661.e9.
- Yang, L., Ainali, C., Tsoka, S., and Papageorgiou, L.G. (2014). Pathway activity inference for multiclass disease classification through a mathematical programming optimisation framework. *BMC Bioinformatics* **15**, 390.
- Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961.
- Yeh, P.J., Hegreness, M.J., Aiden, A.P., and Kishony, R. (2009). Drug interactions and the evolution of antibiotic resistance. *Nat. Rev. Microbiol.* **7**, 460–466.
- Yin, Y., Sun, M., Zhan, X., Wu, C., Geng, P., Sun, X., Wu, Y., Zhang, S., Qin, J., Zhuang, Z., et al. (2019). EGFR signaling confers resistance to BET inhibition in hepatocellular carcinoma through stabilizing oncogenic MYC. *J. Exp. Clin. Cancer Res.* **38**, 83.
- Yu, M.K., Ma, J., Fisher, J., Kreisberg, J.F., Raphael, B.J., and Ideker, T. (2018). Visible machine learning for biomedicine. *Cell* **173**, 1562–1565.
- 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.
- Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., and Liu, X.S. (2015). Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS Comput. Biol.* **11**, e1004498.
- Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Paclitaxel	Selleckchem	S1150
2-deoxyglucose	Selleckchem	S4701
Critical Commercial Assays		
CellTiter-Glo Luminescent Cell Viability Assay	Promega	G7572
Blood and Cell Culture DNA Mini Kit	Qiagen	Cat No./ID: 13323
Deposited Data		
DrugCell hierarchy	This paper	http://ndexbio.org/#/network/a20b3699-4862-11ea-bfdc-0ac135e8bacf
Experimental Models: Cell Lines		
A549	ATCC	CCL-185
A427	ATCC	HTB-53
MCF7	ATCC	HTB-22
Oligonucleotides		
lentiCRISPRv2	Sanjana et al., 2014	Addgene Plasmid #52961
lentiCas9-Blast	Sanjana et al., 2014	Addgene Plasmid #52962
See Table S6 for gRNA sequences		
Software and Algorithms		
limma	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
DrugCell algorithm	This Paper	https://github.com/idekerlab/DrugCell

RESOURCE AVAILABILITY

Lead Contact

Correspondence and requests regarding this manuscript should be sent to and will be fulfilled by the lead investigator Dr. Trey Ideker (tideker@ucsd.edu).

Data and Code Availability

DrugCell code and sample training data are available on GitHub (<https://github.com/idekerlab/DrugCell>) and the version of the codebase used in the manuscript is archived on Zenodo (<https://zenodo.org/badge/latestdoi/250580982>). The trimmed version of the Gene Ontology (GO) used as the architecture for DrugCell is visualized and available for download on the Network Data Exchange (NDEx; URL: <http://ndexbio.org/#/network/a20b3699-4862-11ea-bfdc-0ac135e8bacf>) ([Pratt et al., 2015](#)). Other data sources used in this study are available from their original publications and web portals.

Materials Availability

This study did not generate any unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture and Reagents

A549, A427, and MCF7 cells were retrieved from the American Type Culture Collection (ATCC) and cultured in DMEM + 10% FBS or EMEM + 10% FBS according to ATCC recommendations. All cell lines tested negative for mycoplasma contamination and were authenticated by short tandem repeat (STR) analysis. Paclitaxel (Selleckchem) was dissolved in DMSO (10mM) and diluted in media for use. 2-deoxy-d-glucose (Selleckchem) was dissolved in media (100mM), filtered, and further diluted in media for use.

METHOD DETAILS

Defining a Hierarchy of Genes and Cellular Subsystems

To computationally represent cancer genotypes, we selected the top 15% most frequently mutated genes in human cancers according to the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) among genes annotated to Gene Ontology (GO) terms (Ashburner et al., 2000). This procedure yielded 3,008 genes, henceforth called 'DrugCell genes', which were used in model construction. These genes were organized into a hierarchy of nested gene sets, representing cellular subsystems at different scales, based on terms extracted from the GO Biological Process hierarchy. Terms were retained from GO if they had at least 10 DrugCell genes and were distinct from all child terms, defined as having at least 30 DrugCell genes more than any child (both part_of and is_a hierarchical term relations were considered). Every other term was removed from the hierarchy, and instead its children were assigned directly to its parents to keep the hierarchy connected. To further reduce model complexity, we restricted the hierarchy to a maximal depth of five subsystems by removing all subsystems more than five parent-child relations above the bottom layer subsystems of the hierarchy (subsystems without any children). The resulting hierarchy, composed of 2,086 subsystems, defined the branch of DrugCell for embedding of genotype (left branch in Figure 1A, also called the VNN; Figure 1B).

Pharmacogenomics Data Processing and Morgan Fingerprint Encoding

To obtain a sufficiently large pharmacogenomic dataset for model training, raw drug sensitivity data were retrieved from the Genomics of Drug Sensitivity in Cancer database (GDSC) and the Cancer Therapeutics Response Portal v2 (CTRP) (Seashore-Ludlow et al., 2015; Yang et al., 2013). These data covered a total of 509,294 (cell line, drug) pairs. Among these data, 24,923 pairs redundantly measured in the two repositories were left intact in the training dataset, as such replicates can be beneficial to reduce model over-fitting. Luminescence values were background corrected (media only), normalized to vehicle treatment (DMSO) at each compound concentration, and replicate values averaged. To standardize across the two datasets, we calculated the Area Under dose response Curve (AUC) normalized such that AUC = 0 represents complete cell killing, AUC = 1 represents no effect, and AUC > 1 represents a treatment granting a growth advantage to the cells. Curves were created by connecting individual response points in a piecewise linear fashion, rather than using a sigmoid curve fit. We then normalized the AUC of this piecewise linear fit to the area under a null curve spanning the tested concentration range. The calculated AUC values were in high agreement with previous analyses of this dataset ($r^2 = 0.87$) while correcting for artifacts introduced by forced sigmoid curve fitting seen in other studies (Seashore-Ludlow et al., 2015) (Figure S7A). No batch correction was performed in addition to AUC standardization. The correlation between AUC values present in both datasets was on par ($n = 24,923$; Spearman rho = 0.5) with previous studies (Hatzis et al., 2014; Pozdeyev et al., 2016). To standardize drug representation across datasets, we queried the PubChem entry for each compound used in CTRP or GDSC to obtain an isomeric SMILES notation based on the drug name or InChIKey provided in the dataset. Compounds with no matches in the initial search were manually annotated. To computationally represent chemical structure we used RDKit (<http://www.rdkit.org/>) to calculate a Morgan fingerprint (radius = 2), which decomposes each chemical structure into molecular fragments by iteratively obtaining distinct paths through each atom of the molecule. These fragments were hashed into a bit vector of length 2,048 to be used for model training. Genotypes of each cell line were formulated from non-synonymous coding mutations as previously annotated and used by the Cancer Cell Line Encyclopedia (CCLE; <http://portals.broadinstitute.org/ccle>, 18q2 release) (Barretina et al., 2012). The dataset was filtered to represent only the top 15% most frequently mutated genes ($n = 3,008$). Each cell-line genotype was represented as a bit vector across the 3,008 DrugCell genes indicating the mutational status of each gene in that cell line (0 = wild type; 1 = mutated).

Neural Network Configuration, Training and Evaluation

The DrugCell VNN (the genotype embedding branch; Figures 1A and 1B) was configured following the DCell protocol (Ma et al., 2018) with minor modifications. Each subsystem s in DCell, and also in the hierarchy of subsystems in DrugCell (see above), is assigned a number k of neurons to represent its multidimensional state. This subsystem state, denoted by the output vector $O^{(s)}$ is defined as a function of the states of its c child subsystems and g directly annotated genes, concatenated in the input vector $I^{(s)}$:

$$O^{(s)} = f(W^{(s)}I^{(s)} + b^{(s)})$$

$W^{(s)}$ is a weight matrix of dimensions $k \times (k \times c + g)$ and $b^{(s)}$ is a weight vector of dimension k . $W^{(s)}$ and $b^{(s)}$ provide the parameters to be learned for subsystem s . The function f is a non-linear transformation based on hyperbolic tangent and batch normalization. Training of parameters is performed using an objective (loss) function based on mean-squared error and an optimization procedure based on standard gradient descent and back-propagation. All parameters are initialized uniformly at random between -0.001 and 0.001 .

In what follows, we focus on aspects of DrugCell that significantly build on or depart from the original DCell model (Ma et al., 2018). First, in parallel to the subsystem hierarchy used to embed genotype, DrugCell implements a drug embedding branch configured as a conventional artificial neural network with three hidden layers, with the neurons of each layer fully connected to the next (these three layers have 100, 50, and 6 neurons respectively, see Figure 1C). The input vector to this ANN is the 2,048-bit Morgan fingerprint of a drug (described above) and is fully-connected to the first hidden layer with 100 neurons. The final layer is a set of six neurons representing the drug embedding learned by DrugCell. These six neurons are concatenated with the six-neuron genotype embedding

(see above) and fed to an additional hidden layer of six neurons, which feeds a final output layer of a single neuron representing the predicted drug response, $O^{(DC)}$, measured as a continuous valued AUC (see Figure 1A and Pharmacogenomics Data Processing and Morgan Fingerprint Encoding section, above). Second, the number of neurons per subsystem k (VNN branch, see above) is selected by training and evaluation of a progression of neural network models with increasing values of this parameter ($k = 1, 3, 6, 9, 12$; Figure S7B). The DrugCell model used for all subsequent analysis is configured with $k = 6$, as this value yielded the best Spearman rho between actual and predicted drug responses across all (cell-line, drug) pairs. The DrugCell model is implemented using the PyTorch library and trained using three GPU servers (two servers with Nvidia RTX 2080Ti with 4352 CUDA cores and 11Gb GDDR6 RAM; one server with Tesla K80 with 4992 CUDA cores and 24Gb GDDR5 RAM).

Model predictive performance was evaluated using a standard training / validation / test procedure. The 509,294 (cell line, drug) pairs in the data were divided into five groups of approximately equal size. Five separate models were created, in which each of these groups was held out as the test data, and the remaining four groups were pooled for training and validation. During the training phase of each model, 5,000 random (cell line, drug) pairs were further withheld for use as a validation set on which model predictive performance was used as an early terminating condition; all remaining samples were designated as training. Each model was trained through a maximum of 300 epochs; performance on the validation data was evaluated after each epoch and training was terminated early in the event of decreasing model performance. The performance of each model was measured using Spearman rho between actual and predicted drug responses (AUC) in the test data and the final overall performance was reported ($\rho = 0.80$) as the average rho across the five models. Following evaluation of model performance (Figure 2), we used a model trained using all 509,294 (cell line, drug) pairs to ensure maximal predictive power and interpretability (Figures 3 and 7).

Implementation of Alternative Models for Comparison of Predictive Performance

We compared DrugCell to several alternative models trained using the same data as DrugCell: an elastic net (Figure 2B) and two fully connected neural network models (Figures 2C and 2D). A similar test procedure for 5-fold cross validation to that described above was used for evaluation of all these models. The elastic net model was implemented using the ElasticNetCV function in the scikit-learn library with $cv = 5$. A black-box neural network model (“Matched” in Figure 2C) was designed to have an identical hierarchical structure as DrugCell, but with the gene annotation inputs (gene-to-subsystem assignments) randomly shuffled. The predictive performance was reported as an average Spearman (ρ) across 10 such random models. A second black-box model (“Tissue only” in Figure 2D) was a fully connected neural network model, whose input was a 2,049-bit vector concatenating the 2,048-bit Morgan fingerprint representation of each drug and a single bit indicating the tissue of origin for each cell line. All elements in the input layer were fed to a stack of five hidden layers, of which each has 1,000, 500, 200, 100, 50 neurons respectively. The final hidden layer of 50 neurons was connected to a single neuron representing the predicted drug response output. We further compared DrugCell’s predictive performance to that of five additional models, using the predictive performance reported in the corresponding publications rather than reimplementing those models directly in our study (Ammad-ud-din et al., 2017; Cortés-Ciriano et al., 2016; Iorio et al., 2016; Zhang et al., 2015) (Table S1).

Ranking Important Subsystems in DrugCell

To quantitatively determine important subsystems for drug response prediction, we adopted the Relative Local Improvement in Predictive Power (RLIPP) score as described previously for DCell (Ma et al., 2018). Briefly, for each subsystem in DrugCell we constructed and compared two different L_2 -norm penalized linear regression models of drug response local to that subsystem. The first regression model predicts drug response using the neuron values that represent the state of the subsystem under the different genotypes. The second regression model predicts drug response using the neuron values that represent the states of the subsystem’s children. Both models are optimized to predict drug response, but with consecutive layers of neurons located at and below the subsystem of interest in DrugCell. Performance is calculated as the Spearman correlation (ρ) between the actual and predicted drug responses for each of the two alternative linear regression models (AUC). The RLIPP score is then defined as the ratio of Spearman rho of the first linear model to that of the second linear model. $RLIPP > 1$ reflects that the state of the parent subsystem has more predictive power for drug response than the mere concatenation of the states of its children, indicating the importance of the parent subsystem in learning.

Comparing Important DrugCell Subsystems to Predictive Biomarkers Reported by Alternative Models

Beyond a comparative assessment of predictive accuracy (see above), we wished to compare the genes and subsystems nominated by DrugCell to genetic markers reported previously. For this comparison, we focused on predictive models published by the GDSC (elastic net and random forest; Table S4) (Iorio et al., 2016) in a previous analysis of the same cell-line drug response data as examined by our study (see Pharmacogenomics Data Processing and Morgan Fingerprint Encoding above). We focused on 60 drugs for which GDSC had published predictive gene mutations that were relatively frequent in tumors (top 15% of mutated genes in the integrated GDSC and CTRP dataset, see above). For each of these drugs, we listed the genetic mutations identified as predictive biomarkers in the GDSC study, along with the corresponding DrugCell subsystem containing that gene and its RLIPP score. Separately, we examined the top three subsystems reported by DrugCell according to RLIPP score. For every gene in one of these subsystems, we determined the maximum weight connecting that gene to the neuron of that subsystem that is most relevant to the observed drug response (AUC); the top three genes by weight were reported (Table S4). To select the most relevant neuron to drug response, we first found the principal component that has the strongest Spearman correlation with the observed drug response. We then determine the neuron with the highest loading (eigenvalue) to that principal component as the most relevant neuron to that response.

Viability Assays

Cell viability assays were conducted according to the manufacturer's specifications for CellTiter-Glo Luminescent Cell Viability Assay (Promega). Cells were seeded at 1,000 cells/well in a 384-well microtiter plate and treated after 24 hours. Drugs were diluted in the respective culture medium at the indicated concentrations. Cells were treated for 72h before the addition of CellTiter-Glo reagent and read on a Synergy HT Multi-Detection Microplate Reader (Biotek).

Combinatorial CRISPR-Cas9 Gene Knockouts and Systematic Evaluation

For gene knockout experiments, the CRISPR-Cas9 nuclease was stably integrated at the AAVS1 'safe harbor' locus in MCF7 cells. LentiCas9-Blast (Addgene plasmid # 52962; <http://n2t.net/addgene:52962>; RRID:Addgene_52962) and lentiCRISPR v2 (Addgene plasmid # 52961; <http://n2t.net/addgene:52961>; RRID:Addgene_52961) were gifts from Dr. Feng Zhang (Sanjana et al., 2014). MCF7-Cas9 cells were tested for *Mycoplasma* contamination, expanded, and frozen into multiple aliquots so that experiments could be performed at low passage numbers. Cells were grown in DMEM, 10% FBS, and hygromycin to select for Cas9 expression, which was confirmed by capillary western (Wes, Protein Simple). A custom library of double gRNA constructs (gene + non-targeting, gene + gene) was used which covers all single and pairwise combinations of 3 primary genes (*MEK1*, *PARP1*, and *TP53*) versus 176 secondary genes. These secondary genes were designed to be broadly representative of major cancer-related processes including proliferative signaling, cell cycle progression, transcription regulation, and DNA repair with special attention to druggable targets and tumor suppressor genes. Double (primary, secondary) gRNA constructs were designed as described previously (Shen et al., 2017) with three distinct 20-bp gRNAs per target gene (Table S6) along with three non-targeting controls for a total of $3 \times 3 = 9$ constructs per gene or gene pair. The library was packaged into lentiviruses, and MCF7 cells were infected at an MOI of 0.3 to ensure each cell had zero or one double gRNA construct. Puromycin selection (2.5 $\mu\text{g/mL}$) was started two days after transduction and the concentration was reduced by half upon each splitting to a final concentration of 0.625 $\mu\text{g/mL}$, which was maintained for the remainder of the experiment. Following initial puromycin selection, cells were maintained in exponential growth by harvesting and removing a fraction of cells every two days. DNA was extracted from cells after 21 days of growth with a Blood and Cell Culture DNA Mini Kit (Qiagen) according to manufacturer protocols. To assess the relative frequencies of gRNAs before and after selection, integrated DNA encoding the gRNA sequence was PCR amplified and prepared for HiSeq4000 sequencing (Illumina) according to manufacturer protocols. Standard Illumina primers were used for library preparation, and sequencing was conducted to generate 100-bp reads in a paired-end fashion. After sequencing, data quality was assessed with FastQC. Fitness effects of gene knockouts were determined as previously described (Shen et al., 2017) and normalized to the median fitness for non-targeting guides. Experiments were performed in biological duplicate.

To systematically validate the identified mechanisms of sensitivity to trametinib, olaparib and nutlin-3, we first ranked subsystems by importance in DrugCell simulation of each compound (RLIPP analysis, see above). This ranking was filtered to retain the top five subsystems that contained sufficient (three or more) secondary genes in our CRISPR library. These subsystems were all among the top 25 overall subsystems (top 1%) identified for each drug. We then examined the fitness effects resulting from pairwise knockout of the major target of each compound (*MAP2K1*, *PARP1* and *TP53*) together with each CRISPR library gene present in a top subsystem (up to a maximum of five genes). These pairwise knockout effects were compared to the effects of pairwise knockout of the major target of each compound together with knockout of genes in five random subsystems selected from among those with low RLIPP scores < 2 .

QUANTIFICATION AND STATISTICAL ANALYSIS

Assessing the Correspondence of Learned Subsystem Embeddings to Measured Subsystem Activities

To assess whether the subsystem states that DrugCell had learned are representative of experimentally measured activities of these subsystems, we adapted an expression-based analysis similar to that piloted by our previous DCell proof-of-concept (Ma et al., 2018). We obtained reverse phase protein array (RPPA) data (Li et al., 2017) covering 899 cell lines from the Cancer Cell Line Encyclopedia (CCLE), including the majority of cell lines for which genotypes and drug responses were used to train the DrugCell model. For each subsystem, we created a subsystem activity score similar to other methods that have been described for pathway-based gene expression analysis (Hwang, 2012; Yang et al., 2014). Here, we calculated the "RPPA activity" of each subsystem as the simple sum of signal intensities across all proteins and phosphorylation sites mapping to that subsystem. We then trained a random forest regression model to predict this RPPA activity using the top 6 principal components of that subsystem's DrugCell embedding as features. We compared the predictive performance of these individual subsystem models to models trained to predict the RPPA activity of random sets of genes of matched sizes (Figure S3A).

Differential Expression Analysis

We performed differential expression analysis, which is commonly used to identify pathways regulating drug sensitivity (Kang et al., 2004; Nutt et al., 2000; Suzuki et al., 2014), to identify pathways mediating paclitaxel sensitivity. Raw RNAseq count data were obtained from CCLE (<http://portals.broadinstitute.org/ccle>) for the top 25 most paclitaxel sensitive and 25 most paclitaxel resistant cell lines. Raw counts were transformed to log2 counts per million (log-CPM) and genes with low expression levels were removed (CPM < 0.1) as previously described (Chen et al., 2016). Data were normalized using the trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010). Differential expression was determined by linear modeling using limma (Ritchie et al., 2015). Pathways

enriched for differentially expressed genes were determined using DAVID (Huang et al., 2007). We identified 125 genes that were significantly differentially expressed ($q < 0.05$) when contrasting the 25 most paclitaxel-sensitive cell lines with the 25 most paclitaxel-resistant cell lines (Figure S5B). Pathways enriched for these genes included RNA splicing and cell division (Figure S5C), consistent with previous studies (Bani et al., 2004; Liu et al., 2017; Moos and Fitzpatrick, 1998), as well as pathways responding to ionizing radiation and DNA replication.

Synergy Determinations

Drug combination profiling data across a diverse cell line panel were obtained from DeepSynergy (<http://www.bioinf.jku.at/software/DeepSynergy/>) (Preuer et al., 2018), which used the Loewe model of additivity (Loewe, 1953) to evaluate the interaction of 583 different drug combinations across 39 human cancer cell lines. We used these DeepSynergy data to systematically evaluate the ability of DrugCell to pair a primary drug D_1 with a synergistic second agent D_2 by targeting top subsystems mediating sensitivity to the primary drug (see text and Figure 5). We collected protein target information for drugs from the Therapeutic Target Database (Wang et al., 2020), yielding targets for 283 drugs on which DrugCell had been trained in cell lines and 32 drugs considered by Deep Synergy. The 25 drugs in the intersection of these sets were used for systematic evaluation. For each drug in the set of 25 drugs $D_1 = \{d_i | i = 1 \dots 25\}$, we gathered a set of predicted synergistic target genes, $G_i = \bigcup_{j=1}^5 GOA(s_j^i)$, based on their membership in the top 5 subsystems by RLIPP score ($s_1^i, s_2^i, \dots, s_5^i$), where $GOA(s)$ is a set of genes contained in a subsystem s according to the Gene Ontology Annotation. We then collected a set of secondary drugs, $\{y_1^i, y_2^i, \dots, y_m^i\}$, targeting any gene in G_i and compiled a set of synergistic drug combinations, $\{d_i\} \times D_2^i = \{(d_i, y_1^i), (d_i, y_2^i), \dots, (d_i, y_m^i)\}$. Across all 25 primary drugs in D_1 , the set of synergistic secondary drugs, $D_2 = \bigcup_{i=1}^{25} D_2^i$, led to 75 predicted synergistic drug pairs with corresponding observed DeepSynergy scores. We repeated the process for the bottom 5 RLIPP subsystems to predict a set of non-synergistic secondary drugs, $D_2' = \bigcup_{i=1}^{25} D_2'^i$, leading to 70 predicted non-synergistic drug pairs with corresponding observed DeepSynergy scores. Finally, we compared the distribution of the synergy scores of the predicted synergistic pairs, the predicted non-synergistic pairs, and the remaining 76 drug pairs in DeepSynergy (Figure 5J).

Translation of Continuous Cell Response (AUC) to Binary Cell Response

In addition to Spearman correlation, we characterized DrugCell's predictive performance by the ability to separate cells into binary sensitive versus resistant response classes (Figures 5H–5J, Table S7). For this purpose, we binarized DrugCell's continuous predictions of drug response (AUC) as follows. Let $O_i(d)$ represent the actual response of cell line i exposed to drug d , reflecting the area under dose response curve (AUC), and let $O_i^{(DC)}(d)$ represent the corresponding predictive output of DrugCell. We then seek a drug-specific threshold, t_d , that maximizes balanced accuracy over all cell lines:

$$t_d = \operatorname{argmax}_x \left(\operatorname{mean} \left(\frac{|\{i | O_i^{(DC)}(d) \leq x \text{ and } O_i(d) \leq x\}|}{|\{i | O_i(d) \leq x\}|}, \frac{|\{i | O_i^{(DC)}(d) > x \text{ and } O_i(d) > x\}|}{|\{i | O_i(d) > x\}|} \right) \right)$$

DrugCell's prediction is then translated to a binary drug response ($B_i^{(DC)} \in \{0: \text{sensitive}, 1: \text{resistant}\}$) by use of t_d :

$$B_i^{(DC)} = \begin{cases} 1, & O_i^{(DC)}(d) > t_d \\ 0, & \text{otherwise} \end{cases}$$

Identification of Boolean Logic Combinations

We developed an approximate Boolean logic representation of how two subsystems, *Regulation of PI3K activity* and *Negative regulation of ERK1/ERK2 cascade* (henceforth called subsystems s and t), mediate the prediction of etoposide response in DrugCell. To achieve this Boolean representation, the continuous DrugCell prediction of drug AUC for each cell-line sample i ($O_i^{(DC)} \geq 0$) was translated to a binary drug response ($B^{(DC)} \in \{0: \text{etoposide sensitive}, 1: \text{etoposide resistant}\}$) by use of a threshold (see above). Here, a threshold of 0.82 was selected as it maximizes balanced accuracy when using DrugCell for binary classification of etoposide response over all samples:

		Observed Etoposide Response (AUC > 0.82?)		
		0	1	
DrugCell Predicted Response ($B^{(DC)} : O^{(DC)} > 0.82?$)	0	500	58	Balanced accuracy = MEAN [(500 / 724), (372 / 430)] = 0.78
	1	224	372	Odds Ratio = (500 / 58) / (224 / 372) = 14.32

In the above table, the same threshold was applied to both the predictions (rows) and the observations (columns). We also translated the multi-dimensional output vector of each subsystem ($O_i^{(s)}, O_i^{(t)}$) to a binary state ($B_i^{(s)}, B_i^{(t)} \in \{0: \text{unaltered}, 1: \text{altered}\}$) using $O^{(s)}$ or $O^{(t)}$ as features to classify $B^{(DC)}$ using a k -nearest neighbor (KNN) classifier with $k = 10$ (Cover and Hart, 1967). The output of this classifier was taken as the binary value of the subsystem, $B_i^{(s)}$ or $B_i^{(t)}$. For each subsystem, we selected three exemplary genes with high importance to the subsystem output, creating a vector of gene binary mutation states:

$$\vec{v}_i^j = (x_i^j, y_i^j, z_i^j), \text{ where } j \in \{s, t\} \text{ and } x, y, z \in \{0: \text{unmutated}, 1: \text{mutated}\}$$

These gene exemplars were defined as the three gene inputs most heavily weighted by DrugCell in connection to the neuron of s or t with the highest coefficient of variation over all i . The above procedure thus yielded binarized values for six genes, two subsystems, and one drug response. For each possible combination of binary gene inputs, \vec{v}^j , the (typically multiple) corresponding samples were examined to compute a consensus value $C_{\vec{v}^j}^{(j)}$ for the states of the two subsystems and the DrugCell output ($j \in \{s, t, DC\}$) according to the following rule:

$$C_{\vec{v}^j}^{(j)} = \begin{cases} 1, & P(B^{(j)} = 1 \mid \vec{v} = \vec{v}^j) > P(B^{(j)} = 1) \\ 0, & \text{otherwise} \end{cases}$$

This process yielded a logical truth table which was expressed as a minimal set of Boolean logic gates (Figure 4F) using the technique of Karnaugh maps (Karnaugh, 1953).

PDX Tumor Analysis

For each PDX tumor, measured mutations in DrugCell genes were used as input to DrugCell to predict the response to 6 drugs belonging to compound classes that DrugCell had previously seen (abraxane, binimetinib, encorafenib, INC-280, BKM-120, and BYL-719) and had combination data available, which altogether had been treated in 13 pairwise combinations with secondary drugs from diverse target classes. Since AUC data does not exist for *in vivo* experiments, tumor size was used as a surrogate for AUC. We then performed RLIPP analysis to identify subsystems mediating response to each of the 6 drugs included in this analysis. For each of the 13 available drug combinations, we defined a set of pathways that would lead to the design of that particular combination. We then scanned over 100 different RLIPP values, and at each cutoff compared the identified pathways with the pathways defined for each of the tested combinations to see if it had been identified. We considered a combination to be ‘effective’ if it significantly improved progression free survival as compared to the tested single drugs ($p < 0.05$, log-rank test). The observed PFS of each of these (primary, secondary) combinations was used to evaluate sensitivity and specificity as the number of top ranking DrugCell subsystems was progressively increased, yielding estimates of prediction sensitivity and specificity along a ROC curve for the combination panel.

Breast Cancer Patient Analysis

We obtained aggregated clinical trial data (Smyth et al., 2020) from Project GENIE (Genomics Evidence Neoplasia Information Exchange), an international genomics registry and data sharing platform established by the American Association for Cancer Research. This resource contained mutational profiling data and clinical outcomes for 457 metastatic breast cancer patients following multiple rounds of therapy. We removed patients from the dataset if they had not been treated with a targeted therapy (mTOR or CDK4/6 inhibitors). Such filtering produced a total of 221 estrogen receptor (ER)-positive metastatic breast cancer patients who had undergone treatment with an mTOR inhibitor (everolimus), a CDK4/6 inhibitor (ribociclib), or both compounds in any round of therapy. We predicted patient response to either mTOR or CDK4/6 inhibition using our pre-trained DrugCell model and the mutational profiles of each patient. We classified a patient as DrugCell (+) if they were predicted sensitive (\leq median predicted AUC across all patients) to either therapy (and had been treated with that particular therapy). Conversely, we classified patients as DrugCell (–) if they were predicted insensitive to both therapies. We used a log-rank test ($p < 0.05$) to determine the significance of the associated treatment outcomes (overall survival).